



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS DA EDUCAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

EVERTON RODRIGUES BARBOSA

**MODELO COLABORATIVO PARA A CONSTRUÇÃO E A PUBLICAÇÃO DE  
TESAUROS NO CONTEXTO DO *LINKED OPEN DATA*:  
APLICAÇÃO NO DOMÍNIO DA MÚSICA**

FLORIANÓPOLIS

2021

Everton Rodrigues Barbosa

**MODELO COLABORATIVO PARA A CONSTRUÇÃO E A PUBLICAÇÃO DE  
TESAUROS NO CONTEXTO DO *LINKED OPEN DATA*:  
APLICAÇÃO NO DOMÍNIO DA MÚSICA**

Tese submetida ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, para a obtenção do título de Doutor em Ciência da Informação.

**Orientador:** Prof. Dr. Angel Freddy Godoy Viera.

**Linha de pesquisa:** Informação e Tecnologia.

FLORIANÓPOLIS

2021

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Barbosa, Everton Rodrigues

Modelo colaborativo para a construção e a publicação de Tesouros no contexto do Linked Open Data : aplicação no domínio da Música / Everton Rodrigues Barbosa ; orientador, Angel Freddy Godoy Viera, 2021.  
373 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós Graduação em Ciência da Informação, Florianópolis, 2021.

Inclui referências.

1. Ciência da Informação. 2. Tesouros. 3. Linked Open Data. 4. Organização da Informação Musical. 5. Recuperação da Informação Musical. I. Godoy Viera, Angel Freddy. II. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Ciência da Informação. III. Título.

Everton Rodrigues Barbosa

**Modelo colaborativo para a construção e a publicação de Tesouros no contexto do**  
***Linked Open Data*: aplicação no domínio da Música**

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Moisés Lima Dutra, Dr.  
Universidade Federal de Santa Catarina

Profa. Camila Monteiro de Barros, Dra.  
Universidade Federal de Santa Catarina

Prof. Rogério Aparecido de Sá Ramalho, Dr.  
Universidade Federal de São Carlos

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Ciência da Informação.

---

Prof. Adilson Luiz Pinto, Dr.  
Coordenador do Programa de Pós-Graduação em Ciência da Informação

---

Prof. Angel Freddy Godoy Viera, Dr.  
Orientador

Florianópolis, 2021



Dedico este trabalho aos meus pais, Aldo e Cleonilde.

## AGRADECIMENTOS

Gratidão a Deus, força maior que me deu fôlego de vida, me conduz por caminhos desconhecidos e faz realizar sonhos como este que estou vivendo.

O doutorado me reservou ensinamentos, momentos que marcarão para sempre a minha existência. Este é o final de um ciclo acadêmico muito especial que não se resume à pesquisa científica, pois durante todo o processo pude conhecer meus limites pessoais e adquirir forças para superá-los, me tornando um ser humano melhor.

No decurso dessa jornada, pessoas especiais passaram por mim e deixaram o seu “pouquinho”, que quando somado a outros tantos me fizeram grande, contribuindo não apenas com a minha trajetória acadêmica ou profissional, mas, também, com o aperfeiçoamento do meu caráter.

Entre essas pessoas estão os meus queridos pais, Cleonilde e Aldo, para os quais dedico este trabalho com muito amor. Obrigado por acreditarem em mim e apoiarem minhas escolhas. Sem vocês eu não seria o que sou hoje.

Gratidão a Micael, meu companheiro de jornadas, e ao amor incondicional recebido por Ravi, meu amigo leal/ filho de quatro patas. Vocês me ensinam diariamente a ser mais paciente e a acreditar que tempos melhores virão, mesmo diante de tantas adversidades.

Sou grato também aos colegas e amigos que o doutorado me deu: Adriana, Jean, Josiane e Raquel. Obrigado pelas palavras de motivação, pelos momentos de descontração e pela troca de conhecimentos. Entre os laboratórios e bancos da UFSC, bares e trilhas de Floripa sorrimos juntos, compartilhamos aflições e vibramos com as pequenas conquistas. Esses momentos não se apagarão da minha memória.

Agradeço a todos os professores do PGCIN que contribuíram com a minha formação acadêmica, em especial ao meu orientador Angel por todo o apoio.

Não seria possível chegar até aqui sem o apoio da Universidade Federal do Rio Grande do Norte (UFRN), a qual me orgulho de fazer parte. Gratidão à Escola de Música da UFRN e a todos os colaboradores da Biblioteca Setorial Pe. Jaime Diniz, especialmente Elizabeth e David. Vocês são inspiração para mim.

A todos estes, o meu respeito e a minha profunda gratidão.

“Sem conhecimento, o homem permanece sempre muito próximo do pólo dos que não percebem a extensão do mundo em que vivem, em que circulam” (BOCCEGA, 2002, p. 12).

## RESUMO

Esta pesquisa tem como foco a construção colaborativa e a publicação de tesouros como *Linked Open Data*. Os tesouros foram escolhidos por serem considerados os tipos de Sistemas de Organização do Conhecimento com maior adesão ao modelo de dados *Simple Knowledge Organization System* e ao conjunto de práticas associadas ao paradigma do *Linked Data*. Assim, em razão da necessidade de atualização constante e a participação efetiva de diversos profissionais atuando de forma coordenada parte-se do pressuposto de que a adoção de práticas colaborativas pode agilizar a práxis de construção e manutenção. Ademais, construir um tesouro torna-se ainda mais desafiador quando se busca atender a demandas de informação de usuários heterogêneos. Nesse ponto, destaca-se o domínio da Música como o campo a ser estudado nesta pesquisa. Isto posto, organizar a informação musical é uma tarefa complexa em virtude das várias possibilidades de aplicação da música. Logo, admite-se que ela deve ser organizada com o propósito de atender às demandas de usuários especialistas e não especialistas simultaneamente. Dessa forma, o objeto de estudo da tese é a construção colaborativa e a publicação de tesouros no âmbito do *Linked Open Data*. De modo geral, objetiva-se: **Propor um modelo colaborativo para a construção e a publicação de tesouros no contexto do *Linked Open Data*, tendo como área de aplicação o domínio da Música.** A investigação é caracterizada como de natureza aplicada, exploratória e descritiva e de abordagem qualitativa. Para a realização do estudo, os dados foram coletados a partir de: questionários semiestruturados aplicados aos gestores de vocabulários controlados do domínio da Música em unidades de informação especializadas no Brasil; pesquisa documental, em que fez-se a análise de manuais e políticas de indexação dessas unidades; e pesquisa bibliográfica, em que foram sucedidas buscas sistemáticas em bases de dados para coletar e averiguar a produção científica sobre os temas tesouros e *Linked Data*. Os dados foram avaliados e processados com o auxílio das ferramentas Nvivo e Mendeley. Os métodos de análise empregados foram a Teoria Fundamentada em Dados e a Análise de conteúdo de Bardin. Conclui-se que o modelo colaborativo desenvolvido representa uma contribuição para a área da Ciência da Informação. A proposta se destaca pela reunião das abordagens mais atuais de construção de tesouros atreladas às demandas de sua publicação como *Linked Open Data*, sendo o principal contributo o desenvolvimento de um arcabouço teórico representado pela estruturação dos princípios norteadores. Esses princípios funcionam como o núcleo ou a base teórico-metodológica para o desdobramento de um ambiente colaborativo voltado para a construção e para a publicação de tesouros como *Linked Open Data*.

**Palavras-chave:** Tesouros. *Linked Open Data*. Práticas colaborativas. Organização da Informação Musical. Recuperação da Informação Musical.

## ABSTRACT

This research focuses on collaborative construction and the publication of thesauri as Linked Open Data. Thesaurus were chosen because they are considered the types of Knowledge Organization Systems with greater adherence to the Simple Knowledge Organization System data model and the set of practices associated with the Linked Data paradigm. Thus, due to the need for constant updating and the effective participation of several professionals acting in a coordinated manner, it is assumed that the adoption of collaborative practices can accelerate the construction and maintenance praxis. Furthermore, building a thesaurus becomes even more challenging when seeking to meet the information demands of heterogeneous users. At this point, the domain of Music stands out as the field to be studied in this research. That said, organizing musical information is a complex task due to the various possibilities of music application. Therefore, it is admitted that it must be organized with the purpose of simultaneously meeting the demands of expert and non-specialist users. In this way, the object of study of the thesis is the collaborative construction and the publication of thesauri within the ambit of Linked Open Data. In general, the objective is: **To propose a collaborative model for the construction and publication of thesaurus in the context of Linked Open Data, with the domain of Music as its application area.** The investigation is characterized as of an applied, exploratory and descriptive nature and with a qualitative approach. To perform the study, data were collected from: semi-structured questionnaires applied to managers of controlled vocabularies in the Music domain in specialized information units in Brazil; documentary research, in which manuals and indexing policies for these units were analyzed; and bibliographic research, in which systematic searches were performed in databases to collect and verify the scientific production on thesaurus and Linked Data themes. The data were evaluated and processed with the assistance of the Nvivo and Mendeley tools. The methods of analysis used were Grounded Theory and Content Analysis by Bardin. It is concluded that the developed collaborative model represents a contribution to the area of Information Science. The proposal stands out for bringing together the most current approaches to building thesaurus linked to the demands of its publication as Linked Open Data. The main contribution being the development of a theoretical framework represented by the structuring of the guiding principles. These principles act as the nucleus or the theoretical-methodological basis for the unfolding of a collaborative environment focused on the construction and the publication of thesauri as Linked Open Data.

**Keywords:** Thesaurus. Linked Open Data. Collaborative practices. Organization of Musical Information. Musical Information Retrieval.

## RESUMEN

Esta investigación se centra en la construcción colaborativa y la publicación de tesauros como *Linked Open Data*. Se eligieron Tesauros porque se consideran los tipos de Sistemas de Organización del Conocimiento con mayor adherencia al modelo de datos del *Simple Knowledge Organization System* y al conjunto de prácticas asociadas al paradigma de *Linked Data*. De este modo, por la necesidad de actualización constante y la participación efectiva de varios profesionales actuando de manera coordinada, se asume que la adopción de prácticas colaborativas puede agilizar la praxis de construcción y mantenimiento. Además, construir un tesaurus se vuelve aún más desafiante cuando se busca satisfacer las demandas de información de usuarios heterogéneos. En este punto, el dominio de la Música se destaca como el campo a estudiar en esta investigación. Dicho eso, organizar la información musical es una tarea compleja debido a las diversas posibilidades de aplicación de la música. Por tanto, se admite que debe organizarse con la finalidad de atender simultáneamente las demandas de usuarios expertos y no especialistas. Así, el objeto de estudio de la tesis es la construcción colaborativa y la publicación de tesauros en el ámbito de *Linked Open Data*. De modo general, el objetivo es: **Proponer un modelo colaborativo para la construcción y publicación de tesauros en el contexto de *Linked Open Data*, con el dominio de la Música como área de aplicación.** La investigación se caracteriza por ser de carácter aplicado, exploratorio y descriptivo y con un enfoque cualitativo. Para la realización del estudio se recogieron datos de: cuestionarios semiestructurados aplicados a gestores de vocabularios controlados en el dominio Música en unidades de información especializadas en Brasil; investigación documental, en la que se analizaron manuales y políticas de indexación de estas unidades; e investigación bibliográfica, en la que se realizaron búsquedas sistemáticas en bases de datos para recopilar y conocer la producción científica sobre tesauros y temas de *Linked Data*. Los datos se evaluaron y procesaron con la ayuda de las herramientas Nvivo y Mendeley. Los métodos de análisis empleados fueron la Teoría Fundamentada en Datos y el análisis de contenido de Bardin. Se concluye que el modelo colaborativo desarrollado representa una contribución al área de Ciencias de la Información. La propuesta destaca por reunir los enfoques más actuales para la construcción de tesauros vinculados a las exigencias de su publicación como *Linked Open Data*, siendo la principal contribución el desarrollo de un marco teórico representado por la estructuración de los principios rectores. Estos principios actúan como núcleo o base teórico-metodológica para el desarrollo de un entorno colaborativo enfocado a la construcción y publicación de tesauros como *Linked Open Data*.

**Palabras clave:** Tesaurus. *Linked Open Data*. Prácticas colaborativas. Organización de la información musical. Recuperación de información musical.

## LISTA DE FIGURAS

Figura 1 - Mapa conceitual dos principais temas abordados na tese.....	28
Figura 2 - Fluxograma dos processos colaborativos na construção de tesouros .....	63
Figura 3 - Estrutura de camadas da Web Semântica .....	66
Figura 4 - Exemplos de Triplos RDF e identificadores URI.....	69
Figura 5 - Exemplo de serialização em RDF/XML.....	70
Figura 6 - Exemplo de serialização em Notation 3 (N3).....	70
Figura 7 - Exemplo de serialização em <i>Turtle</i> .....	71
Figura 8 - Exemplo de serialização em N-Triples .....	71
Figura 9 - Exemplo de serialização em JSON-LD .....	72
Figura 10 - Serviço Endpoint do DBpedia .....	73
Figura 11 - Nuvem LOD .....	77
Figura 12 - Gráfico de elementos do <i>Thesaural Markup Language</i> (TML).....	82
Figura 13 - Exemplo de marcação utilizando os elementos TML.....	83
Figura 14 - Modelagem do AGROVOC baseado no OWL .....	88
Figura 15 - Fragmento da representação textual no UNESCO Thesaurus.....	90
Figura 16 - Representação do SKOS em RDF/XML .....	91
Figura 17 - Extrato do UNESCO Thesaurus representado em RDF <i>Graph</i> .....	92
Figura 18 - Conceitos publicados como LOD .....	93
Figura 19 - Utilização das propriedades do SKOS-XL no AGROVOC .....	96
Figura 20 - Utilização das propriedades do ISO-THES no UNESCO Thesaurus.....	99
Figura 21 - Exemplo de aplicação do UNESKOS no UNESCO Thesaurus .....	101
Figura 22 - Fatores que influenciam a percepção humana da música .....	103
Figura 23 - Página principal do DBTune.org .....	106
Figura 24 - Página da artista no BBC Music .....	107
Figura 25 - Catálogo online da RISM .....	108
Figura 26 - <i>Musical Instrument Museums Online</i> (MIMO) .....	109
Figura 27 - Página principal do MIDI <i>Linked Data</i> .....	110
Figura 28 - Página principal do DOREMUS.....	111
Figura 29 - Página principal do <i>MusicBrainz</i> .....	112
Figura 30 - Página principal do <i>AcousticBrainz</i> .....	113
Figura 31 - Página principal do <i>The Music Ontology</i> .....	114
Figura 32 - Página principal do <i>Linked Jazz</i> .....	115

Figura 33 - Página principal do <i>Sheet Music Consortium</i> .....	117
Figura 34 - Página inicial do Jamendo .....	118
Figura 35 - Caracterização da pesquisa .....	122
Figura 36 – Visão geral das práticas colaborativas na construção dos vocabulários controlados .....	167
Figura 37 – Princípios para a construção e publicação de tesouros no contexto do LOD.....	169
Figura 38 - Árvore de palavras do princípio do consenso .....	173
Figura 39 - Nuvem de palavras do princípio do consenso .....	174
Figura 40 - Árvore de palavras do princípio do controle editorial .....	176
Figura 41 - Nuvem de palavras do princípio do controle editorial.....	179
Figura 42 - Árvore de palavras do princípio da representação multilíngue .....	181
Figura 43 - Nuvem de palavras do princípio da representação multilíngue.....	184
Figura 44 - Árvore de palavras do princípio da interoperabilidade.....	186
Figura 45 - Nuvem de palavras do princípio da interoperabilidade .....	189
Figura 46 - Árvore de palavras da categoria consistência .....	191
Figura 47 - Árvore de palavras da categoria integridade.....	193
Figura 48 - Nuvem de palavras do princípio da Avaliação da Qualidade.....	194
Figura 49 - Árvore de palavras do princípio do acesso .....	196
Figura 50 - Nuvem de palavras do princípio do Acesso.....	198
Figura 51 - Árvore de palavras do princípio do reuso .....	200
Figura 52 - Nuvem de palavras do princípio do reuso .....	202
Figura 53 - Gráfico hierárquico dos princípios mais citados entre os estudos analisados ....	204
Figura 54 - Cluster por similaridade de palavras.....	205
Figura 55 - Gráfico hierárquico dos estudos que contemplam o maior número de princípios .....	206
Figura 56 - Modelo colaborativo para a construção e publicação de tesouros do domínio da Música como <i>Linked Open Data</i> .....	207
Figura 57 - Diagrama geral do modelo colaborativo para a construção e publicação de tesouros do domínio da Música como <i>Linked Open Data</i> .....	211
Figura 58 – Dimensão dos grupos de gestão do tesouro .....	212
Figura 59 – Dimensão da comunidade do LOD do domínio da Música .....	214
Figura 60 – Dimensão da Comunidade de Usuários .....	216
Figura 61 – Etapas da atividade de Planejamento .....	217
Figura 62 - Visão geral do fluxo de tarefas na edição dos tesouros .....	224



Figura 63 - Etapas da atividade de compilação de termos .....	233
Figura 64 - Fluxograma da atividade de compilação de termos candidatos.....	234
Figura 65 - Exemplo de utilização de bases de conhecimento para o reconhecimento de entidades nomeadas em fragmentos textuais.....	241
Figura 66 - Etapas da atividade de seleção de termos .....	243
Figura 67 - Árvore de decisão para a seleção de termos .....	247
Figura 68 - Coordenação e subordinação de conceitos .....	249
Figura 69 – Exemplo de grupos e subgrupos de conceitos .....	250
Figura 70 - Etapas da atividade de preparação dos dados para a publicação .....	255
Figura 71 - Visão geral da proposta de modelagem conceitual do tesouro .....	257
Figura 72 - Exemplo de modelagem para os dados dos tesouros usando os modelos SKOS, SKOS-XL, ISO-THES, UNESKOS, RDFS e FOAF .....	260
Figura 73 - Exemplo de modelagem de dados (esquemas conceituais e relações semânticas) .....	262
Figura 74 - Uso do código 303 com negociação de conteúdo.....	266
Figura 75 - Formato de serialização RDF/XML .....	268
Figura 76 - Criando projetos na ferramenta <i>OpenRefine</i> .....	270
Figura 77 - Representação dos dados a partir de vocabulários semânticos.....	271
Figura 78 - Etapas da atividade de interligação dos conjuntos de dados .....	273
Figura 79 – Modelo de Hub para o mapeamento de conceitos do tesouro com outros vocabulários.....	276
Figura 80 - Conjuntos de dados do domínio da Música que possuem entidades vinculadas ao Wikidata .....	280
Figura 81 - Etapas da atividade de validação dos dados .....	283
Figura 82 - Serviço de validação sintática para documentos RDF do W3C .....	287
Figura 83 - Serviço de validação SKOS do W3C .....	288
Figura 84 - Etapas da atividade de publicação .....	290
Figura 85 - Informação sobre os dados do tesouro AGROVOC .....	291
Figura 86 - Fragmento da descrição legível por máquina dos dados do AGROVOC.....	293
Figura 87 - Navegação e visualização da versão SKOS do UNESCO Thesaurus .....	295
Figura 88 - Diagramação do MIMO Thesaurus na forma de árvore de conceitos .....	296
Figura 89 - Exemplo de consulta SPARQL no vocabulário Getty.....	297
Figura 90 - Resultados da consulta SPARQL no vocabulário Getty.....	298
Figura 91 - Atividade de <i>Feedback</i> .....	299

## LISTA DE QUADROS

Quadro 1 - Modelos computacionais e estatísticos aplicados a construção automática de tesouros .....	46
Quadro 2 - Etapas para construção e manutenção de tesouros .....	49
Quadro 3 - Atores, materiais e métodos distribuídos por processos .....	61
Quadro 4 - Melhores práticas para a publicação de dados na <i>web</i> .....	78
Quadro 5 - Formatos e iniciativas para representação dos tesouros na <i>web</i> .....	85
Quadro 6 – Padrões para a representação de registros bibliográficos .....	86
Quadro 7 - Principais propriedades do SKOS <i>Core Vocabulary</i> .....	89
Quadro 8 - Principais tesouros publicados em LOD que utilizam o SKOS .....	94
Quadro 9 - O vocabulário SKOS-XL .....	95
Quadro 10 – Exemplos de classes do vocabulário ISO-THES.....	98
Quadro 11 - UNESKOS <i>Vocabulary Specification</i> .....	100
Quadro 12 - Principais desafios e oportunidades para os dados conectados do domínio da Música .....	121
Quadro 13 - Unidades de Informação especializadas em Música no Brasil .....	124
Quadro 14 - Respostas das unidades de informação em relação ao uso de vocabulários controlados na indexação.....	126
Quadro 14 - Resultados das buscas nas bases de dados .....	128
Quadro 16 – Etapas da pesquisa.....	131
Quadro 17 –Fases, etapas e procedimentos para a análise e tratamento dos dados conforme a Teoria Fundamentada em Ddados .....	133
Quadro 18 - Perfil dos especialistas que avaliaram o modelo.....	134
Quadro 19 – Vocabulários controlados, gestores e os tipos de unidades de informação em que atuam .....	136
Quadro 20 - Origem dos vocabulários controlados .....	138
Quadro 21 - Equipe técnica que atua na construção dos vocabulários controlados .....	139
Quadro 22 – Estruturação dos termos nos vocabulários controlados.....	140
Quadro 23 - Percorso metodológico na construção dos vocabulários controlados .....	141
Quadro 24 - Profissionais envolvidos com a construção dos vocabulários controlados .....	143
Quadro 25 – Modificações ou sugestões de termos do vocabulário.....	144
Quadro 26 - Canais de comunicação e procedimentos para acolher as sugestões da comunidade .....	145

Quadro 27 - Participação das unidades de informação em redes cooperativas de catalogação ou indexação .....	146
Quadro 28 - Seleção dos termos candidatos à inclusão no vocabulário controlado.....	148
Quadro 29 - Uso de vocabulários existentes para a seleção de termos candidatos .....	149
Quadro 30 - Adaptação dos termos compilados a partir de vocabulários controlados existentes .....	150
Quadro 31 - Responsáveis pela compilação dos termos do domínio da Música.....	151
Quadro 32 - Competências básicas atribuídas aos profissionais responsáveis pela compilação de termos do domínio da Música.....	152
Quadro 33 - Etapas em que podem ocorrer a compilação de termos .....	153
Quadro 34 - Compilação e tradução de termos em outros idiomas.....	154
Quadro 35 - Especificação de termos .....	155
Quadro 36 - Processamento automático, semiautomático ou manual aplicado à validação dos termos .....	158
Quadro 37 - Consenso sobre a inclusão de termos no vocabulário .....	159
Quadro 38 – Importância da participação dos usuários na etapa de testes e avaliação dos vocabulários controlados do domínio da Música .....	161
Quadro 39 - Controle distribuído e compartilhado do vocabulário controlado.....	162
Quadro 40 – Recursos de colaboração e etapas em que são utilizados .....	163
Quadro 41 – Características dos sistemas de gestão e uso de tecnologia do Linked Data .....	163
Quadro 42 - Percepção da colaboração nas etapas de construção dos vocabulários controlados .....	165
Quadro 43 - Funções dos membros da equipe editorial .....	220
Quadro 44 - Abordagens para o controle de simultaneidade nas ferramentas de edição dos tesouros .....	225
Quadro 45 - Tipos de licenças abertas para a publicação de tesouros como LOD.....	229
Quadro 46 – Disciplinas e subdisciplinas do domínio da Música.....	244
Quadro 47 - Critérios para a definição de termos preferenciais do tesouro .....	252
Quadro 48 - Exemplo de mapeamento de metadados .....	263
Quadro 49 - Exemplo de padrões de URI para um tesouro como LOD.....	265
Quadro 50 - Conjuntos de dados e propriedades para a execução do mapeamento no DBPedia, Wikidata e MusicBrainz .....	277
Quadro 51 – Clareza na descrição e relação entre as dimensões, princípios e atividades de construção.....	302

Quadro 52 – Estruturação das etapas e conformidade às práticas do LOD.....	303
Quadro 53 – Aplicação no domínio da música, em outros contextos ou domínios .....	304
Quadro 54 – Resultado da avaliação dos especialistas.....	305
Quadro 55 - Critérios de inclusão e exclusão .....	341

## LISTA DE ABREVIATURAS E SIGLAS

AAC	<i>The American Art Collaborative</i>
ALISE	<i>Association for Library and Information Science Education</i>
ANSI	<i>American National Standards Institute</i>
API	<i>Application Programming Interface</i>
BBC	<i>British Broadcasting Corporation</i>
BIBFRAME	<i>Bibliographic Framework</i>
BnF	Biblioteca Nacional da França
CC	<i>Creative Commons</i>
CERES	<i>California Environmental Resources Evaluation System</i>
CI	Ciência da Informação
CIDOC	<i>Committee for Documentation of the International Council of Museums</i>
CRM	<i>Conceptual Reference Model</i>
CSV	<i>Comma-Separated Values</i>
DOREMUS	<i>Doing Reusable Musical Data</i>
EURECOM	Centro de Pesquisa em Sistemas de Comunicação
FRBR	<i>Functional Requirements for Bibliographic Records</i>
GEMET	<i>General Multilingual Environmental Thesaurus</i>
GRID	<i>Global Research Identifier Database</i>
GTO	<i>Generic Thesaural Ontology</i>
HTML	<i>HyperText Markup Language</i>
HTTP	<i>HyperText Transfer Protocol</i>
IAML	<i>International Association of Music Libraries, Archives and Documentation Centres</i>
IDF	<i>Inverse Document Frequency</i>
IM	Informação Musical
IMS	<i>International Musicological Society</i>
ISO	<i>International Organization for Standardization</i>
JSON-LD	<i>JavaScript Object Notation JavaScript Object Notation for Linked Data</i>
LCSH	<i>Library of Congress Subject Headings</i>
LIRMM	<i>Laboratory of Informatics, Robotics and Microelectronics of Montpellier</i>
LOD	<i>Linked Open Data</i>

MADS	<i>Metadata Authority Description</i>
MARC	<i>Machine Readable Cataloging</i>
MeSH	<i>Medical Subject Headings</i>
MIDI	<i>Musical Instrument Digital Interface</i>
MIMO	<i>Musical Instrument Museums Online</i>
NCIt	<i>National Cancer Institute Thesaurus</i>
NISO	<i>National Information Standards Organization</i>
OBO	<i>Open Biomedical Ontologies</i>
OC	Organização do Conhecimento
OCLC	<i>Online Computer Library Center</i>
OIM	Organização da Informação Musical
OLV	<i>Open Link Virtuoso</i>
OPAC	<i>Online Public Access Catalog</i>
OWL	<i>Web Ontology Language</i>
PLN	Processamento de Linguagem Natural
PP	Philharmonie de Paris
RDA	<i>Resource Description and Access</i>
RDF	<i>Resource Description Framework</i>
RF	Rádio France
RI	Recuperação da Informação
RIM	Recuperação da Informação Musical
RISM	<i>Répertoire International des Sources Musicales</i>
SGML	<i>Standard Generalized Markup Language</i>
SKOS	<i>Simple Knowledge Organization System</i>
SMS	<i>Sheet Music Consortium</i>
SOC	Sistemas de Organização do Conhecimento
SPARQL	<i>SPARQL Protocol and RDF Query Language</i>
SRI	Sistemas de Recuperação da Informação
SWAD-E	<i>Semantic Web Advanced Development for Europe</i>
SWDWG	<i>Semantic Web Deployment Working Group</i>
TGT	Teoria Geral da Terminologia
TFD	Teoria Fundamentada em Dados
TML	<i>Thesaural Markup Language</i>

UCLA	<i>University of California, Los Angeles</i>
URI	<i>Uniform Resource Identifiers</i>
USGS/BRD	<i>United States Geological Survey Biological Resources Division</i>
W3C	<i>World Wide Web Consortium</i>
WS	<i>Web Semântica</i>
XML	<i>eXtensible Markup Language</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	23
1.1	DELIMITAÇÃO DO PROBLEMA E DO OBJETO DE ESTUDO.....	24
1.2	JUSTIFICATIVAS.....	29
1.3	OBJETIVOS.....	35
1.3.1	Objetivo geral.....	35
1.3.2	Objetivos específicos.....	35
1.4	ESTRUTURA DA TESE .....	36
<b>2</b>	<b>FUNDAMENTOS TEÓRICOS</b> .....	37
2.1	A CONSTRUÇÃO DE TESAUROS .....	37
2.1.1	O conceito como elemento central do tesouro.....	38
2.1.2	Garantias para a representação de conceitos ou seleção de termos.....	41
2.1.3	Métodos e etapas de construção .....	44
2.1.4	Abordagens colaborativas na construção de tesauros .....	50
2.2	O <i>LINKED DATA</i> , SEUS PRINCÍPIOS E TECNOLOGIAS .....	64
2.2.1	Identificadores Uniformes de Recursos (URI).....	67
2.2.2	O <i>Resource Description Framework</i> (RDF) .....	68
2.2.3	Sintaxes de serialização para os documentos RDF .....	69
2.2.4	SPARQL: protocolo e linguagem de consulta RDF .....	72
2.2.5	Atribuição de licenças para o reuso dos dados .....	74
2.2.6	Melhores práticas para a publicação de dados na <i>web</i> .....	77
2.3	TESAUROS NA <i>WEB</i> .....	80
2.3.1	Tesauros representados em XML, RDF e OWL .....	81
2.3.2	Tesauros representados em SKOS.....	89
2.3.3	Extensões do vocabulário SKOS .....	95
2.4	INICIATIVAS DO <i>LINKED DATA</i> NO DOMÍNIO DA MÚSICA .....	102



2.4.1	Projetos de <i>Linked Data</i> no âmbito da Informação Musical .....	105
2.4.2	Desafios e oportunidades para os dados conectados no domínio da Música .....	118
3	<b>PERCURSO METODOLÓGICO</b> .....	122
3.1	CARACTERIZAÇÃO DA PESQUISA.....	123
3.2	UNIVERSO E AMOSTRA.....	124
3.3	COLETA DOS DADOS.....	128
3.3.1	<b>Pesquisa Bibliográfica</b> .....	128
3.3.2	<b>Questionário semiestruturado</b> .....	129
3.3.3	<b>Pesquisa documental</b> .....	130
3.4	ETAPAS DA PESQUISA .....	131
3.5	ANÁLISE E TRATAMENTO DOS DADOS .....	132
3.6	AVALIAÇÃO DO MODELO .....	134
4	<b>ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS</b> .....	136
4.1	PRÁTICAS COLABORATIVAS NA CONSTRUÇÃO DE VOCABULÁRIOS DO DOMÍNIO DA MÚSICA.....	136
4.1.1	<b>Caracterização das Unidades de Informação: acervo e sistemas de gestão</b> .....	137
4.1.2	<b>Aspectos gerenciais e metodológicos dos vocabulários controlados</b> .....	138
4.1.3	<b>Elaboração e manutenção do vocabulário controlado</b> .....	142
4.1.4	<b>Compilação ou levantamento de termos</b> .....	148
4.1.5	<b>Especificação ou supressão de termos</b> .....	154
4.1.6	<b>Registro e validação de termos</b> .....	156
4.1.7	<b>Testes e avaliação</b> .....	160
4.1.8	<b>Ferramentas colaborativas e representação dos vocabulários controlados na <i>web</i></b> .....	162
4.1.9	<b>Percepção da colaboração nas etapas de construção</b> .....	164
4.1.10	<b>Considerações gerais sobre as práticas colaborativas na construção de vocabulários controlados do domínio da Música</b> .....	166

4.2	PRINCÍPIOS PARA A CONSTRUÇÃO E A PUBLICAÇÃO DE TESAuros COMO <i>LINKED OPEN DATA</i> .....	168
4.2.1	Estruturação e análise dos princípios norteadores.....	169
4.2.2	Considerações gerais sobre os princípios evidenciados na literatura .....	203
5	MODELO COLABORATIVO PARA A CONSTRUÇÃO E A PUBLICAÇÃO DE TESAuros DO DOMÍNIO DA MÚSICA COMO <i>LINKED OPEN DATA</i> .....	207
5.1	PLANEJAMENTO.....	217
5.1.1	Definição dos objetivos, escopo e métodos de construção do tesouro .....	218
5.1.2	Papeis de atuação dos profissionais envolvidos .....	220
5.1.3	Seleção de modelos de dados e ferramentas para a gestão e a publicação do tesouro.....	221
5.1.4	Definição de licenças abertas para o reuso dos dados.....	226
5.1.5	Políticas de gestão, atualização e revisão .....	230
5.2	COMPILAÇÃO DE TERMOS .....	232
5.2.1	Coleta de dados .....	235
5.2.2	Pré-processamento e transformação dos dados.....	237
5.2.3	Mineração de texto e dados .....	239
5.3	SELEÇÃO DE TERMOS.....	242
5.3.1	Análise de termos candidatos .....	243
5.3.2	Categorização e relações lógico-semânticas .....	248
5.3.3	Especificação e registro de termos preferenciais .....	251
5.4	PREPARAÇÃO DOS DADOS PARA A PUBLICAÇÃO .....	254
5.4.1	Modelagem e mapeamento dos dados.....	255
5.4.2	Padrões para os identificadores únicos (URI).....	265
5.4.3	Conversão dos registros para o formato RDF .....	268
5.4.4	Armazenamento e gerenciamento dos dados RDF .....	272
5.5	INTERLIGAÇÃO DOS CONJUNTOS DE DADOS .....	273
5.5.1	Mapeamento entre vocabulários .....	274

<b>5.5.2 Reconciliação dos dados</b> .....	278
<b>5.6 VALIDAÇÃO DOS DADOS</b> .....	282
<b>5.6.1 Restrições de integridade para o tesouro</b> .....	283
<b>5.6.2 Serviços de validação de dados</b> .....	287
<b>5.7 PUBLICAÇÃO</b> .....	289
<b>5.7.1 Informações sobre o conjunto de dados do tesouro</b> .....	290
<b>5.7.1 Busca, exploração e visualização dos dados</b> .....	294
<b>5.8 FEEDBACK</b> .....	298
<b>6 AVALIAÇÃO DO MODELO COM ESPECIALISTAS</b> .....	302
<b>7 CONSIDERAÇÕES FINAIS</b> .....	307
<b>REFERÊNCIAS</b> .....	312
<b>APÊNDICES</b> .....	340
<b>APÊNDICE A – CRITÉRIOS DE INCLUSÃO E EXCLUSÃO PARA AS BUSCAS SISTEMÁTICAS NAS BASES DE DADOS</b> .....	341
<b>APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO</b>	342
<b>APÊNDICE C – QUESTIONÁRIO</b> .....	345
<b>APÊNDICE D – AMOSTRA TEÓRICA: DOCUMENTOS ANALISADOS</b> .....	352
<b>APÊNDICE E – INSTRUÇÕES PARA A ELABORAÇÃO DO PARECER</b> .....	360
<b>APÊNDICE F – PARECERES DOS ESPECIALISTAS</b> .....	361

## 1 INTRODUÇÃO

Na perspectiva do crescente volume de dados, o conceito de *Linked Data* surge em 2006 como um conjunto de práticas voltadas para a publicação e a estruturação de dados conectados na *web*, favorecendo a leitura e consumo tanto por humanos quanto por computadores. Esse conceito evoluiu e a partir de 2009 discute-se a importância de publicar dados como *Linked Open Data* (LOD), ou seja, dados conectados disponibilizados sob uma licença aberta com o objetivo de favorecer sua livre reutilização.

Nesse cenário, persiste um desafio: a necessidade de organizar a informação na *web* para que os diversos públicos possam recuperar conteúdo relevante mediante uma necessidade de informação. Atendendo a esse propósito, os projetos de *Linked Data* têm se beneficiado dos Sistemas de Organização do Conhecimento (SOC). Esses instrumentos, outrora aplicados à organização da informação em ambientes informacionais tradicionais – bibliotecas, arquivos e museus – se mostram úteis também nas iniciativas de organização da informação em ambientes digitais, principalmente no contexto da Web Semântica (WS). No entanto, para que a *web* possa se aproveitar de todos os benefícios dos SOC é preciso que eles estejam estruturados semanticamente como instâncias de metadados.

De modo geral, os SOC variam de estruturas planas a multidimensionais com certo grau de controle introduzido – pois recorre à conversão da linguagem natural para uma linguagem artificial –, exercem papel fundamental na indexação e classificação de conteúdo e, da mesma forma, oferecem suporte à recuperação da informação, inclusive no contexto das tecnologias digitais (ZENG, 2008; BRÄSCHER; CARLAN, 2010). São exemplos de SOC as taxonomias, folksonomias, ontologias, os sistemas de classificação e os tesouros. Este último se destaca por formar uma estrutura de termos correlacionados capazes de auxiliar na tarefa de indexação e na busca nos Sistemas de Recuperação da Informação (SRI).

Dessa maneira, pensando na capacidade de leitura e processamento dos SOC na *web*, o *Semantic Web Deployment Working Group* (SWDWG), do *World Wide Web Consortium* (W3C), introduziu o *Simple Knowledge Organization System* (SKOS), o qual surge como um modelo de dados que formaliza a estrutura básica dos SOC para um esquema de dados legível por máquina, dando suporte à representação, uso e interoperabilidade dos vocabulários controlados na *web* de dados conectados (ZOGHLAMI; KERHERVÉ; GERBÉ, 2011).

Esta pesquisa tem como foco o uso de tesouros estruturados em SKOS e seus modelos de extensão aplicados às iniciativas de organização e recuperação da informação em ambientes digitais, conforme recomenda o conjunto de melhores práticas do *Linked Data*. Assim, partindo

do pressuposto que os tesouros são particularmente úteis para representar domínios de especialidade, o domínio da Música foi escolhido como o campo a ser estudado. Por conseguinte, organizar a informação musical é uma tarefa complexa devido as várias possibilidades de aplicação da música, presente não apenas na vida cotidiana das pessoas, mas também na pesquisa acadêmica, no processo de ensino-aprendizagem e na prática instrumental.

Admite-se que a informação musical deve ser organizada com o propósito de atender a demandas de informação de um público heterogêneo, nomeadamente representado por usuários especialistas e não especialistas. Outrossim, infere-se que publicar dados sobre arte, música, arquitetura e objetos do patrimônio cultural, como LOD, tem sido uma necessidade e tendência mundial (PATRÍCIO, 2012; GRIMOÛARD, 2014; HARPRING, 2015; KNOBLOCK et al., 2017; FINK, 2018). Essa iniciativa implica não somente no uso de tecnologias e padrões para a descrição de objetos ou recursos musicais na *web*, mas também no emprego de vocabulários para representá-los ou descrevê-los.

Nessa ótica, os tesouros se destacam entre os SOC por serem os instrumentos de indexação e recuperação da informação com maior adesão às práticas do *Linked Data* (PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012). No entanto, sua elaboração enfrenta alguns desafios, que serão melhores discutidos na sequência.

## 1.1 DELIMITAÇÃO DO PROBLEMA E DO OBJETO DE ESTUDO

Em 2004, o W3C recomendou que as bibliotecas publicassem seus conjuntos de dados bibliográficos utilizando tecnologias da WS. Desde então, inúmeros modelos de dados têm sido desenvolvidos e experimentados nos projetos de LOD nesse âmbito. De acordo com Torre-Bastida, González-Rodríguez e Villar-Rodríguez (2015, p. 114, tradução nossa)<sup>1</sup>, “[...] as experiências com vocabulários controlados, ontologias, metadados e busca são os candidatos preferidos para implementar os princípios que promulgam o paradigma do *Linked Data*.”

No que se refere aos vocabulários controlados, os tesouros são particularmente importantes em domínios, os quais apresentam significativa quantidade de conhecimento compilado (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 217). Segundo Hanandeh (2013), a atualização constante dos tesouros contribui para tornar os SRI mais efetivos, pois o corpo de saberes dos diversos campos científicos evolui continuamente. Não obstante, Mari Vállez et al.

---

<sup>1</sup> “la experiencia en vocabularios controlados, ontologías, metadatos y búsquedas convierte a los expertos del sector bibliotecario en candidatos perfectos para implementar los principios que promulga el paradigma linked data” (TORRE-BASTIDA; GONZÁLEZ-RODRÍGUEZ; VILLAR-RODRÍGUEZ, 2015, p. 114).

(2015) afirmam que a construção desses vocabulários pode ser oneroso em razão da necessidade de envolvimento humano, implicando em custo e tempo para a sua elaboração.

Em vista disso, a estrutura de conceitos dos tesauros busca sistematizar e/ou representar um domínio específico do conhecimento. Por essa razão, esses instrumentos são elaborados utilizando a literatura científica da área como base para a coleta de termos que representam os conceitos do domínio. O controle terminológico é um dos requisitos que oferece rigor à elaboração e à manutenção, impedindo que os termos selecionados sejam incorporados ou até usados erroneamente (PINHO, 2014). Ademais, construir um vocabulário controlado torna-se ainda mais desafiador quando o SRI busca atender demandas de informação de usuários especialistas e não especialistas simultaneamente (SILVA, 2015; HARPRING, 2016).

Nesse ponto, destaca-se a área de Recuperação da Informação Musical (RIM), definida como “[...] um esforço de pesquisa multidisciplinar que procura desenvolver esquemas inovadores de busca baseados em conteúdo, novas interfaces e mecanismos de entrega em rede para tornar o vasto conteúdo musical do mundo acessível a todos.” (DOWNIE, 2004, p. 12, tradução nossa)<sup>2</sup>.

A RIM é um campo de pesquisa que busca soluções para melhor processar e recuperar a Informação Musical. Sua agenda de pesquisa está concentrada em dois focos principais: no computador e no usuário (SCHEDL; GÓMEZ; URBANO, 2014).

Quando tratamos da Informação Musical, aspectos ligados à organização e à representação da informação se apresentam de forma complexa, pois a Música é um campo do conhecimento e, ao mesmo tempo, se caracteriza como uma manifestação artístico-cultural. A música serve ao entretenimento da sociedade e, também, é uma área de ensino com agendas de pesquisa bem definidas no âmbito científico (DANNENBERG, 2001).

Uma das principais questões envolvidas na indexação e descrição de coleções musicais são apresentadas por Kelly (2010), a qual declara que as bases de dados precisam favorecer a busca e a recuperação de conteúdo musical, de modo que os índices possam ser pesquisados por músicos treinados e não treinados. À vista disso, o aumento das pesquisas *online*, tanto por especialistas quanto por não especialistas, também cria novas oportunidades para o profissional da informação que busca a melhor forma de obter recursos para os usuários. Muitos dos desafios da RIM decorrem da necessidade de oferecer meios relacionados ao contexto do usuário e ao cruzamento de informações na *web* (KELLY, 2015).

---

<sup>2</sup> “*Music Information Retrieval (MIR) is a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world’s vast store of music accessible to all*” (DOWNIE, 2004, p. 12).

O potencial para o uso dos dados conectados e tecnologias da WS no domínio da Música permite que compositores, artistas, gêneros e outros conceitos relacionados aos recursos musicais possam se apresentar vinculados (DIX et al., 2014). Porém, os projetos de dados conectados no campo da Música enfrentam o desafio de atender às demandas de duas categorias de públicos: os grupos de especialistas, como críticos, pesquisadores e intérpretes; e o público geral. Assim, esses projetos necessitam de infraestrutura tecnológica essencial que compreenda o paradigma representacional dos dados, a mecânica para sua reutilização e as políticas de governança (BROWN et al., 2014).

Contudo, Fillon et al. (2014) declaram que um gerenciamento eficiente e dinâmico dos metadados associados aos recursos na *web* também é fundamental. Essa manutenção, ao ocorrer de maneira colaborativa, otimiza o processo contínuo de coleta e o enriquecimento semântico dos recursos musicais. Os dados conectados se apresentam como uma extensão natural dos modelos colaborativos subjacentes à criação e à disponibilização de dados pelas bibliotecas, visto serem partilháveis, extensíveis e facilmente reutilizáveis (PATRÍCIO, 2012).

No entanto, os princípios existentes do *Linked Data* são exíguos quando se edita de forma colaborativa os mesmos armazenamentos distribuídos de dados conectados, mais precisamente executando modificações concorrentes em um contexto de colaboração social. Assim, a edição colaborativa social no contexto do *Linked Data* é definida por Zarzour, Al-Ayyoub e Jararweh (2018, p. 90)<sup>3</sup> como [...] uma situação em que mais de um usuário trabalha em conjunto nas mesmas réplicas de dados e interage compartilhando operações para atingir um objetivo comum.” Desse modo, torna-se vital adotar métodos eficientes para a replicação e o gerenciamento eficiente dos dados (ZARZOUR; SELLAMI, 2012, 2014a, 2014b).

Recursos e tecnologias do LOD são aplicados em projetos comerciais, setores científicos e governamentais, porém sua aplicação nos repositórios de patrimônio cultural ainda é incipiente (FROSTERUS et al., 2013; VAN HOOLAND et al., 2013; VAN HOOLAND; VERBORGH, 2015; KNOBLOCK et al., 2017; FINK, 2018). Destacam-se alguns projetos desenvolvidos nos países da União Europeia e Estados Unidos, como são os casos do Europeana e do *The American Art Collaborative* (AAC), respectivamente.

Esses projetos dispõem de plataformas para integrar a busca e a recuperação da informação em acervos de instituições de patrimônio cultural de suas regiões. Acredita-se que os benefícios das tecnologias do LOD devem ser expandidos, e por isso a implementação de

---

<sup>3</sup> “Social collaborative editing can be defined as a situation where more than one user work together on the same replicas of data and interact by sharing operations in order to achieve a common goal” (ZARZOUR; AL-AYYOUB; JARARWEH, 2018, p. 90).

plataformas colaborativas é uma tendência nesses países, pois facilitam as pesquisas em várias coleções simultaneamente (EUROPEANA FOUNDATION, 2015; SCHOLZ, 2019; FINK, 2018).

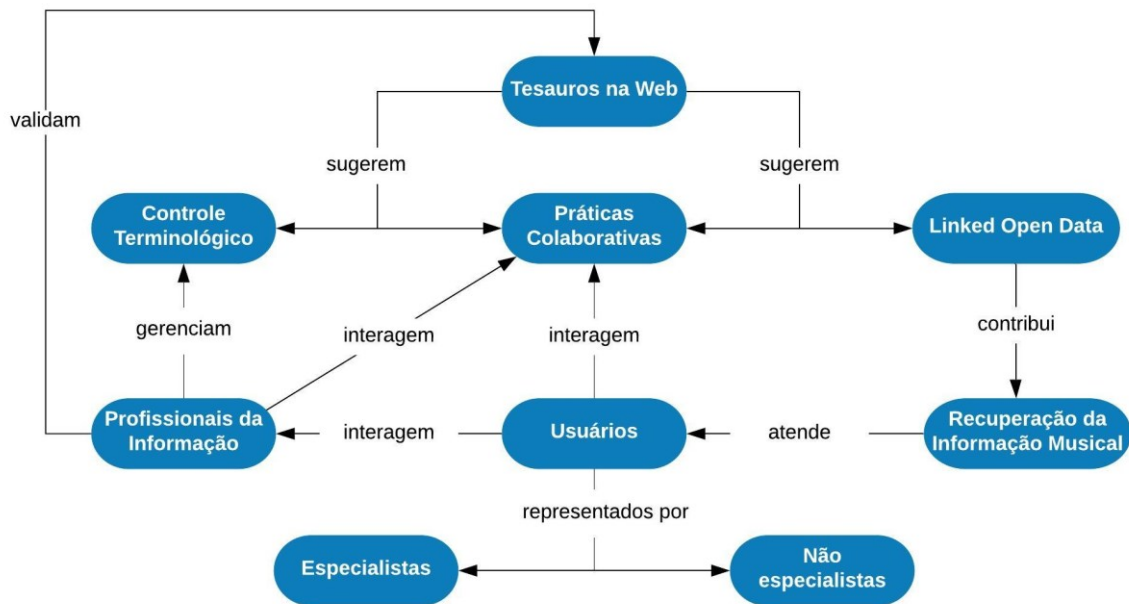
Quanto ao aspecto colaborativo, os projetos para a publicação de SOC como LOD procuram estabelecer direcionamentos que vão além do uso de tecnologias e recomendações oficiais do W3C. Em relação aos tesouros, a necessidade de atualização constante demanda no engajamento de todos os profissionais envolvidos, os quais devem operar colaborativamente para que os procedimentos dispensados obedeçam às diretrizes estabelecidas pela organização que elabora a terminologia.

Essa ação colaborativa parece ser uma alternativa positiva para agilizar no processo de construção e manutenção de tesouros no contexto do LOD e, por consequência, otimizar a Organização da Informação Musical em se tratando do uso adequado da terminologia. Nesse aspecto, os projetos de *Linked Data* no domínio da Música são caracterizados pela interconectividade entre conjuntos de dados distribuídos na *web*, visto que buscam estabelecer relacionamentos, facilitar a interoperabilidade e permitir que os dados possam ser reutilizados por outras aplicações. Para isso, necessita também de uma equipe multidisciplinar que atue de forma coordenada (FUTRELLE; DOWNIE, 2003; PAGE et al., 2013; NURMIKKO-FULLER et al., 2016).

Isto posto, o objeto de estudo da tese é a **construção colaborativa de tesouros no âmbito do LOD**. Estão intrínsecas à concepção do objeto de estudo três desafios principais – discutidos de forma isolada na literatura e que sugerem o alinhamento de práticas colaborativas – considerando o domínio da Música como área de aplicação (ilustrado na Figura 1): a necessidade de atualização constante dos tesouros; a participação efetiva de pessoas e organizações tanto nas atividades de construção como na publicação de tesouros como LOD; e a fundamental elaboração de um vocabulário que atenda às necessidades de busca e recuperação da informação de usuários especialistas e usuários não especialistas.



**Figura 1** - Mapa conceitual dos principais temas abordados na tese



**Fonte:** Elaborado pelo autor (2021).

A Figura 1 apresenta o mapa conceitual dos principais temas abordados na tese com vistas a melhor ilustrar seus relacionamentos. Observa-se que o conceito de tesauros na *web* sugere a adoção de padrões e tecnologias do LOD que, por sua vez, contribui dentre os vários domínios com a Recuperação da Informação Musical. O domínio foi selecionado devido a necessidade de atender demandas de usuários heterogêneos, representados por especialistas e não especialistas.

Esses sujeitos interagem com os profissionais da informação responsáveis pelo controle da terminologia do domínio. Ambos podem contribuir com a construção dos tesauros na *web* por via das práticas colaborativas. Por conseguinte, percebe-se que tanto o Controle Terminológico como a iniciativa do LOD sugerem a adoção de Práticas Colaborativas, podendo ser considerado como uma possível solução para a problemática em torno da elaboração de tesauros. Nesse caso, especificamente dedicados à Organização e Recuperação da Informação Musical.

Acrescenta-se a necessidade de validar a estrutura do vocabulário, pois uma vez estruturados como instância de metadados estes precisam atender a critérios de integridade, ou seja, os modelos de dados utilizados para representar o tesouro na *web* precisam ser aplicados corretamente para que a qualidade dos dados e da terminologia não seja prejudicada. Assim, a

validação é executada pelos profissionais da informação com o auxílio de tecnologias próprias para a verificação da integridade.

Diante da complexidade dos temas tratados e inter-relacionados, os pressupostos teóricos conduziram à formulação do seguinte questionamento: **Como elaborar de forma colaborativa tesouros do domínio da Música utilizando tecnologias do LOD?**

Esse questionamento leva à formulação de perguntas subjacentes – elucidados a partir da pesquisa bibliográfica e documental – que esta pesquisa busca responder. A saber: **Como as equipes de Controle Terminológico entendem, definem, implementam e praticam o trabalho colaborativo na construção de vocabulários controlados do domínio da Música? Quais tecnologias ou ferramentas colaborativas são utilizadas para integrar as equipes de Controle Terminológico e agilizar a elaboração de tesouros do domínio da Música?**

O estudo em realce está concentrado na construção colaborativa e na publicação de tesouros no contexto do *Linked Open Data*. Parte do pressuposto de que a vinculação de vocabulários controlados na *web* de dados é resultado do trabalho cooperante entre organizações e grupos de pessoas. Assim, a manutenção e o gerenciamento da terminologia nesses espaços demandam certo controle, visto que além de ser fruto de um processo compartilhado pode atender a demandas de informação de comunidades diversificadas.

De modo restrito, o foco desta pesquisa será dado aos procedimentos e tecnologias para a construção colaborativa e a publicação de tesouros do domínio da Música, sendo essa uma área que lida com um significativo volume de dados conectados e atende a múltiplos usuários.

## 1.2 JUSTIFICATIVAS

Serão apresentadas na sequência as justificativas científicas e pessoais desta pesquisa. A relevância científica deste estudo emerge a partir da compreensão dos fenômenos da Organização do Conhecimento (OC) e da RI, áreas da Ciência da Informação (CI) que se preocupam com a entrega de conteúdo relevante em coleções estáticas e dinâmicas. Nesse sentido, a CI enquanto campo interdisciplinar carece do suporte teórico-metodológico para fundamentar discussões em torno de outros campos do conhecimento (SARACEVIC, 1996; ARAÚJO, 2009).

No que se refere à **justificativa científica**, estudos sobre a OC no contexto do LOD confluem no uso de tecnologias, padrões ou modelos de dados desenvolvidos com a finalidade

de melhor organizar a informação. A alternativa encontrada para enriquecer semanticamente os recursos na *web*, considerando o paradigma do LOD, é utilizar ontologias e tesouros, os quais devem ser atualizados com frequência e fortemente condicionados ao uso de padrões e linguagens específicas (PICKLER, 2007; KOUTSOMITROPOULOS; SOLOMOU, 2018).

Do ponto de vista da OC, a linguagem é colocada como um meio de transmissão de sentido que funciona como signo convencionado para representar um objeto ou fenômeno do mundo real (GARCÍA-GUTIÉRREZ, 2011). De acordo com García Gutiérrez (2011), ao evoluir biológica e culturalmente a humanidade aumentou sua capacidade de atribuir significados e inferir relacionamentos mais complexos aos fenômenos, modificando rapidamente sua capacidade conceitual pelo uso da linguagem natural.

Na OC, a manutenção dos tesouros está condicionada à aplicação de procedimentos voltados para o Controle Terminológico de um domínio do conhecimento, visto que a função do vocabulário controlado é minimizar os erros de tradução da língua natural (PINHO, 2014). A principal motivação para se construir um tesouro fundamenta-se na ideia de usar um vocabulário controlado para a indexação e busca. Entre as vantagens estão a “[...] normalização de conceitos de indexação, redução de ruído, identificação de termos de indexação com significado semântico claro e recuperação baseada em conceitos no lugar de palavras.” (BAEZA-YATES; RIBEIRO-NETO, 2013, p. 217).

Porém, o sujeito que traduz a linguagem natural para a linguagem controlada carrega uma formação social, isto é, um conjunto de valores que são transmitidos culturalmente e condicionados pelo seu espaço e tempo de atuação. Dessa forma, Cintra et al. (2002, p. 11) concordam que embora aplicados os princípios da “[...] racionalização de escolhas e de procedimentos para uma utilização eficaz do sistema de conceitos [...]” permanece o questionamento: como evitar desvios nessa tradução?

A linguagem permite que se expresse pensamentos, ideias e emoções. Porém, pelo fato de estar relacionada com aspectos cognitivos e altamente abstratos ela pode ser incompleta, ambígua ou limitada, dificultando a compatibilização entre a intenção do autor e os usuários do sistema de informação (NARUKAWA, 2011).

O tesouro – enquanto linguagem de controle – não está livre de interferências culturais e da subjetividade de quem o elabora e exige um trabalho quase contínuo de atualização que inclui operações de supressão, reagrupamento ou adição de novos termos. Esse procedimento deve ser permanente e torna os tesouros instrumentos dinâmicos “[...] capazes de incorporar os avanços do conhecimento e as modificações de significado de termos já existentes.” (CINTRA et al., 2002, p. 47).

Além da diversidade de metodologias para a construção de um tesouro, seu projeto de elaboração envolve necessariamente uma equipe multidisciplinar que inclui

[...] profissionais com o conhecimento do domínio a ser representado e profissionais com o conhecimento sobre como modelar um domínio, estes últimos geralmente oriundos da área de Ciência da Informação. Pode-se ainda acrescentar que a construção de um tesouro está sujeita a fatores semelhantes aos que ocorrem no desenvolvimento de um projeto de *software*, tais como: levantamento de requisitos, modelagem do problema, controle de versões e manutenção (CAMPOS et al., 2006, p. 72).

Nesse sentido, espera-se que o responsável por desenvolver recursos tecnológicos para a elaboração de tesouros considere aspectos operacionais e metodológicos e, ao mesmo tempo, crie ferramentas inteligentes para que todos os processos e atores cooperem de maneira integrada (CAMPOS et al., 2006). Naturalmente, essa elaboração se caracteriza como um processo iterativo e envolve especialistas de domínio, que apesar de serem orientados por diretrizes bem definidas não estão eximidos da arbitrariedade individual.

[...] o processo de desenvolvimento e manutenção do tesouro depende de uma equipe de especialistas de domínio (também denominados curadores de tesouro). Essas equipes podem consistir em dez ou mais pessoas que colaboram simultaneamente. Portanto, deve-se ter cuidado para garantir que as decisões sobre casos limítrofes ocorram em consenso e de acordo com as diretrizes pré-estabelecidas. Deve-se evitar que modificações feitas por um curador sejam desfeitas por outro sem comunicação e discussão (NOHAMA et al., 2012, p. 12, tradução nossa)<sup>4</sup>.

Nessa dinâmica, há especialistas em documentação em diversos domínios do conhecimento que produzem manualmente a estrutura de conceitos de um tesouro, porém a construção manual desse instrumento é demorada. Muitas vezes carregado de subjetividade, podendo comprometer sua estrutura e efetividade. Outrossim, podem haver controvérsias em relação às decisões limites, como: a pertinência de um termo para um domínio, a inclusão de termos compostos ou derivados, o uso de sinônimos, de ambiguidade e de termos adicionais que estão no limiar do domínio, entre outros (CAMPOS et al., 2006; NARUKAWA, 2011; NOHAMA et al., 2012).

---

<sup>4</sup> “Usually, the thesaurus development and maintenance process depends on a team of domain experts (also named thesaurus curators). Such teams may consist of ten or more people that collaborate simultaneously. Therefore, care must be taken to assure that the decisions on borderline cases occur in consensus and in accordance with pre-established guidelines. It must be avoided that modifications done by one curator are undone by another one without communication and discussion” (NOHAMA et al., 2012, p. 12).

Santos et al. (2010) reiteram que para manter os vocabulários controlados atualizados é importante que todos os atores sociais envolvidos trabalhem de forma colaborativa. Ademais, as políticas, procedimentos e tecnologias devem confluir para que o tesouro construído represente benefício para a RI. O aspecto colaborativo evidente, assim como a necessidade de atualização constante, faz com que sejam introduzidas na área de elaboração de tesouros recursos tecnológicos avançados a fim de tornar esses instrumentos mais efetivos para RI, especialmente em ambientes dinâmicos, como a *web*.

Com base na literatura (CAMPOS et al., 2006; LIMA et al., 2006; LI; SONG; PARK, 2009; SANTOS et al., 2010; SONG et al., 2011; NARUKAWA, 2011; NOHAMA et al., 2012; HANANDEH, 2013; PETRAKI; KAPETIS; YANNAKOUDAKIS, 2015; HARPRING, 2016; BERTI JUNIOR et al., 2017), pode-se verificar que a área necessita desenvolver metodologias para minimizar erros de tradução e concomitantemente agilizar a elaboração de tesouros. Esta pesquisa parte do pressuposto de que os profissionais envolvidos na elaboração de tesouros, como *Linked Data*, precisam trabalhar de forma colaborativa. Assim, deve-se enfatizar que representar um domínio do conhecimento e formalizá-lo na estrutura do tesouro envolve decisões limites sobre o uso da linguagem, sendo essa uma das etapas em que o caráter colaborativo pode ser mais evidente.

Concernente à entrega de conteúdo relevante na *web* de dados conectados, os serviços de informação têm incorporado padrões do LOD com o objetivo de melhor organizar e disponibilizar o conteúdo na Internet, dando ênfase ao aspecto semântico dos recursos digitais (BAUER; KALTENBÖCK, 2011). Como vimos, o LOD é um conceito que preconiza práticas cooperantes e integradas, e as metodologias colaborativas para o contexto dos dados conectados variam conforme as tecnologias e áreas de aplicação (ZARZOUR; AL-AYYOUB; JARARWEH, 2018).

Na área de Bibliotecas os projetos enfrentam algumas dificuldades, especialmente em virtude da ausência de estudos que elenquem as melhores práticas para publicar dados bibliográficos conectados na *web*.

Permanecem muitas perguntas não respondidas sobre onde e como as bibliotecas e outras instituições de patrimônio cultural publicarão dados vinculados. O ecossistema de dados vinculados emergentes depende de uma variedade de protocolos, especificações e sistemas que podem não ser familiares para a equipe de tecnologia da informação nas bibliotecas. Embora os dados vinculados permitam que as bibliotecas publiquem seus dados diretamente na Web a partir de seus próprios servidores locais, e alguns o farão, muitas bibliotecas não terão o conhecimento ou a capacidade de

implementar dados vinculados sem usar serviços hospedados por fornecedores (COYLE et al., 2017, p. 10, tradução nossa)<sup>5</sup>.

Na perspectiva do LOD, vocabulários controlados no formato SKOS especificam registros de metadados e enriquecimento de recursos de bibliotecas por meio de anotações semânticas, além de servir como base para o desenvolvimento de ontologias (ÇAĞDAŞ; STUBKJÆR, 2015). Por conseguinte, publicar dados conectados expressos em SKOS representa uma vantagem para o setor de bibliotecas, visto que os tesouros estão ligados aos recursos que eles também indexam. Além disso, a publicação em LOD oferece o benefício de um único ponto de acesso usando linguagens de consulta padrão amplamente implantadas em aplicativos de computação (CARACCILO et al., 2011).

Na pesquisa bibliográfica exploratória preliminar – realizada no âmbito desta pesquisa – constatou-se que inúmeras instituições internacionais publicam dados de tesouros usando as tecnologias do LOD, conectando-os a outros conjuntos de dados. Verificou-se, ainda, que o emprego do modelo SKOS conta com a iniciativa de poucas instituições brasileiras. Catarino (2014, p. 17) afirma que “[...] pouco se tem publicado no Brasil a respeito deste modelo de construção de esquemas de conceito [...]”.

De igual modo, Santos e Moreira (2018, p. 381) concordam que a produção científica sobre o SKOS “[...] ainda é timidamente explorada.” Não obstante, o tema demanda de estudos que especifiquem melhores práticas, no sentido de auxiliar os profissionais quanto à publicação de vocabulários controlados em SKOS, visto que uma investigação mais completa acerca das melhores práticas de gerenciamento continua sendo discutida pelo Grupo de Trabalho de Boas Práticas para os Dados na *Web*<sup>6</sup> do W3C (W3C, 2014d).

Assim, mudar para uma cultura de maior compartilhamento de dados implica olhar para além das necessidades institucionais particulares e pensar em modelos que atendam às demandas de comunidades que têm objetivos em comum. Nesse ponto, acredita-se que desenvolver um modelo colaborativo para a elaboração de tesouros fará com que as instituições dedicadas à Organização da Informação Musical (OIM) no Brasil ofereçam melhores condições de acesso e uso da informação no contexto do LOD, tendo em conta que o comportamento do usuário ao recuperar IM mudou devido ao uso diversificado da Internet.

---

<sup>5</sup> “*There remain many unanswered questions about where and how libraries and other cultural heritage institutions will publish linked data. The emerging linked data ecosystem relies on a range of protocols, specifications, and systems that may not be familiar to information technology staff in libraries. Although linked data allows libraries to publish their data directly to the web from their own local servers, and some will, many libraries will not have the expertise or capacity to implement linked data without using services hosted by vendors*” (COYLE et al., 2017, p. 10).

<sup>6</sup> Do inglês: *Data on the Web Best Practices Community Group*.

Para Allik et al. (2013, p. 178, tradução nossa)<sup>7</sup>, “Há uma crescente conscientização na comunidade de informática musical da necessidade de representações comuns usando tecnologias da Web Semântica e dados vinculados.” As próximas gerações de sistemas de RIM devem fornecer meios para que os pesquisadores localizem não apenas a obra, mas também o conjunto maior de informações contextuais ao seu respeito (GRACY; ZENG; SKIRVIN, 2013). Para o favorecimento da pesquisa contextual, os conjuntos de dados devem incorporar padrões da WS, possibilitando, assim, a vinculação de registros e recursos musicais na *web* (NURMIKKO-FULLER et al., 2016).

Quanto à **justificativa pessoal**, a possibilidade de contribuir com a área da OIM e RIM caracteriza uma oportunidade de aperfeiçoamento profissional, pois enquanto bibliotecário da Escola de Música da Universidade Federal do Rio Grande do Norte a referida temática tem sido objeto de estudo e deriva de experiências obtidas ao longo de oito anos de atuação no campo da organização e da representação da IM nessa instituição. Logo, observada a necessidade de atender a demandas do público diversificado acredita-se que o aspecto colaborativo é um requisito potencialmente útil para otimizar tanto a OIM quanto a RIM.

A evolução das tecnologias digitais, especialmente por intermédio da *web*, tem provocado mudanças no cenário global cujas informações são geradas e transmitidas rapidamente, modificando o cotidiano de pessoas e organizações (CASTELLS, 2016). Desse modo, as redes de conexões ampliam o desenvolvimento de SRI, que buscam atender demandas específicas de usuários de informação em diversos domínios de interesse (BAEZA-YATES; RIBEIRO-NETO, 2013).

Entende-se que as ferramentas colaborativas podem contribuir com a agilidade na manutenção de tesouros. Para isso, deve atender as demandas de construção, que se configura por meio dos fluxos de trabalho, padrões e modelos de dados, tarefas da rede de especialistas, profissionais da informação e usuários.

Na perspectiva do LOD, as práxis colaborativas englobam iniciativas de cooperação entre pessoas e instituições, bem como ferramentas tecnológicas capazes de promover a coordenação de atividades e a integração de equipes de trabalho que editam simultaneamente um conjunto de dados vinculados. A ausência de uma matriz, constructo ou modelo teórico que faça emergir os principais fundamentos da prática colaborativa para a elaboração de tesouros no formato do LOD é colocada como ponto principal desta pesquisa

---

<sup>7</sup> “*There is growing awareness in the music informatics community of the necessity of common representations using Semantic Web technologies and linked data*” (ALLIK et al., 2013, p. 178).

Assim, a hipótese lançada nesta tese é de que **as práticas colaborativas agilizam o processo de construção e publicação de tesouros do domínio da Música no contexto do LOD e contribuem com a OIM e a RIM na web.**

Para responder a esse questionamento central seria preciso compreender como ocorrem as práticas colaborativas no contexto de elaboração de vocabulários controlados do domínio da Música, identificar as etapas de construção, o fluxo de tarefas, os papéis de cada profissional envolvido e como ocorre a interação usuários-instituições. Do mesmo modo, deve-se analisar como as tecnologias do LOD estão sendo aplicadas nesse processo contributivo, já que a efetividade da produção pode ser influenciada pelo nível de integração de todos os atores envolvidos na manutenção e na edição dos dados conectados.

### 1.3 OBJETIVOS

Diante da contextualização, problemática e justificativas apresentadas seguem os objetivos geral e específicos da tese.

#### 1.3.1 Objetivo geral

Propor um modelo colaborativo para a construção e a publicação de tesouros no contexto do LOD, tendo como área de aplicação o domínio da Música.

#### 1.3.2 Objetivos específicos

Para o alcance do objetivo geral, foram definidos os seguintes objetivos específicos:

- a) Identificar práticas colaborativas na construção de vocabulários controlados do domínio da Música em unidades de informação especializadas no Brasil;
- b) Estruturar, com base na literatura, os princípios que norteiam o processo de construção colaborativa e publicação de tesouros como *Linked Open Data*;
- c) Representar, por meio de um modelo, os procedimentos e tecnologias voltadas para a construção colaborativa e publicação de tesouros do domínio da Música no contexto do *Linked Open Data*;
- d) Avaliar o modelo com especialistas da área.



## 1.4 ESTRUTURA DA TESE

A tese está estruturada em sete seções, além das referências e apêndices. A primeira seção apresenta a introdução, composta pela contextualização do tema, delimitação do problema e objeto de estudo, justificativas, objetivo geral e específicos e a descrição da estrutura da pesquisa.

A segunda seção anuncia os fundamentos teóricos e está dividida em quatro subseções que tratam, especificamente: os métodos e processos de construção de tesouros; os principais fundamentos e tecnologias do LOD; as tecnologias da WS para publicar tesouros no contexto do LOD; e as principais iniciativas de LOD no domínio da Música, com destaque dos principais projetos, desafios e oportunidades para os dados conectados nesse domínio.

A terceira seção aborda os procedimentos metodológicos definidos para o alcance dos objetivos da pesquisa. A quarta seção lança os resultados encontrados com a pesquisa de campo e bibliográfica, os quais buscaram, respectivamente, identificar práticas colaborativas na construção de vocabulários do domínio da Música e estruturar um conjunto de princípios que norteiam a construção e a publicação de tesouros como LOD. Na quinta seção é difundido a proposta de modelo. Na sexta seção são apresentados os resultados da avaliação do modelo com especialistas da área. E, por último, na sétima seção são expressas as considerações finais.

## 2 FUNDAMENTOS TEÓRICOS

Os fundamentos teóricos a seguir permitem conhecer o significado e o alcance dos principais temas abordados na pesquisa: a) a construção de tesouros; b) o conjunto de padrões, tecnologias e princípios que norteiam a publicação de dados como *Linked Data*; c) os tesouros na *web* representados a partir do padrão SKOS e seus modelos de extensões; d) os projetos, oportunidades e desafios do *Linked Data* para o domínio da Música.

### 2.1 A CONSTRUÇÃO DE TESAURUS

A palavra tesouro vem do latim “*thesaurus*”, significa tesouro. O termo originou-se na Grécia e está relacionado a armazenagem/ repositório. No princípio, esse termo era usado para designar dicionários especializados de uma determinada área do conhecimento e com o passar do tempo passou a ser utilizado para designar um tipo de linguagem documentária destinada à indexação de documentos. Acredita-se que o primeiro emprego da palavra tesouro data de 1531, com a publicação do *Thesaurus linguae latinae*, por Robert Estiene (SIMÕES, 2008).

Um dos trabalhos que introduziram a elaboração de tesouros foi a publicação de Peter Markok Roget, intitulada *Thesaurus of English Words and Phrases*. Na época Roget, até então secretário da *Royal Society*, teve a ideia de elaborar um instrumento que permitisse recuperar palavras a partir de seus conceitos a fim de facilitar a atividade literária. Assim, ele elaborou uma sequência de termos arranjados conforme as ideias que os termos/conceitos expressavam, ou seja, seus significados. Em 1936, o *Oxford English Dictionary* definiu a expressão inglesa *thesaurus* para nominar dicionário, enciclopédia (VICKERY, 1960; SIMÕES, 2008).

No contexto da RI, o termo em referência foi empregado pela primeira vez por Peter Luhn, em 1957. O tesouro, na concepção do que é hoje, tem sua evolução registrada na década de 1950, quando surgiram vários tipos de indexação de assunto, como o Sistema Uniterm, desenvolvido por Mortiner Taube. A ideia do Uniterm era utilizar um conjunto limitado de termos para a indexação de documentos. Desse modo, um tesouro tem um enfoque ligeiramente diferente na RI se comparado aos tesouros lingüísticos, pois está preocupado muito mais em organizar conceitos e seus relacionamentos ao invés de simplesmente os termos (VOSS, 2006).

Da mesma maneira, seu efetivo uso representou um grande avanço para o campo da RI e acelerou o desenvolvimento de tesouros documentários em diversas áreas do conhecimento (VICKERY, 1960; CINTRA et al., 2002; SIMÕES, 2008). O tesouro é um tipo de SOC em que os conceitos que cobrem um domínio do conhecimento são representados por termos

organizados de forma sistemática em três níveis de relacionamento: hierárquico, associativo e de equivalência. Eles podem ser monolíngues ou multilíngues e têm como principais objetivos promover a consistência na indexação de conteúdo e facilitar a pesquisa e a navegação em recursos de busca (ZENG, 2008; HARPRING, 2016).

Nesse sentido, o tesouro é um instrumento cujo os termos estão organizados de maneira lógico-semânticas. Seu processo de estruturação não se volta apenas para o controle do vocabulário, mas sim para a representação conceitual de um determinado domínio em observância ao significado dos termos. Portanto, é preciso esclarecer a diferença entre termos e conceitos e a importância desses dois elementos para a construção do tesouro.

### **2.1.1 O conceito como elemento central do tesouro**

O tesouro é considerado um tipo de linguagem artificial, pois não é fruto de um processo natural e evolutivo como a linguagem natural, sendo esta última designada por Dalhberg (1978, p. 101) como “línguas utilizadas nas necessidades da vida diária”. Diferente da linguagem natural, os tesouros apresentam regras para a sua estruturação e utilização – assumindo funções claras – qual seja: atender às solicitações de busca dos usuários de um sistema de informação.

Um vocabulário controlado, na sua forma mais simples, não se preocupa em diferenciar termos de palavras, visto que ele se volta apenas para o controle de polissemia ou expressa tão somente as relações existentes entre sinônimos ou quase sinônimos. Nos vocabulários controlados do tipo tesouro o conceito é visto como um elemento central.

Palavra é a menor unidade léxica, cujo significado se depreende do contexto em que ela figura mas que, tomada isoladamente, pode ter vários significados. O uso de palavras na indexação/recuperação é inadequado pela ambigüidade que elas carregam. Neste ponto começa a artificialidade da linguagem documentária: a partir de alguns princípios, escolhe-se uma determinada palavra ou expressão para representar um único conceito, ou idéia. Quando isto se dá, tem-se então, não mais uma "palavra" mas um "termo". Dito de outra maneira, o termo equivale a um conceito + uma designação. Esta designação pode ser constituída por uma ou mais unidades léxicas. O controle dos termos é, portanto, necessário para que a cada um deles não se atribua mais do que um conceito e, também, para que a cada conceito não se atribua mais de um termo (GOMES, 1990, p. 15).

À vista disso, o tesouro faz uso dos termos para representar os conceitos de um domínio, explicitando as relações estruturais e semânticas existentes entre esses elementos. Os termos são escolhidos apenas para nomear um conceito e suas demais formas variantes são

especificadas no tesouro como termos não preferenciais, os quais formam um conjunto de remissivas que conduzem aos termos autorizados (descritores).

A noção de conceito é discutida no campo da linguística e examinada também no campo da Terminologia, Biblioteconomia e Ciência da Informação, especialmente em função da elaboração de produtos terminológicos e documentários. No campo da Terminologia encontra-se a abordagem da Teoria Geral da Terminologia (TGT) desenvolvida por Eugen Wüster, e na Biblioteconomia e Ciência da Informação sofre influência da Teoria da Classificação Facetada de Ranganathan (1967) e da Teoria Analítica do Conceito, ou simplesmente Teoria do Conceito de Dahlberg (1978). Essas diversas abordagens podem contribuir com a construção de tesouros, porque possuem princípios comuns, constituídos por elementos que estão na base da formação dos SOC, que são os conceitos (CAMPOS, 1994).

A Teoria da Classificação Facetada pressupõe a classificação por facetas e foi elaborada na década de 1930 pelo Bibliotecário indiano Shiyali Ramamrita Ranganathan. Uma das bases dessa teoria é a análise de facetas, expressão usada por Ranganathan para se referir ao processo de fragmentar um assunto em diversos aspectos ou partes que representam as categorias fundamentais de um assunto: Personalidade, Matéria, Energia, Espaço, Tempo (RANGANATHAN, 1967).

Na abordagem analítico-sintética de Ranganathan um assunto complexo é analisado e fragmentado até que seja formado apenas por assuntos simples que, posteriormente, podem ser combinados para representar o assunto ou o conteúdo de um documento. Igualmente, os tesouros são elaborados com o propósito de facilitar a indexação e a recuperação de conteúdo em sistemas de informação. Em ambos os casos, a unidade a ser manipulada é o conceito, pois esse é considerado a unidade de conhecimento representado por facetas nos sistemas de classificação ou por termos nos tesouros (MACULAN; LIMA, 2017).

Nesse sentido, Tristão, Fachin e Alarcon (2004, p. 165) consideram que “o uso de categorias na organização de conceitos e, em consequência, na elaboração de uma classificação é um recurso para o entendimento da natureza do conceito e para a formação das estruturas conceituais. As categorias possibilitam a sistematização do conhecimento.”

Quanto à Teoria Analítica do Conceito, Dahlberg (1978) estabelece princípios para que os conceitos possam ser identificados e compreendidos e objetiva facilitar a recuperação de informações, tendo o conceito como elemento norteador. Para a autora, o conceito é uma unidade de conhecimento que expressa abstrações da realidade.

Com a evolução da linguagem natural o ser humano desenvolveu a capacidade de se relacionar com os diversos objetos do mundo real. Dahlberg (1978) considera que esses objetos

são considerados itens de referência sobre os quais podem ser realizadas afirmações ou enunciados verdadeiros representados por formas verbais. Por conseguinte, Dalhberg (1978, p. 102) afirma que o conceito é constituído de “[...] elementos que se articulam numa unidade estruturada”. Contudo, um conceito é elucidado a partir da compreensão de como esses elementos estão articulados em uma tríade: o referente (objeto); a forma verbal (termo/nome); e as características a eles atribuídas.

Essa abordagem pode contribuir com o entendimento do conceito como elemento central dos tesouros, tendo em vista que é a partir da definição do conceito que são estabelecidas as relações lógico-semânticas entre os diferentes termos que comporão tal instrumento. Por exemplo, sobre o conceito de “Música” pode-se afirmar: é uma forma de arte; constituída pela combinação de diversos sons e ritmos; prática cultural e humana. Esses enunciados, quando relacionados, ajudam a explicitar o significado do conceito “Música” e a revelar diversos outros objetos relacionados a ele.

A ideia de conceito e construção de linguagens artificiais também se apoia em estudos no campo da terminologia, a qual pode ser compreendida como uma disciplina, prática ou produto que é resultado da prática. Para Cabré (1995, p. 2, tradução nossa)<sup>8</sup>, a terminologia “[...] como disciplina se ocupa do estudo de termos especializados; como uma prática, é o conjunto de princípios utilizados na compilação de termos; como um produto, é o conjunto de termos de uma especialidade particular”.

No campo da terminologia, o conceito é tido como uma “unidade de pensamento, constituída por um conjunto de características atribuídas a um objeto ou uma classe de objetos, que pode ser representado por um termo ou um símbolo” (BOUTIN-QUESNEL et al., p.18, tradução nossa).<sup>9</sup>

Uma das abordagens de maior influência no campo da terminologia moderna é a Teoria Geral da Terminologia (TGT) desenvolvida por Wüster, que propôs um novo olhar acerca das questões metodológicas e normativas das terminologias especializadas. Uma das contribuições de Wüster foi buscar compreender a terminologia como “[...] uma ferramenta capaz de eliminar ambiguidades no âmbito das comunicações científicas e técnicas” (CERVANTES, 2009, p. 126).

<sup>8</sup> “*Es de sobras conocida la polisemia del término terminología, que nos remite por lo menos a tres nociones: a) a la disciplina; b) a la práctica; c) al producto generado por esa práctica. Como disciplina, es la materia que se ocupa de los términos especializados; como práctica, es el conjunto de principios encaminados a la recopilación de términos; como producto, es el conjunto de términos de una determinada especialidade*” (CABRÉ, 1995, p. 2).

<sup>9</sup> “*Unité de pensée constituée d’un ensemble de caractères attribués à un objet ou à une classe d’objets et qui peut s’exprimer par un terme ou par un symbole*” (BOUTIN-QUESNEL, 1985, p.18).

No que concerne a elaboração de tesouros, o controle do vocabulário busca eliminar as ambiguidades entre os termos. Assim, a abordagem de Wüster pode contribuir com a construção de tesouros pois considera que o conceito é formado pela comunidade que o utiliza, está diretamente relacionado ao contexto de uso e é representado por meio de um termo unívoco, ou seja, um determinado conceito só pode ser representado por um único termo, parte do princípio da monossignificação.

O princípio da monossignificação pode ser pouco útil quando se trata da representação e vinculação de vocabulários controlados na *web*, pois estes podem estar estruturados distintamente. No entanto, essa discussão se torna necessária para que o problema da ambiguidade seja mitigado ao menos numa perspectiva intra vocabulários.

Em linhas gerais, o conceito resulta de uma abstração do ser humano que utiliza da linguagem para se referir a fenômenos ou objetos do mundo real. Estes, por sua vez, são constituídos de um conjunto de elementos característicos que servem para identificar as qualidades desses objetos. Logo, o conceito é para o tesouro um elemento que norteia toda a estruturação, visto que a representação de um domínio do conhecimento está condicionada à compreensão e à articulação de conceitos expressos por termos e relacionados entre si conforme seus significados.

Desta maneira, a seleção de termos que irá compor o tesouro deve obedecer a critérios para que este possa representar um determinado domínio do conhecimento. Esses critérios estão ancorados nas garantias para a representação dos domínios apresentadas na sequência.

### **2.1.2 Garantias para a representação de conceitos ou seleção de termos**

De acordo com Svenonius (2001), a definição do domínio assegura as escolhas do vocabulário utilizado. Assim, são considerados alguns critérios na fase de seleção de termos para que os tesouros alcancem objetivos, entre os quais se destacam as garantias literária, estrutural e do usuário. Essas garantias são vistas como uma justificativa para a representação de um conceito ou para a seleção de um termo conforme três aspectos principais: a ocorrência dos termos na literatura; a necessidade de estruturação dos termos nas hierarquias; e a adequação às solicitações de informação por usuários de um sistema de recuperação de informações.

O termo garantia literária (do inglês “*Literary warrant*”) foi introduzido por Wyndlan Hulme, ao considerar que [...] um cabeçalho de classe é garantido apenas quando foi demonstrado que ele existe na literatura, e o teste da validade de um cabeçalho é o grau de

precisão com a qual ele descreve a área de assuntos comuns para a classe” (HULME, 1911, p. 46)<sup>10</sup>. Parte do pressuposto que a literatura da área vai determinar a seleção de termos da linguagem de indexação. Por conseguinte, a frequência com que os termos aparecem na literatura são indicativos da temática e se tornam candidatos à inclusão no vocabulário controlado.

Segundo Barité (2007), a garantia literária deve ser vista como um princípio orientado pelo consenso da comunidade científica sobre o uso da linguagem. Dessa maneira, são admitidos primordialmente nos tesouros ou demais linguagens documentárias termos registrados na literatura especializada da área, em razão de que eles são responsáveis por legitimar os novos conhecimentos gerados por essa comunidade.

No entanto, essa garantia não é suficiente para que a seleção de termos ocorra, pois nem sempre um termo presente na literatura é utilizado pelos usuários ao tentarem encontrar o que necessitam. Isto posto, parte-se do princípio de que os termos também devem ser selecionados conforme a frequência com que são aplicados na tarefa de recuperação da informação, atendendo assim ao critério da garantia do usuário.

Segundo a ANSI/NISO (2005, p. 167, tradução nossa)<sup>11</sup>, a garantia do usuário tem em conta os “[...] frequentes pedidos de informação sobre o conceito ou pesquisas de texto livre sobre o termo por usuários de um sistema de armazenamento e recuperação de informações”. Embora alguns destes utilizem termos livres nas tentativas de busca por assunto, os tesouros podem ser empregados como instrumentos que levam-nos a melhor direcionar suas solicitações de pesquisa, e para isso pode ser necessário acomodar o vocabulário por eles utilizados na estruturação do tesouro.

A busca pela compatibilização entre a linguagem encontrada na literatura e a usada pelo usuário é persistente. Os termos devem ser incluídos no vocabulário sem que sua estrutura funcional seja prejudicada. Por isso, pode haver a necessidade de selecionar termos que não se apoiam totalmente na garantia do usuário ou na garantia literária, mas sim em termos que apresentam uma função útil para a estruturação do tesouro. Desse modo, surge o critério da garantia estrutural, que encontra na estrutura das hierarquias e classes de assunto a justificativa

---

<sup>10</sup> “A class heading is warranted only when a literature in book form has been shown to exist, and the test of the validity of a heading is the degree of accuracy with which it describes the area of subject- matter common to the class” (HULME, 1911, p. 46).

<sup>11</sup> “Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of frequent requests for information on the concept or free-text searches on the term by users of an information storage and retrieval system” (ANSI/NISO, 2005, p. 167).

para a inclusão de um termo, podendo serem admitidos no vocabulário a fim de manter os elos entre as hierarquias ou facilitar a disposição de termos mais específicos (SVENONIUS, 2001).

Além das garantias já mencionadas, a literatura apresenta outros tipos de garantias, como a garantia educacional, a garantia cultural, a garantia documental, a garantia terminológica e a garantia organizacional (BEGHTOL, 1986, 1995; BARITÉ, 2000, 2007; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010).

A garantia educacional, segundo Beghtol (1986), está ancorada nos diferentes tipos de necessidades informacionais de indivíduos e estabelecidas em consonância com um conjunto de valores educacionais. Parte do pressuposto de que as necessidades de informação estão condicionadas às diferentes culturas. Se assemelha à Garantia Cultural à medida que considera os valores culturais como requisito para estabelecer qualquer tipo de representação do conhecimento, ou seja, qualquer sistema de conceitos deve considerar os valores dos indivíduos daquela cultura.

Parece senso comum considerar que indivíduos pertencentes a diferentes culturas têm diferentes necessidades de informação e diversas formas de interpretação do conhecimento estabelecido. Na verdade, grande parte da pesquisa aplicada tem referências locais, e há segmentos do conhecimento que são relevantes apenas em um espaço geográfico (e, portanto, cultural) e não em outro, seja em geologia de solos, na música ou usos e costumes sociais (BARITÉ, 2007, p. 4-5, tradução nossa)<sup>12</sup>.

A garantia cultural é um princípio considerado em domínios cujas referências regionais irão determinar como o conhecimento deve ser sistematizado nos vocabulários. Logo, pode haver variações linguísticas que implicam na representação de um determinado conceito e ser preciso estruturá-los conforme os usos locais.

Barité (2000) ainda introduz os princípios da garantia terminológica e da garantia documental. A garantia terminológica considera os termos utilizados por autores amplamente reconhecidos na comunidade científica, assumindo que eles são autoridades no assunto e por esta razão têm condições de legitimar a escolha dos termos mais representativos da área. Quanto à garantia documental discorre que ao invés de se basear apenas nos postulados teóricos do

---

<sup>12</sup> “Parece de sentido común considerar que individuos que pertenecen a diferentes culturas, cuentan con diferentes necesidades de información y formas diversas de interpretación del conocimiento establecido. De hecho, gran parte de la investigación aplicada tiene referencias locales, y hay segmentos de conocimiento que son relevantes sólo en un espacio geográfico (y por tanto, cultural) y no en otro, tanto si se habla de la geología de suelos como de la música o los usos y costumbres sociales” (BARITÉ, 2007, p. 4-5).



domínio apresentados na literatura a seleção de termos deve se apoiar nas especificidades e na terminologia tratada na documentação da área.

A garantia organizacional consiste na seleção de termos conforme as características e o contexto da organização que desenvolve e aplica o vocabulário controlado. Conforme Barité (2007), vocabulários controlados podem ser construídos para atender as necessidades de comunicação entre as pessoas de uma entidade, seja para manter a eficiência no trabalho, seja para o cumprimento de objetivos corporativos. Algumas instituições utilizam linguagens muito específicas, com termos próprios e aplicação em contextos particulares. Por esse motivo considera-se a necessidade de desenvolver vocabulários customizados que atendam aos propósitos de troca de informação entre os membros da organização.

As garantias funcionam para nortear a seleção de termos que representam o domínio ao qual a linguagem de indexação pretende representar. Não obstante, esses critérios devem ser aplicados consoante com opções metodológicas e etapas para a construção de tesouros, descritas tanto na literatura como em instruções normativas.

### **2.1.3 Métodos e etapas de construção**

Os métodos e as etapas para a construção de tesouros são especificadas em normas nacionais e internacionais, entre as quais se destacam: ISO 25964:2011; ANSI/NISO Z.39.19-2005; NF Z47-100; NP 4036-1992; UNE 50-106-90; BS 8723-2: 2005 (SIMÕES, 2008). Entre as normas citadas, as tradicionalmente utilizadas no Brasil são a ISO e ANSI/NISO (LIMA et al., 2006; GABRIEL JÚNIOR; LAIPELT, 2017).

De modo geral, a elaboração de tesouros envolve a coleta e a classificação hierárquica dos termos e emprega duas abordagens principais: estatística e linguística. Estas podem ser desenvolvidas por especialistas humanos, de forma manual ou com o auxílio de máquinas que ajudam a compilar e agrupar os termos candidatos por meio de procedimentos automáticos e semiautomáticos.

Os procedimentos automáticos aplicam métodos estatísticos para a computação de similaridades de palavras e a formação da estrutura hierárquica dos termos. Quanto ao procedimento semiautomático, especialistas humanos e técnicas computacionais são operadas conjuntamente, cabendo a esses sujeitos validarem uma disposição hierárquica de termos resultantes do processamento automático de documentos (HANANDEH, 2013; PETRAKI; KAPETIS; YANNAKOUDAKIS, 2015). Contudo, uma única abordagem não é suficiente para solucionar os problemas colocados pela construção de tesouros (IBEKWE-SANJUAN, 2006).

Os métodos automáticos se tornaram evidentes entre as décadas de 1970 e 1980, quando pesquisas buscaram otimizar tempo e custos de manutenção para elaborar vocabulários mais precisos e atualizados. Ao longo do tempo, diversas abordagens linguísticas e estatísticas foram desenvolvidas empregando também métodos semiautomáticos de construção (REES-POTTER, 1989; GREFENSTETTE, 1994; MORIN; JACQUEMIN, 2004; SCHNEIDER; BORLUND, 2004).

Em conformidade com Frakes e Baeza-Yates (1992), o método automático usa técnicas de Processamento de Linguagem Natural (PLN) que se baseia em modelos computacionais e estatísticas para o cálculo de similaridades entre as palavras de um *corpus* de texto. Essa técnica busca converter a linguagem natural, expressa nos documentos, em representações formais, compreensíveis por computador, e objetiva formar uma estrutura hierárquica de termos representativos com base na frequência que aparecem em uma coleção de referência ou *corpus* de amostra.

A atividade automática requer uma fase de pré-processamento, na qual diferentes técnicas de PLN podem ser aplicadas. Por conseguinte, devem ser executadas análises léxicas (normalização das estruturas textuais), eliminação de palavras desnecessárias ou sem sentido (*stopwords*), remoção de afixos (prefixos ou sufixos) e o agrupamento dos substantivos para posteriormente construir as categorias de termos.

Os modelos clássicos de RI consideram que cada documento é descrito por um conjunto de palavras representativas que fornecem uma visão lógica dos documentos quando essas representações computacionais são criadas por meio do PLN e podem ser utilizadas na construção de tesouros (BAEZA-YATES; RIBEIRO-NETO, 2013).

A maioria dos modelos ou métodos que empregam as técnicas de PLN tem como base a similaridade dos termos. Mediante o coeficiente de similaridade de termo é possível identificar termos frequentes em fragmentos de texto e analisar seus relacionamentos. Porém, a coocorrência do termo pode não ser uma fonte confiável de estimativa para termos de baixa frequência. Isto se infere à dispersão de dados, principal problema da construção automática (PARK; CHOI, 1996).

A origem dos métodos de abordagem estatística para a coleta de termos advém dos estudos no campo da RI com as versões dos índices IDF (*Inverse document frequency*), usados para selecionar termos coocorrentes em unidades de texto (frase, parágrafo, texto completo) (CROUCH; YANG, 1992; IBEKWE-SANJUAN, 2006). Posteriormente, diversos trabalhos no campo da linguística computacional começaram lidar com o cálculo de coocorrência para o agrupamento em classes de palavras (CHURCH; HILL; HANKS, 1990; SMADJA, 1993).

Dessa forma, ao longo do tempo inúmeros estudos têm aplicado diferentes modelos estatísticos e computacionais para coletar e classificar os termos, observando não só frequência e relacionamentos no nível sintático, mas também no nível semântico. O Quadro 2 mostra de forma resumida alguns modelos encontrados na literatura. É prudente afirmar que dois ou mais destes podem ser aplicados em conjunto, e a comprovada eficácia de suas aplicações está condicionada à realização de testes.

**Quadro 1** - Modelos computacionais e estatísticos aplicados na construção automática de tesauros

Modelos	Breve descrição	Autores
<i>Vector Space Model</i> (Modelo de Espaço Vetorial)	Modelo estatístico usado para medir a similaridade de termos. Nesse modelo, tanto os documentos quanto os termos são representados como vetores, e a similaridade entre um documento e um termo é dada pelo cosseno entre o vetor do termo e o vetor do documento.	(XU; YU, 2010); (SONG et al., 2011) (TSENG, 2002)
<i>Mutual Information</i> (Informação Mútua)	Modelo estatístico que estima a relação entre as palavras que compõem o texto para mediar sua similaridade por meio de normas de associação. A taxa de associação pode ser ampliada para fornecer estimativas robustas em outros idiomas.	(CHURCH; HILL; HANKS, 1990) (HAN et al., 2013)
<i>Collocation map</i> (mapa de disposição)	O modelo é uma instância específica da rede bayesiana para codificar relações de dependência de termos. O mapa de disposição é um gráfico acíclico em que cada nó corresponde a um termo, e a evidência estatística entre os termos é codificada em cada borda. Assim, o mapa de disposição é usado para decidir a similaridade entre grupos de termos.	(PARK; CHOI, 1996). (HAN; HAN; CHOI, 1992)
<i>Neural network</i> (Redes Neurais)	Modelo computacional inspirado no sistema nervoso central (cérebro) utilizado na inteligência artificial e aprendizado de máquina para reconhecer padrões. Aplicado a construção automática, o modelo é capaz de classificar documentos de texto (ou hipertexto) em conjuntos de categorias ou códigos de assuntos predefinidos.	(ZHANG; ZHOU, 2006; XU; YU, 2010)
<i>Fuzzy logic</i> (Lógica Fuzzy)	Modelo estatístico computacional empregado para classificação de texto que introduz o relacionamento semântico entre termos usando o conceito de grau de associação. Calcula a relevância difusa de cada termo nos contextos usando uma Função de Relevância de Contexto. Pode ser aplicado à construção de tesauros multilíngues para introduzir um grau de	(FUGINI, 1995; CHAU; YEH, 2001)

	relacionamento semântico das relações lexicais entre termos multilíngues.	
<i>Latent Semantic Indexing (LSI)</i>	Modelo computacional que utiliza decomposição de valor singular de uma matriz de coocorrência para extrair automaticamente os principais termos que contribuem para a indexação de documentos.	(DEERWESTER et al., 1990) (HAGIWARA; OGAWA; TOYAMA, 2005)

**Fonte:** Elaborado pelo autor (2021).

Apesar dos ganhos, a abordagem da coocorrência apresenta limitações porque algumas unidades de texto podem ser sintaticamente inválidas. Desse modo, Grefenstette (1994) afirma que os modelos estatísticos de coocorrência funcionam como um método para “associação de palavras de primeira ordem”, pois apresentam palavras isoladas e não revelam os relacionamentos entre as unidades de uma mesma classe, exigindo assim, conforme Lancaster (2004), a intervenção de humanos para que possam ser aplicadas na construção de tesouros.

A abordagem linguística, apesar de também fazer uso de modelos estatísticos, é fundamental em critérios linguísticos para a seleção e relação de termos. Classes de termos produzidos pela abordagem linguística podem ser qualificadas como “associações de segunda ordem”, pois levam em conta não apenas as semelhanças sintáticas das palavras, mas também suas funções gramaticais (IBEKWE-SANJUAN, 2006).

À vista disso, três abordagens linguísticas específicas são usadas para a construção de tesouros:

- a) Semelhança funcional (abordagem distributiva): parte da hipótese de que quando ocorre frequentemente similaridades sintáticas em unidades de texto estas formam algum tipo de “classes semânticas”. Logo, as unidades têm maior probabilidade de apresentarem algum tipo de relacionamento (HINDLE, 1990; GREFENSTETTE, 1994).
- b) Eviência sintática interna: baseada na função gramatical das palavras e em padrões léxico-sintáticos (RUGE, 1992; GREFENSTETTE, 1997; WOODS, 1997);
- c) Aquisição de relações semânticas baseada em marcadores de relacionamento semânticos e recursos semânticos externos: parte da hipótese subjacente de que as relações semânticas podem ser expressas mediante uma variedade de padrões lexicais e sintáticos evidentes em vocabulários consolidados como tesouros de domínio, ontologias e outros recursos semânticos (HEARST, 1992).

Embora as abordagens automáticas tenham surgido com a função de agilizar a construção de tesouros, elas não garantem a precisão nos níveis de relacionamentos conceituais. Geralmente, a técnica de PLN é aplicada para calcular a coocorrência de palavras e gerar o agrupamento de termos semelhantes. No entanto, essa técnica pode ser considerada limitada, principalmente se o domínio é representado por idiomas morfologicamente complexos que exigem alto nível de desambiguação de termos.

Para além das abordagens estatísticas e linguísticas, outras abordagens são explicitadas nas normas internacionais. A ANI/NISO Z39.19-2005 (R2010), por exemplo, categoriza os métodos de construção em quatro tipos de abordagens: abordagens de Comitê; abordagem empírica; combinação de métodos; e assistência de máquinas.

Na abordagem de comitê, os especialistas no domínio elaboram uma lista dos principais termos da área para posteriormente indicar as relações lógico-semânticas existentes entre eles. Esse método é realizado com o auxílio de especialistas do domínio experientes na construção de vocabulários controlados. As listas de termos podem ser obtidas a partir da análise de diversas fontes de referência ou enviadas pela comunidade de usuários do sistema.

A abordagem empírica considera os métodos dedutivos e indutivos para compilar termos. Essa abordagem conta com a participação de especialistas do domínio para selecionar os termos representativos que são extraídos da literatura no decurso da indexação, enquanto que no método indutivo a coleta de termos é realizada mediante a análise da literatura da área.

Dahlberg (1978) considera ainda a possibilidade de aplicar um terceiro tipo de método, denominado método relacional ou método analítico. Esse método consiste na análise da definição e da estrutura dos conceitos que integram o sistema de conceitos do domínio, os quais são analisados sob o ponto de vista da definição dos conceitos e dos relacionamentos que possuem entre si. Há, ainda, a possibilidade de aplicar um método híbrido ou combinação de métodos, considerando a possibilidade de usar as abordagens indutiva, dedutiva e relacional conjuntamente.

A assistência de máquinas ou de sistemas de computador durante a construção de tesouros pode ser utilizada para identificar termos candidatos, registrar ou computar a frequência com que estes aparecem no *corpus* de amostra da coleção ou, ainda, analisar e extrair termos utilizados por usuários durante a busca. Assim, abordagens automáticas e semiautomáticas devem ser consideradas a fim de auxiliar o processo intelectual dos profissionais envolvidos.

Sendo o tesouro o instrumento empregado tanto na indexação como na recuperação de informação, as equipes de gestão precisam analisar a participação da comunidade de usuários e

indexadores. Para isso, podem ser desenvolvidas diversas abordagens que incluam a participação efetiva desses atores sociais como, por exemplo, a implementação de comitês de usuários.

Os tesouros devem ser pensados como recursos que facilitam o acesso à informação em diversos ambientes e por diversos públicos. Com a universalização desse acesso, especialmente em razão do desenvolvimento de ambientes informacionais digitais, as normas para a construção de tesouros se aprimoraram ao longo do tempo para melhor atender às demandas de representação na *web*, interoperabilidade com outros vocabulários e representação de termos e conceitos em diversos idiomas. Este último parece ser o maior desafio nos tempos atuais, pois alcançar o princípio da representação multilíngue no nível do conceito é essencialmente complexo.

Quanto aos tesouros multilíngues, a norma ISO 25964-1 (2011) orienta que podem ser desenvolvidos de três maneiras: a) tradução de um tesouro monolíngue já existente; b) a combinação de vários tesouros monolíngues distintos já existentes; c) a construção simultânea das várias versões linguísticas do tesouro multilíngue. Os projetos devem considerar o contexto de aplicação, especialmente a existência de tesouros monolíngues que cobrem o domínio, sua utilização na indexação de documentos da coleção, o tempo para a construção e os recursos financeiros e humanos disponíveis.

Com base nas normas supracitadas, os principais procedimentos para a construção de tesouros foram sumarizados e elencados no Quadro 2:

**Quadro 2** - Etapas para construção e manutenção de tesouros

Etapas	Objetivos
Planejamento	Fase preliminar em que são estabelecidas as orientações gerais. Nessa etapa são definidos os objetivos do tesouro, escopo, público-alvo, métodos de construção, coleção da amostra e uso de recursos disponíveis (humano, financeiro e materiais).
Compilação de termos	Reunir os termos candidatos à inclusão no tesouro a partir de um <i>corpus</i> (coleção) de documentos.
Processamento automático e semiautomático de documentos	Utilizar máquinas, técnicas e sistemas para o pré-processamento de documentos tentando normalizar textos e identificar termos potenciais em um <i>corpus</i> .
Registro dos termos	Manter o registro individual de cada termo admitido no tesouro, especificando a fonte, o nome de todas as autoridades consultadas, a data de inclusão e referências a termos sinônimos, genéricos, específicos e relacionados.
Verificação dos termos	Verificar se o termo é correto antes de admiti-lo no tesouro, bem como revisar suas relações com outros termos da mesma hierarquia. Devem ser examinadas fontes de referência como: tesouros existentes, dicionários e

	enciclopédias especializadas, especialistas de domínio e outras fontes auxiliares.
Especificidade dos termos	Selecionar termos altamente específicos restritos ao domínio principal do tesouro, porque o uso excessivo de termos de outras áreas pode tornar o tesouro difícil de manusear.
Admissão, supressão e validação dos termos	Incluir termos que ainda não são utilizados na indexação e excluir os muito abrangentes, sem conteúdo semântico bem definido, bem como os que não são muito frequentes, visto eles serem considerados ineficazes na recuperação da informação. Validar as alterações realizadas pela equipe nas etapas anteriores.
Decisões editoriais	Decidir sobre os padrões de metadados, tecnologias utilizadas, armazenamento e a forma de apresentação do conteúdo do tesouro: diagramação, interfaces gráficas, árvore de conceitos e outros recursos.
Testes, avaliação e validação	Realizar testes com o usuário objetivando observar se os termos são apropriados para realizar buscas nos sistemas de recuperação da informação, ou seja, se os termos se adequam às expectativas dos usuários. Verificar a consistência estrutural e as relações lógicas entre os conceitos do tesouro.

**Fonte:** Elaborado pelo autor (2021) a partir de (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010; INTERNATIONAL STANDARD ORGANIZATION, 2011; 2013).

As etapas elencadas acima podem ser executadas de maneira automática, manual ou semiautomática. Mas, para além da total automatização de procedimentos as abordagens linguísticas por via de métodos semiautomáticos ganham espaço diante da necessidade de confluir os modelos baseados no cálculo de coocorrência de palavras e os modelos baseados nas análises lexicais e semânticas de componentes textuais.

A escolha entre as opções de abordagem dependerá de recursos materiais, financeiros e humanos disponíveis. Para uma melhor compreensão de como são aplicadas as etapas e os métodos de construção com vistas a atender aos propósitos desta pesquisa, serão apresentadas na sequência algumas abordagens colaborativas encontradas na literatura que põem em destaque a participação de diferentes atores sociais envolvidos na construção.

#### 2.1.4 Abordagens colaborativas na construção de tesouros

Essencialmente, os tesouros são construídos por muitas mãos. Desse modo, para que essa ferramenta possa representar um determinado domínio do conhecimento as equipes de gestão devem não apenas seguir os métodos e as etapas apresentadas nas normas e na literatura, mas também fazer com que os atores sociais envolvidos colaborem sistematicamente para a sua construção. Assim, as abordagens que põem em realce as práticas colaborativas durante a

construção dos tesouros são: a) a construção automática, semiautomática e a interoperabilidade com outros vocabulários; b) a cooperação com organizações e o uso de vocabulários existentes; c) as redes de profissionais e especialistas do domínio; d) a participação da comunidade de usuários.

### **a) Construção automática, semiautomática e a interoperabilidade com outros vocabulários**

De modo geral, a decisão sobre incluir ou não um conceito no tesouro é do grupo de gestão auxiliado por especialistas de domínio. Em conformidade com Granada, Vieira e Lima (2012) a coleta semiautomática de termos candidatos à inclusão inicia com a escolha de um conjunto de palavras representativas derivadas da aplicação de técnicas automáticas em fragmentos textuais para que posteriormente sejam analisadas por especialistas. Essa análise pode ser realizada por meio de uma interface de avaliação.

A interface de avaliação é um método característico da abordagem semiautomática, permite que os especialistas visualizem a semelhança dos termos e classifique-os conforme outros termos semelhantes encontrados. A interface ajuda a entender o contexto do termo pois a qualquer momento o avaliador pode visualizar onde ele aparece nos fragmentos textuais do *corpus* e verificar sua pertinência. Outras ferramentas semelhantes são apresentadas na literatura e possuem o mesmo objetivo: auxiliar de maneira rápida e eficiente na seleção de termos para a inclusão no tesouro.

Cheung et al. (2011) apresenta um mecanismo semiautomático usado para extrair conceitos de sites de tesouros públicos. A ferramenta permite que as atualizações realizadas nos tesouros públicos sejam levadas ao conhecimento de gestores do tesouro local, que podem decidir sobre a atualização do vocabulário. Com isso, os gestores podem acompanhar periodicamente a evolução da terminologia com base em tesouros existentes. No mesmo sentido, Moreira e Moura (2006) desenvolveram uma ferramenta semiautomática que aplica técnicas de PLN para construir e gerenciar um tesouro do domínio da Ciência da Informação com base em tesouros existentes, cujos termos são extraídos a partir de documentos impressos ou coletados diretamente da *web*.

Com o crescente volume de informação em formato digital se torna difícil aplicar abordagens baseadas em texto para criar tesouros na *web*. Desse modo, é necessário projetar técnicas escalonáveis para estimar associação de termos em grandes quantidades de dados.



Logo, as abordagens devem garantir que os relacionamentos semânticos extraídos desses *corpora* de base *web* sejam relevantes, precisos e contemplem ao máximo o domínio que se pretende representar.

Nessa oportunidade, estudos têm se aproveitado das técnicas de mineração de dados, de texto e de documentos na *web* para agilizar a construção de tesouros em cenários de grandes volumes de informação. No entanto, é preciso esclarecer a diferença básica entre os tipos de mineração. A mineração de dados é uma técnica que explora dados estruturados para identificar padrões consistentes. Já mineração de texto identifica padrões em dados não estruturados, escritos em linguagem natural. Quanto à mineração *Web* usa técnicas de mineração de dados para encontrar padrões em documentos ou recursos na *web*. Esse último tipo pode ainda se desmembrar em outros três de abordagem subjacentes: mineração da estrutura, mineração de uso e mineração de conteúdo (SCOZ, 2014).

Como exemplo, Ohishi et al. (2006), Yin, Zhu e Chen (2013) e Nakayama, Hara e Nishio (2006; 2007) utilizaram em seus estudos corpus enciclopédicos de base *web* para extrair relações semânticas particulares entre termos e suas descrições a partir de técnicas de processamento automático. Nas pesquisas, a técnica de mineração *Web* é colocada como principal meio para superar as dificuldades advindas do grande volume de dados.

Com base na descrição e localidade dos *hiperlinks*, Yin, Zhu e Chen (2013) proporam um método de construção de tesouro baseado no agrupamento de estrutura de *links* do Wikipédia. Para os autores, as plataformas Wiki se mostram vantajosas para os propósitos de construção de tesouros em razão do número massivo de páginas e as frequentes atualizações sucedidas a partir da colaboração dos usuários na *web*. O conteúdo desses *corpora* abrange temas tradicionais ou raros e termos mais recentes, colocando à prova a qualidade de alguns documentos. Para tentar sanar essa dificuldade, técnicas de processamento automático também são operadas para excluir expressões de baixa qualidade.

A partir do conhecimento coletivo disponibilizado, sua ampla rede de conceito e variada cobertura de tópicos a Wikipedia pode ser usada como um *corpus* de extração de conhecimento. Nela, os conteúdos de livre acesso são utilizados como coleção de referência e submetidos a um mecanismo de mineração *Web* que busca codificar relações semânticas explícitas e implícitas em uma grande escala de conceitos e identificar termos representativos em suas páginas (YI; ZHU; CHEN, 2013; NAKAYAMA; HARA; NISHIO, 2006; 2007).

Nesse mesmo contexto, a *Biblioteca Nazionale Centrale di Firenze*, por meio de um acordo colaborativo com a *Wikimedia* Itália, buscou associar termos do *Nuovo Soggettario Thesaurus* e vincular a conceitos equivalentes da Wikipedia. Nessa oportunidade, *links*

recíprocos podem ser implementados, permitindo a vinculação entre os dois conjuntos de dados. Nessa experiência, as conexões da Wikipedia como uma ferramenta multidisciplinar de indexação de assuntos geraram o primeiro protótipo de interoperabilidade já produzido entre bibliotecas e enciclopédias eletrônicas livres.

A interoperabilidade entre o *Nuovo Soggettario Thesaurus* e a Wikipedia continua sendo uma atividade constantemente sujeita à análise e à revisão intelectual. Após uma fase inicial de associação de equivalentes conceituais mediante procedimentos manuais, os procedimentos de associação automática também foram utilizados para verificar as relações de equivalência exatas dos termos (VITI, 2018).

Os relacionamentos entre os termos de diferentes vocabulários ganham força com as tecnologias da WS, em especial do *Linked Data*. Esse é o principal requisito para promover a interoperabilidade entre os vocabulários controlados e outros conjuntos de dados, inclusive dos tesouros, uma vez que conectados e representados em modelos de dados facilitam a leitura e o processamento por sistemas de computador, podem permitir que os dados sejam transportados de um sistema para outro e, além disso, favorecem a busca e a recuperação da informação em contextos mais amplos.

Nesse contexto, a construção de tesouros na *web* encontra nos métodos escalonáveis uma oportunidade para mitigar o problema de tempo, custos e necessidade periódica de atualização. Ainda, alcança na interoperabilidade a oportunidade de integrar seus vocabulários a outros conjuntos de dados, ajudando a compor uma grande rede de conceitos conectados e disponíveis para que demais aplicações na *web* também possam reutilizá-los.

Os exemplos acima revelam que pode ser necessário aplicar abordagens semiautomáticas, nas quais especialistas humanos avaliam e selecionam os termos recolhidos por ferramentas automáticas. A maior evidência de colaboração se dá entre os profissionais da informação e especialistas de domínio que são convidados a analisar os resultados dos agrupamentos de termos e sinalizar para o sistema se as inferências automáticas são pertinentes.

## **b) Cooperação com organizações e uso de vocabulários existentes**

Conforme já abordado, elaborar tesouros é uma atividade complexa, demanda tempo e exige constante atualização. Além disso, pode ser considerada uma praxe cara, pois necessita de esforço humano e ferramentas eficientes de manutenção. Isto posto, a cooperação com

organizações e o uso de vocabulários existentes na compilação e termos se torna uma alternativa positiva, no sentido de minimizar seus esforços de construção.

Segundo Wright, Harrison e Watkins (2015) reutilizar um tesouro existente, adequado e que já possua uma comunidade de usuários se torna vantajoso, visto evitar a proliferação de diversos vocabulários cobrindo um único domínio. Nessa perspectiva, Sholikahet al. (2017) acreditam que utilizar diferentes fontes, incluindo *corpus* multilíngues, para compilar termos e construir tesouros pode melhorar a cobertura do vocabulário. No estudo em alusão, os autores consideraram a cooperação com instituições que já possuem tesouros para extrair termos candidatos em vocabulários existentes, sejam eles mantidos por serviços de informação, centros de documentação ou agências internacionais.

Nos casos acima em que os léxicos de livre acesso são reutilizados, entidades podem colaborar com a construção do tesouro sem que nenhuma parceria formal esteja firmada, logo esse processo pode se caracterizar como uma colaboração não-formal ou indireta. Quando formalizada, alguns estudos mostram que a colaboração ocorre entre redes de especialistas ou organizações de uma mesma área de domínio.

Como exemplo aponta-se o *STW Thesaurus for Economics*, produzido originalmente em cooperação entre os quatro principais fornecedores públicos e privados alemães de serviços de informações comerciais e econômicas. O tesouro é atualizado regularmente e elaborado por um grupo de especialistas do domínio da *German National Library of Economics* e do *Leibniz Information Centre for Economics* que seguem o uso da terminologia internacional no campo das Ciências Econômicas.

Parcialmente financiado pelo *German Federal Ministry of Economics*, o tesouro foi desenvolvido objetivando fornecer acesso uniforme ao conteúdo por ela publicado. Nesse projeto, uma equipe de desenvolvedores e engenheiros de *software* trabalham continuamente em novas aplicações para que o tesouro se apresente como uma ferramenta terminológica integrada à serviços de informação nas instituições cooperantes (KEMPF; NEUBERT, 2016).

Uma outra evidência da prática colaborativa formal foi encontrada na pesquisa de Qi, Luo e Zhu (2010), que discutem a elaboração de uma plataforma para a gestão e compartilhamento do *Chinese Science and Technology Vocabulary System* (CSTVS). É apresentado o quadro técnico dedicado à criação, gerenciamento, manutenção e validação do vocabulário científico chinês mediante uma rede de especialistas.

No que se refere à coleta e ao processamento de termos, o vocabulário chinês é apoiado por três aspectos: cooperação com instituições que já possuem tesouros; compilação de termos

a partir de agências de notícias e periódicos especializados; e descoberta de novos termos a partir de serviços documentais (jornais, periódicos da área, enciclopédias e tesouros existentes).

Com o crescimento das práticas do LOD, uma forma alternativa de compilar termos é utilizar conjuntos de dados mantidos e disponibilizados de forma aberta por instituições. Há na atualidade diversos projetos que aplicam tecnologias e ferramentas da WS para estruturar os tesouros e disponibilizá-los como dados abertos conectados, como demonstrado em (CARACCILO et al., 2011, 2013; DEBEVERE et al., 2011; LAPORTE; MOUGENOT; GARNIER, 2012; BACA; GILL, 2014; CHARLES; FREIRE; ISAAC, 2014; ABECKER et al., 2015; ÇAĞDAŞ; STUBKJÆR, 2015; BINDING; TUDHOPE, 2016; COBB, 2016). Esses projetos servem como fontes de referência para as instituições que desejam construir ou atualizar seus vocabulários a partir de um conjunto de dados consolidados e adequados com as práticas do LOD.

### **c) Redes de profissionais e especialistas de domínio**

O controle terminológico nos vocabulários controlados se configura, na maioria dos casos, por meio de redes de profissionais e especialistas de domínio que atuam de forma integrada. A atualização desses produtos terminológicos é uma tarefa complexa, fruto de um trabalho sistemático que envolve uma equipe multidisciplinar com vasta experiência na área de especialidade (CAMPOS et al., 2006; LIMA et al., 2006; FUJITA, 2012; MACULAN et al., 2015).

Os grupos são formados por pesquisadores, profissionais da informação, linguistas e até usuários da comunidade. O desafio para essas comunidades está em como promover a cooperação dos membros da rede de especialistas por meio de tecnologias digitais que garantam agilidade na elaboração, controle e validação da estrutura conceitual dos vocabulários (CAMPOS et al., 2006; LIMA et al., 2006).

De modo geral, os procedimentos são apoiados por ferramentas colaborativas construídas para integrar as etapas em um fluxo de trabalho em que administradores (grupo editorial) e especialistas de domínio registram e validam os termos compilados. A equipe de gerenciamento é responsável pela auditoria na etapa de inclusão ou exclusão de um termo, e por essa razão alguns estudos apresentam soluções práticas para sanar problemas de manutenção dos tesouros, pois entendem que a interação entre as equipes de especialistas do domínio é um componente fundamental para se obter decisões consensuais.

Algumas abordagens são discutidas a fim de sanar o que Nohama et al. (2012) chamam de erros dinâmicos, ocasionados pela falta de transparência e comunicação ineficientes entre os especialistas humanos. O método utilizado no estudo é a análise de *logs* de usuários, que busca fazer um levantamento de todas as alterações realizadas pelos especialistas no sistema de gestão de tesouros objetivando identificar anomalias na construção. Posteriormente, esses dados são justapostos em um fórum de discussão que reúne todos os colaboradores, tornando possível o compartilhamento dos casos problemáticos e permitindo que as equipes editoriais tomem decisões para que os erros sejam corrigidos e evitados na posteridade.

No mesmo sentido, McCulloch (2005) destaca a importância de consultar uma equipe de indivíduos para aconselhar sobre questões estruturais – concernente às relações semânticas entre os termos – e para fornecer *feedback* sobre versões preliminares do tesouro. Idealmente, esse grupo deve incluir especialistas do domínio e especialistas em tesouros, de preferência com formação em Ciência da Informação. Esse retorno pode ser requerido mediante a realização de reuniões presenciais para a discussão em grupo. Com o propósito de limitar uma discussão extensa, recursos tecnológicos podem ser utilizados, como *e-mail* ou fórum de discussão *online*.

É provável que as opiniões quanto à inclusão ou à exclusão de termos não seja um consenso. Nesse sentido, o autor citado enfatiza que pode ser prático atribuir áreas específicas do tesouro a subgrupos de assessores para limitar a deliberação, com uma única pessoa responsável pelo controle editorial geral. Onde houver diferenças de opinião, o editor deve tomar a decisão final sobre questões como quais termos incluir, qual formato adotar e onde colocá-los na estrutura do tesouro.

Çağdaş e Stubkjær (2015) apresentaram uma proposta baseada no método de comitê aplicado na construção do tesouro do *Cadastre and Land Administration* da União Europeia. No método de comitê, os especialistas do domínio elaboram uma lista dos principais termos candidatos e indicam as relações entre eles, o que permite se aproximar o máximo possível do consenso sobre a relevância de um termo candidato à inclusão no tesouro. No caso os termos candidatos derivam do modelo de dados de informações geográficas, preparado a partir do comitê de padronização editorial e dos relatórios do projeto *Property Formation*, bem como de tesouros existentes, agências cadastrais e departamentos universitários ligados ao projeto.

No cenário dos dados conectados, a contribuição de colaboradores de diversas regiões tem consequências importantes para a manutenção do vocabulário. Kempf e Neubert (2016) relatam que antes da publicação na *web* todos os usuários do *STW Thesaurus for Economics* eram bem conhecidos. Agora, em virtude de sua disponibilidade aberta, o tesouro pode estar em uso em locais sobre os quais a instituição promotora, editores do tesouro, não têm

conhecimento algum. Nessa oportunidade, para manter os colaboradores informados acerca das novas versões ou outras alterações e para ter uma impressão de quem está usando o tesauro regularmente foi estabelecido uma lista de discussão para anúncios de novas versões e outras modificações.

As listas de discussões constroem uma comunidade de especialistas e colaboradores que compartilham da mesma ferramenta terminológica, podendo ser um canal de troca de conhecimentos entre pessoas ou instituições que não têm um vínculo formal, mas utilizam-na. Através desses canais de comunicação as instituições que usam o *STW Thesaurus for Economics* para indexação ou como ferramenta de referência, por exemplo, podem sugerir novos descritores ou traduções alternativas para os rótulos conceituais. Desse modo, pelo fato de estar aberto ao uso público as listas de discussões elevaram a um maior nível de participação no desenvolvimento colaborativo do tesauro.

No mesmo sentido de cooperação, ao descrever a publicação dos vocabulários Getty em LOD, Cobb (2016) destaca a importância de empregar um conjunto de ferramentas que promova a colaboração assíncrona entre os especialistas. No projeto foram desenvolvidas plataformas *Wiki* para a troca de conhecimento, um aplicativo de rastreamento de problemas e um aplicativo para acessar o repositório *Git* e controlar os códigos-fonte relacionados aos conjuntos de dados e reuniões virtuais. Em um esforço contínuo para envolver ainda mais a comunidade estabeleceu-se um fórum público de discussão, esperando que esta o use para questionar, discutir problemas e compartilhar soluções a respeito de aspectos técnicos dos vocabulários.

Os vocabulários são recursos compilados que crescem principalmente mediante contribuições da comunidade de usuários especializados, incluindo projetos de tradução internacional em grande escala (HARPRING, 2018). Desse modo, Plini et al. (2010) acreditam que as plataformas devem garantir à interoperabilidade entre o tesauro e outros sistemas de indexação com a ideia de cooperar para desenvolver um sistema abrangente de organização do conhecimento que possa ser definido como integrado, aberto, multifuncional e multilíngue.

#### **d) Participação da comunidade de usuários**

O objetivo do tesauro é garantir que os usuários possam recuperar de maneira rápida e precisa as informações que necessitam. Portanto, o comportamento de busca desses sujeitos pode ser considerado um componente de análise. Sua participação pode ser importante não só

para a avaliação da linguagem, mas também para a coleta de termos novos. Essa colaboração pode se caracterizar como não-formal ou indireta, porque a comunidade de usuários pode contribuir com a manutenção do tesauro sem a devida intenção.

A fase de testes e a avaliação também se configura como um processo colaborativo no qual os usuários fazem parte. Nela podem ser realizados testes para verificar a consistência dos tesouros a partir do comportamento de busca desses indivíduos, os quais podem ser aplicados em duas etapas: na primeira os profissionais indexam as coleções nas bases de dados utilizando os termos do tesauro construído; e na segunda os usuários são convidados a realizar sucessivas buscas no sistema. Nesse aspecto, os utilizadores são capazes de fornecer *feedbacks* (implícitos ou explícitos) significativos para a equipe de implementação, contribuindo com a atualização permanente da estrutura do tesauro (LAWRENCE et al., 2006).

Nesse enfoque, Lawrence et al. (2006) destacam a importância de criar fóruns de discussão na *web* para receber comentários de profissionais e do público. A análise de *logs* e painéis de especialistas são abordagens indicadas na avaliação. Dessa maneira, realizar reuniões presenciais com os membros do painel de especialistas, por *email* e por fórum da *web* pode enriquecer os debates acerca da inclusão de termos e suceder no consenso entre os pares.

A colaboração não-formal ou indireta de usuários e especialistas é evidenciada nos trabalhos de (CHEUNG et al., 2011; LOSEE, 2007; YOON, 2009; LAPORTE; MOUGENOT; GARNIER, 2012; PU; CHIEN, 2004; MOURA, 2009; LAPORTE et al., 2014). A interação com recursos de busca na *web* é aplicada para viabilizar a atualização dos tesouros, cuja abordagem é útil para preencher a lacuna existente entre os conceitos apresentados pelos autores e os termos conceituais apresentados por um usuário/pesquisador.

O uso de métodos alternativos considera que os usuários podem auxiliar tanto na coleta de novos termos como na avaliação e teste dos termos já registrados no vocabulário (NORUZI, 2018). Esse tipo de colaboração é percebido nas abordagens semiautomáticas de construção majoritariamente por meio de quatro ações: análise de *logs*, marcações sociais (*taggings*), método de associação de palavras e expansão de consultas.

Em se tratando das marcações sociais, alguns sistemas de recuperação permitem que o usuário atribua uma etiqueta (marcação semântica) para representar, por meio de uma expressão livre, o conteúdo de uma página ou dos resultados de busca. A maneira como os temas são categorizados pode ser analisada pela equipe editorial, que julga a possibilidade de incluir novos termos no tesauro (MOURA, 2009; NORUZI, 2018).

Para Noruzi (2018), como as anotações sociais são fortemente condicionadas ao comportamento de busca pode haver diferenças entre os termos mais comumente usados na

literatura e os empregados por usuários. Portanto, a garantia literária continua sendo o princípio orientador para a seleção dos termos preferenciais, e os usuários devem ser consultados para garantir que os termos refletem o uso comum e, também, para determinar termos não-preferidos.

No método de análise de *logs*, as palavras-chaves utilizadas para recuperar conteúdos nos sistemas de busca são avaliadas, visto que algumas expressões podem divergir dos termos controlados usados na etapa de indexação. Essa abordagem é discutida nos estudos (PU; CHIEN, 2004; LEE; HUANG; CHAO, 2007; ZHANG; WOLFRAM, 2009; MORSHED et al., 2010; LAIPELT, 2015; NOHAMA et al., 2012) e consiste em extrair termos relevantes a partir das sessões de consultas dos usuários para posteriormente agrupá-los em classes conceituais.

Uma vantagem da abordagem de *log* de transações para o estudo do comportamento de busca de informações é que um arquivo de *log* da *web* registra com precisão todas as atividades de usuários online realizadas em um servidor. Diferentemente das abordagens de entrevista para coletar dados do usuário, os *logs* de transações podem acumular grandes quantidades de dados para análise. Os termos frequentes dos registros que não aparecem no tesouro podem justificar a inclusão de novos termos (ZHANG; WOLFRAM, 2009).

Utilizando o método de análise de *logs* de usuários, Nohama et al. (2012) propõem solucionar debilidades identificadas na comunicação entre os membros de uma equipe de especialistas que podem ocasionar em problemas de consistência terminológica. O método apontado faz um levantamento de todas as alterações realizadas pelos especialistas no sistema de gestão de tesouros a fim de identificar erros. De posse desses dados, os casos problemáticos são compartilhados em fórum de discussão que reúne todos os colaboradores, possibilitando que decisões sejam tomadas em conjunto para que as falhas sejam corrigidas e evitadas no futuro.

O teste de associação de palavras é apresentado por Spiteri (2005) e ocorre da seguinte forma: os usuários recebem uma lista de termos de estímulo específicos do domínio e são solicitados a fazer associações, especificando como eles acham que esses termos estão relacionados. Feito isso, são geradas categorias de termos que são analisados conforme os tipos de relacionamentos identificados pela maioria dos participantes. Os testes de associação de palavras representam uma ferramenta potencialmente útil na construção de tesouros, principalmente, para envolver usuários finais nesse processo.

O método de expansão de consultas é utilizado por Liebeskind, Dagan e Schler (2018) para dar suporte à construção semiautomática de um tesouro no idioma Hebraico, avaliando a interação de termos do Hebraico moderno e antigo. No método, termos candidatos são extraídos



automaticamente de um corpus textual e, posteriormente, um especialista de domínio decide se esse termo deve ser incluído como uma entrada autorizada do tesouro ou apenas como uma remissiva. O processo iterativo é aplicado sobre a lista de expansões e a cada consulta no sistema pode ocorrer a remoção ou a inclusão de novos termos. Em seguida, o especialista do domínio seleciona termos do Hebraico antigo para o tesouro, adequando-os e incluindo à lista de expansão de consultas.

Em Pu e Chien (2004) é apresentado uma abordagem para investigar a construção de um tesouro automático a partir do interesse de busca dos usuários na *web*, considerando os vocabulários que estes utilizam ao pesquisar. A técnica consiste em duas etapas: a primeira em extrair termos relevantes a partir de seções de consulta; e a segunda em agrupar os termos relevantes em classes conceituais, técnica denominada por eles de “*concept clustering*”. Podem ser extraídos muitos termos conceitualmente relacionados à solicitação de pesquisa, mas que não ocorrem frequentemente em documentos. Como os termos relevantes extraídos podem ser muitos e conter significados diferentes, para torná-los mais compreensíveis ou bem organizados a abordagem proposta usa um método de *cluster* de consulta para organizar ainda mais os termos relevantes em classes.

Um dos recursos importantes da abordagem proposta é utilizar os *logs* das seções de consulta para extrair termos relevantes e organizá-los em *clusters* de conceitos. Ele se baseia na suposição de que alguns termos relacionados como sinônimos e termos alternativos podem aparecer em um conjunto de consultas com sessões semelhantes. Para explorar consultas e documentos da *web* como fonte de *corpus*, a abordagem proposta foi projetada para se integrar a um mecanismo de pesquisa (PU; CHIEN, 2004).

Diante de todas as abordagens apresentadas, algumas evidências são elucidadas e colocam em realce as práticas colaborativas, muitas delas percorridas como possíveis soluções para enfrentar problemas de tempo e custos de manutenção, gerenciamento compartilhado, qualidade e efetividade do tesouro construído. Nesse sentido, a colaboração pode ser caracterizada como:

- a) formal ou direta e não-formal ou indireta;
- b) ocorrer entre organizações, especialistas do domínio, profissionais da informação e usuários do sistema;
- c) de maneira síncrona ou assíncrona;
- d) mediante grupos de pessoas localizados no mesmo espaço ou geograficamente dispersos.

Com efeito, a colaboração pode não estar necessariamente formalizada no método proposto, tendo em vista que os procedimentos para a construção de tesouros são bem especificados nas normas, mas principalmente nas aplicações, ferramentas ou sistemas colaborativos desenvolvidos com o propósito de integrar os procedimentos, permitindo que as ações sejam executadas sob o paradigma de redes. Assim, é possível sistematizar como ocorrem essas ações. O Quadro 3 sumariza quais são os principais atores sociais envolvidos e que materiais e métodos podem ser empregados em cada processo.

**Quadro 3** - Atores, materiais e métodos distribuídos por processos

Processos	Atores	Materiais	Métodos
Planejamento	<ul style="list-style-type: none"> <li>- Especialistas de domínio;</li> <li>- Organizações;</li> <li>- Profissionais da informação;</li> <li>- Usuários do sistema.</li> </ul>	<ul style="list-style-type: none"> <li>- Normas técnicas;</li> <li>- Guias de referência do W3C;</li> <li>- Tesouros existentes;</li> <li>- <i>softwares</i> de gestão de tesouros;</li> <li>- Correio eletrônico (<i>e-mail</i>).</li> </ul>	<ul style="list-style-type: none"> <li>- Fóruns de discussão;</li> <li>- Comitê de especialistas;</li> <li>- Reuniões presenciais.</li> </ul>
Compilação de termos	<ul style="list-style-type: none"> <li>- Especialistas de domínio;</li> <li>- Organizações;</li> <li>- Profissionais da informação;</li> <li>- Usuários do sistema.</li> </ul>	<ul style="list-style-type: none"> <li>- Tesouros existentes;</li> <li>- Dicionários e enciclopédias eletrônicas;</li> <li>- Agências de notícias;</li> <li>- Periódicos especializados;</li> <li>- Conjuntos de dados (<i>datasets</i>).</li> </ul>	<ul style="list-style-type: none"> <li>- Mineração na <i>web</i>, de texto ou dados;</li> <li>- Marcação social (<i>tagging</i>);</li> <li>- Análise de <i>logs</i> de usuários;</li> <li>- Processamento de Linguagem Natural.</li> </ul>
Processamento automático e semiautomático de documentos	<ul style="list-style-type: none"> <li>- Profissionais da informação.</li> </ul>	<ul style="list-style-type: none"> <li>- Tesouros existentes;</li> <li>- Dicionários e enciclopédias eletrônicas;</li> <li>- Agências de notícias;</li> <li>- Periódicos especializados;</li> <li>- Conjuntos de dados (<i>datasets</i>).</li> </ul>	<ul style="list-style-type: none"> <li>- Mineração na <i>web</i>, de texto ou dados;</li> <li>- Processamento de Linguagem Natural.</li> </ul>
Registro de termos	<ul style="list-style-type: none"> <li>- Profissionais da informação.</li> </ul>	<ul style="list-style-type: none"> <li>- Fichas terminológicas;</li> <li>- <i>softwares</i> de gestão de tesouros.</li> </ul>	<ul style="list-style-type: none"> <li>- Fóruns de discussão;</li> <li>- Comitê de especialistas.</li> </ul>
Verificação dos termos	<ul style="list-style-type: none"> <li>- Especialistas de domínio;</li> <li>- Profissionais da informação.</li> </ul>	<ul style="list-style-type: none"> <li>- <i>softwares</i> de gestão de tesouros;</li> <li>- Correio eletrônico (<i>e-mail</i>).</li> </ul>	<ul style="list-style-type: none"> <li>- Análise de <i>logs</i>;</li> <li>- Fóruns de discussão;</li> <li>- Reuniões presenciais.</li> </ul>
Especificidade dos termos	<ul style="list-style-type: none"> <li>- Especialistas de domínio;</li> <li>- Profissionais da informação.</li> </ul>	<ul style="list-style-type: none"> <li>- <i>softwares</i> de gestão de tesouros;</li> <li>- Correio eletrônico (<i>e-mail</i>).</li> </ul>	<ul style="list-style-type: none"> <li>- Fóruns de discussão;</li> <li>- Reuniões presenciais;</li> <li>- Comitê de especialistas.</li> </ul>

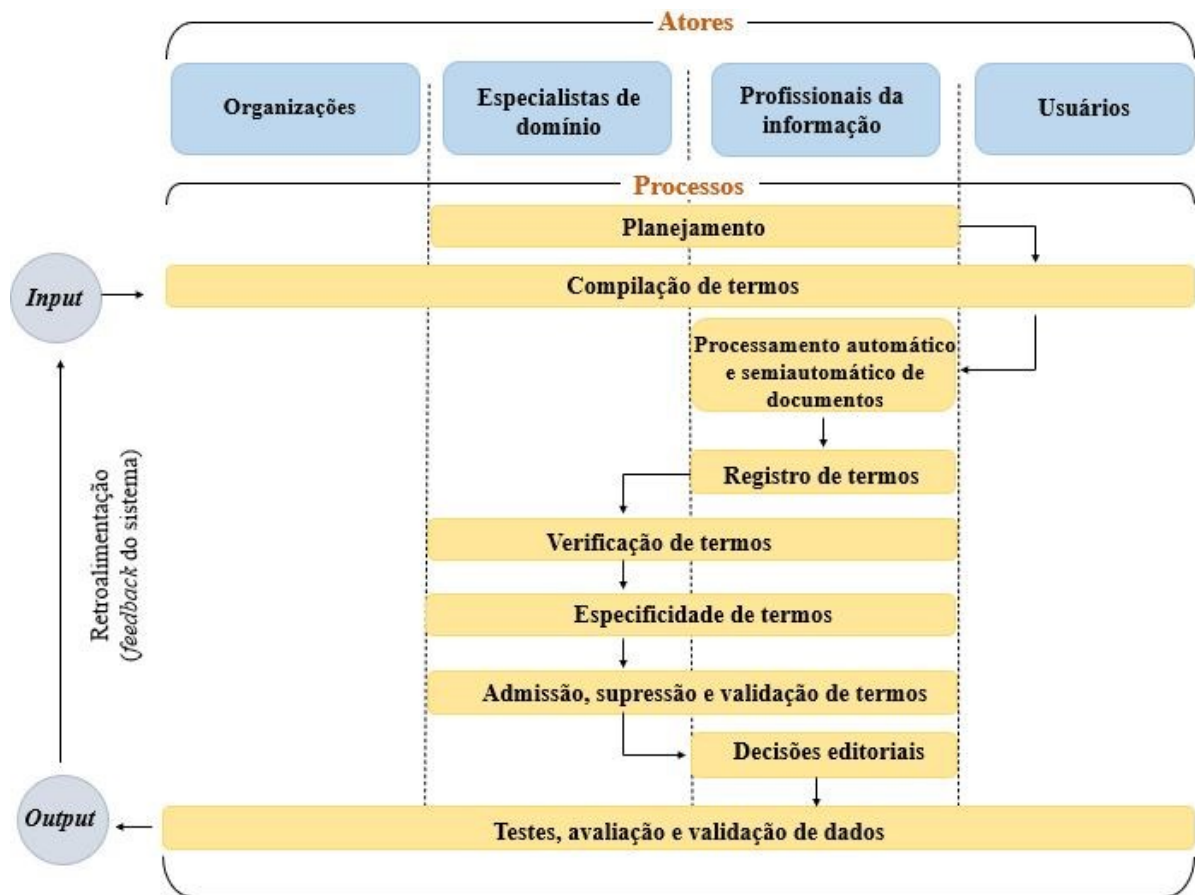
Admissão, supressão e validação de termos	- Especialistas de domínio; - Profissionais da informação.	- <i>softwares</i> de gestão de tesouros; - Correio eletrônico ( <i>e-mail</i> ).	- Fóruns de discussão; - Reuniões presenciais; - Comitê de especialistas.
Decisões editoriais	- Profissionais da informação.	- <i>softwares</i> de gestão de tesouros; - Guias de referência do W3C; - Normas técnicas.	- Fóruns de discussão; - Reuniões presenciais; - Comitê de especialistas.
Testes, avaliação e validação dos dados	- Especialistas de domínio; - Organizações; - Profissionais da informação; - Usuários do sistema.	- Sistemas de Recuperação da Informação; - Redes sociais ad hoc; - <i>softwares</i> de gestão de tesouros; - Serviço de <i>Endpoints</i> ; - Validadores de dados.	- Marcação social ( <i>tagging</i> ); - Análise de <i>logs</i> de usuários; - Regras de validação.

**Fonte:** Elaborado pelo autor (2021).

As práticas são evidentes em todas as etapas, com destaque para as fases de compilação, especificidade, inclusão e supressão de termos, testes e avaliação do vocabulário. Inúmeras iniciativas contam com a colaboração de instituições governamentais, de cultura, ensino ou pesquisa que fornecem o *corpus* necessário para a compilação de termos candidatos à inclusão nos tesouros. Em geral, esses projetos com a iniciativa de dados ou arquivos abertos disponibilizam livremente seus dados, léxicos ou vocabulários na *web* e são utilizados principalmente na abordagem de construção automática e semiautomática.

Diante disso, objetivando melhor ilustrar a integração das atividades de elaboração construiu-se um fluxograma que identifica a participação dos atores envolvidos em cada processo. A Figura 2 mostra que a construção e manutenção de tesouros é um processo cíclico e continuado que inclui um elemento de entrada (compilação de termos), um processamento e um elemento de saída (teste e avaliação), sendo este último capaz de gerar o *feedback* em termos de qualidade do produto gerado.

**Figura 2** - Fluxograma dos processos colaborativos na construção de tesouros



**Fonte:** Elaborado pelo autor (2021).

Os processos em que todos os atores participam se concentram na entrada e na saída, ou seja, nas etapas de compilação de termos e nas etapas de testes, avaliação e validação dos dados. Esses processos apresentam um alto grau de colaboração.

As operações de planejamento, verificação, especificidade e admissão e supressão de termos são consideradas atividades técnicas e exigem a participação tanto de especialistas do domínio como de profissionais da informação. Ainda, exprimem um grau de colaboração médio. Quanto às etapas de registro de termos, processamento automático de documentos e decisões editoriais são executadas por profissionais da informação e por isso apresentam um grau mínimo de colaboração, visto englobarem uma equipe de profissionais especializados das áreas de Informática e Ciência da Informação.

Além de avaliar a consistência semântica dos tesouros, há uma necessidade de verificar a consistência estrutural das cadeias de relacionamentos ou vínculos que representam esses recursos na *web*. Assim, com o propósito de averiguar a qualidade da estrutura desses dados, as instituições responsáveis por manter tesouros publicados como LOD frequentemente

disponibilizam na *web* um serviço de validação de dados, requisito fundamental para que outras organizações possam reutilizá-los em seus repositórios (CARACCIOLO et al., 2011).

O fluxograma mostra que os processos de entrada e saída são pontos de contato com o cliente/usuário, momento que estes contribuem com a elaboração do tesauro e do mesmo modo julgam o serviço mediante testes e avaliação nos SRI. Essas práxis representam um forte potencial de agregação de valor e devem ser observadas ao máximo para que se cumpra o objetivo do tesauro.

As práticas colaborativas empregadas em cada processo devem ser suportadas por sistemas de gestão de tesouros. Destacam-se o desenvolvimento e a aplicação de sistemas integrados que permitem o gerenciamento do vocabulário baseado em *Web Service*, projetos de LOD, ferramentas e sistemas colaborativos que possibilitam a integração de equipes editoriais.

Na atualidade, os padrões e tecnologias da WS representam uma oportunidade para o uso de tesouros na *web*. Ao seguir as recomendações do W3C, os tesouros passam a ser interoperáveis com outros vocabulários, aptos para a reutilização e processamento. Ao integrar suas coleções de dados à comunidade LOD os tesouros favorecem o alinhamento de seus conjuntos de dados com demais aplicações na *web*.

## 2.2 O *LINKED DATA*, SEUS PRINCÍPIOS E TECNOLOGIAS

Desde o surgimento da *web* na década de 1990 até a atualidade, os recursos para o acesso e o uso da informação em ambientes digitais têm se modificado continuamente. Na *web* de documentos os recursos textuais e imagéticos são representados por meio de identificadores globais que aplicam do padrão *Hypertext Markup Language* (HTML) para acessar os dados. Assim, na perspectiva de inserir significado aos conteúdos disponíveis na rede e projetar uma *web* de dados, Berner-Lee, Hendler e Lassila (2001) introduziram no início dos anos 2000 o conceito de WS. Desde então, tecnologias têm sido aprimoradas intentando otimizar os meios para publicar, acessar e intercambiar conteúdos, não somente pela sintaxe, como também pela semântica dos dados.

A maioria dos conteúdos na *web* é projetado para a leitura por humanos, no entanto há aqueles facilmente manipulados a partir de programas de computador. Em vista disso, as tecnologias da WS foram desenvolvidas para permitir que agentes de *softwares* possam rapidamente realizar tarefas para os usuários a partir do significado dos dados disponibilizados na rede (BERNERS-LEE; HENDLER; LASSILA, 2001). Essa iniciativa visa aproximar a

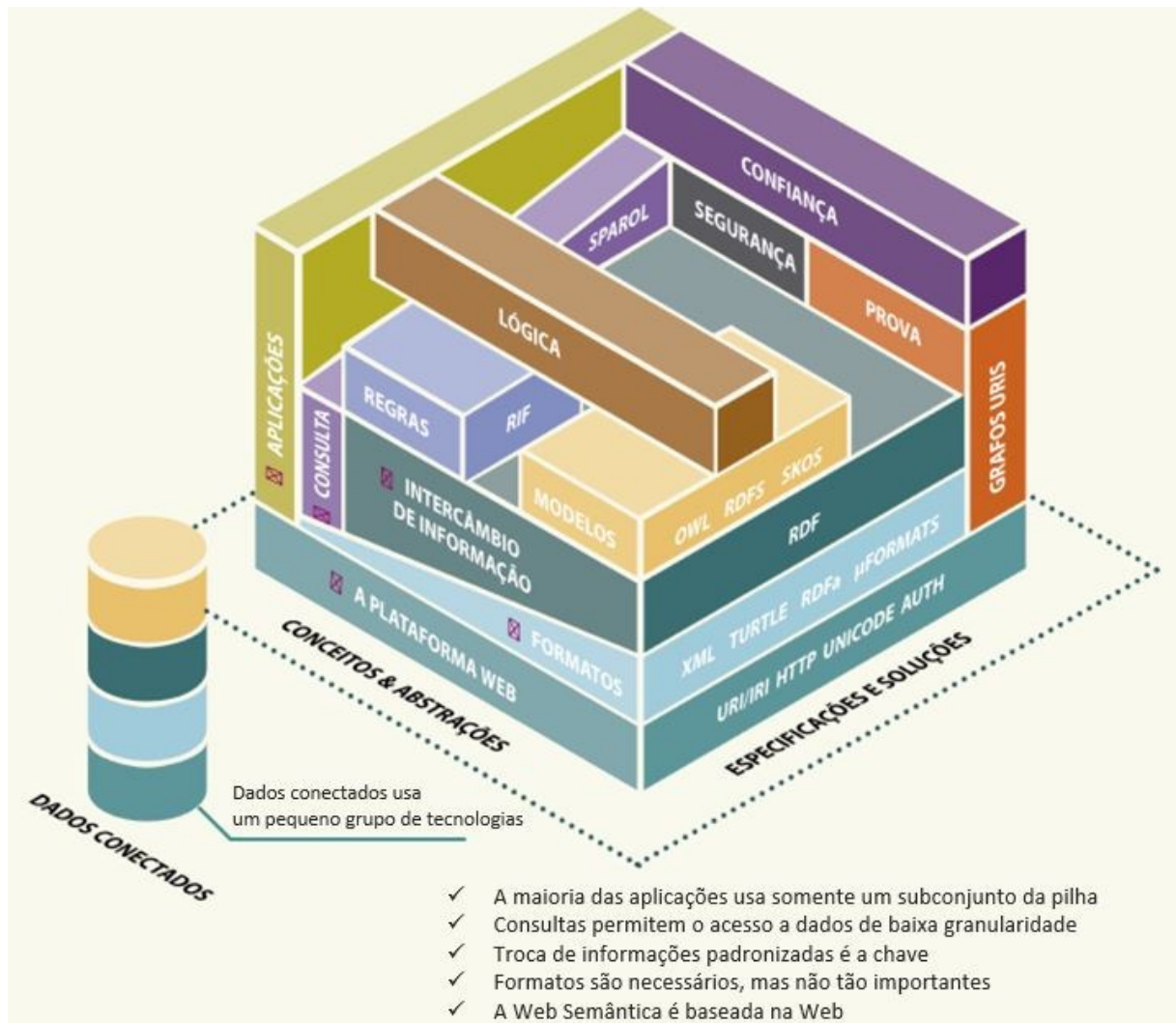
comunicação das pessoas com agentes ou sistemas computacionais e tornar o ciberespaço um ambiente de interação cada vez mais rico.

Para proporcionar esse ambiente de interação, os computadores precisam ter acesso às coleções de informação estruturadas para que por via de um conjunto de regras possam realizar o processamento automático sobre essas coleções. A formalização e estruturação desse acervo promovem a interoperabilidade dos dados, permitindo que as aplicações nesse ambiente compreendam o significado de cada recurso descrito e executem tarefas por meio de associações semânticas. Todavia, é essencial o uso de vocabulários específicos, ontologias e padrões de metadados (LIMA; CARVALHO, 2004).

Dada a pertinência das questões postas no projeto inicial, a WS tem se difundido e agregado diversos modelos de dados que se apoiam nas discussões dos grupos de trabalho do W3C, este último fundado por Tim Berners-Lee em 1994 com a pretensão de ser um consórcio internacional com vistas a desenvolver e lançar ferramentas, tecnologias e recomendações para a *Web* (W3C, 2018).

Na época, Berners-Lee desenvolveu um modelo de camadas da WS elaborado para dar uma visão mais clara em termos de sua implementação, descrevendo as principais tecnologias utilizadas. À medida que a área foi se expandindo a estrutura de camadas foi reavaliada (Figura 3) e, atualmente, demonstra toda a complexidade das tecnologias envolvidas, contemplando conceitos, abstrações, especificações e soluções da WS.

**Figura 3** - Estrutura de camadas da Web Semântica



**Fonte:** Isotani e Bittencourt (2015)

Uma das vantagens da *web* é permitir que diferentes comunidades globais possam compartilhar informações, as quais são representadas por recursos computacionais acessíveis codificados em diversas linguagens que expressam algum tipo de significado (GANGEMI; PRESUTTI, 2006). Nessa perspectiva, o W3C entende que todos os dados na *web* necessitam ser concedidos para processamento, recuperação e reuso, e por isso defende uma rede de dados conectados (*Linked Data*).

O conceito de *Linked Data* foi difundido por Berners-Lee em 2006 e se refere a “[...] um conjunto de práticas recomendadas para publicar e interligar dados estruturados na *web*.” (HEATH; BIZER, 2011, tradução nossa). Os dados conectados utilizam um pequeno grupo de tecnologias representadas na estrutura de camadas da WS e seus princípios são apresentados sob o agrupamento de quatro regras:

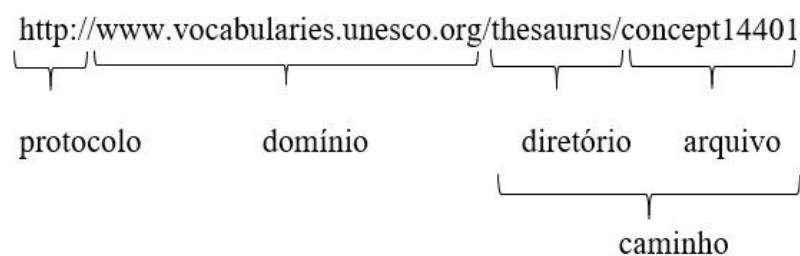
1. Use URIs como nomes para as coisas;
2. Use URIs HTTP, para que as pessoas possam acessar esses nomes;
3. Quando alguém acessar um URI, forneça informações úteis utilizando os padrões (RDF, SPARQL);
4. Inclua links para outros URIs. Para que eles possam descobrir mais coisas. (BERNERS-LEE, 2009, não paginado, tradução nossa).

Para projetar uma *web* de dados conectados algumas especificações apresentadas nas quatro regras e também ilustradas na pilha de linguagens da WS são essenciais. Além dos mecanismos de identificação global, como os *Uniform Resource Identifier* (URI) e o *Hypertext Transfer Protocol* (HTTP), o *Linked Data* tem como base o uso do modelo de dados *Resource Description Framework* (RDF), os tipos de notação ou sintaxes de serialização para representar as triplas RDF e o protocolo SPARQL (ISOTANI; BITTENCOURT, 2015), que serão abordados na sequência.

### 2.2.1 Identificadores Uniformes de Recursos (URI)

Os recursos disponíveis na *web* precisam de um endereço único para que os usuários possam encontrá-los rapidamente. Esse endereço é chamado de URI e consiste em um sistema de identificação comum para habilitar o acesso de forma permanente e confiável aos recursos na *web*. Os URI são identificadores formados por um conjunto de caracteres, sendo muito usados para representar palavras de linguagens naturais (DUERST; SUIGNARD, 2005).

Os URI podem se apresentar em dois tipos: os *Uniform Resource Names* (URN) e os *Uniform Resource Locators* (URL). Embora este último seja o mais conhecido entre os usuários, pois corresponde aos endereços de recursos para acesso *online* (FERREIRA; SANTOS, 2013), o URI é o tipo de esquema de endereçamento mais abrangente. No contexto do *Linked Data*, a recomendação de utilizá-los vai ao encontro quanto à necessidade de promover acesso aos conjuntos de dados de maneira que eles possam ser descobertos e citados permanentemente ao longo do tempo, independente de sua disponibilidade ou formato. O exemplo a seguir ilustra como os URI são estruturados:





Os endereços são compostos por três elementos principais: protocolo para acessar o recurso (HTTP); endereço da máquina que hospeda o recurso, conhecido como domínio; e o próprio recurso, composto por diretório e arquivo. O endereço dá acesso ao documento “concpet14401”, que pode ser um arquivo HTML ou em outros formatos.

O HTTP é um protocolo que surgiu a partir da necessidade de padronizar a comunicação ou o tratamento de pedidos e respostas entre o cliente e um servidor *web*. Nessa oportunidade, ao estabelecer a conexão com o servidor o cliente envia uma requisição, que pode ser representada pelo URI, e o servidor retorna para o cliente as informações sobre o recurso solicitado.

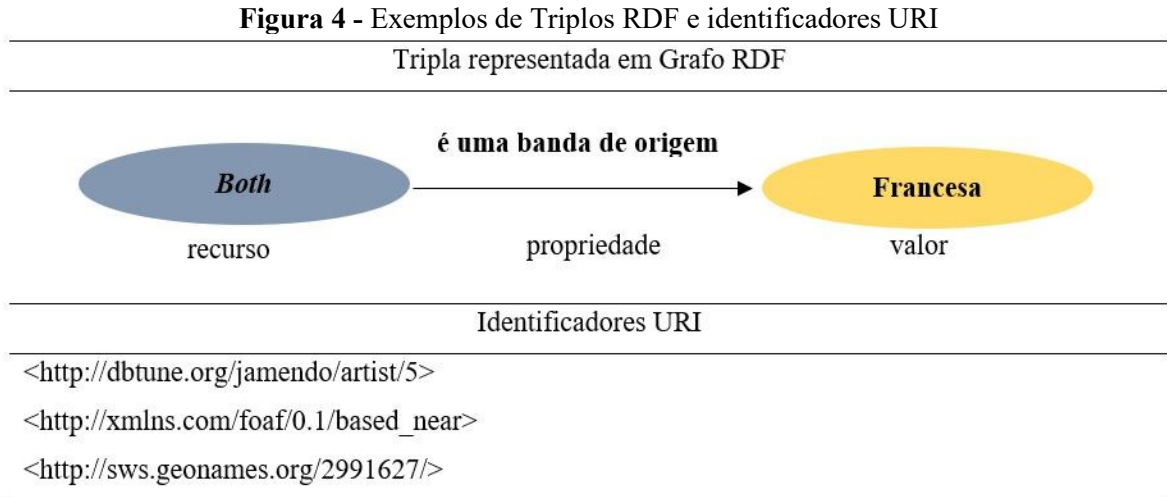
Os identificadores podem assumir muitas formas, todavia para o uso e citação de dados na *web* dependem do uso de HTTP URI (RAUTENBERG et al., 2018). Vale ressaltar que os URI não são capazes de carregar semântica, mas devem ser igualmente compreendidos por humanos, logo recomenda-se que os caracteres empregados para representar o caminho do recurso não sejam passíveis de alteração ao longo do tempo. Assim, deve-se evitar o uso de palavras ou expressões e priorizar caracteres numéricos.

Com isso, destaca-se a importância de aplicar identificadores persistentes para se obter informações sobre os recursos globalmente, uma vez que os URI são a base para a descrição de recursos na *web* a partir da linguagem RDF.

### 2.2.2 O *Resource Description Framework* (RDF)

O RDF é reconhecido como uma linguagem ou modelo para representar metadados sobre recursos da *web* e para caracterizar informações a respeito dos meios que podem ser identificados na rede a partir de três elementos denominados triplas: sujeitos, predicados e objetos. Nesse modelo, o **sujeito** (recurso) e o **objeto** (valor) representam dois recursos relacionados, e o **predicado** (propriedade) configura a natureza de seu relacionamento, formulado de maneira direcional.

O RDF é baseado na ideia de grafos, extensível, e por isso facilita a interligação de dados de diferentes fontes. A Figura 4 ilustra como se configura os relacionamentos entre recursos no modelo RDF, em que cada elemento é identificado por um URI (W3C, 2014a).



**Fonte:** Elaborado pelo autor (2021).

Conforme a ilustração, cada recurso na *web* é identificado por seus URI e utilizados para qualificar de maneira uniforme os recursos no modelo RDF. Os dados podem estar armazenados em um banco de dados de triplas (*triple store*) com um esquema de mapeamento para o RDF. Assim, podem haver diversas triplas que fazem parte de um ecossistema de dados na *web* usadas para explorar essa infraestrutura, visto que a partir delas diferentes fontes de dados podem ser combinadas para formar um único grafo.

A aplicação do RDF é uma recomendação do W3C que disponibiliza em sua página oficial um conjunto de documentos instrutivos que incluem: guia introdutório; descrição do modelo de dados RDF; especificação técnica do RDF; e documentos que especificam a serialização do RDF em formatos específicos (W3C, 2014a, 2014b, 2014c).

O W3C (2005a) enfatiza que “[...] a maneira mais simples de publicar dados RDF é criar um ou mais documentos RDF contendo seus dados e publicá-los na *web* por meio de um servidor HTTP”. Desse modo, os URI identificam os recursos declarados a partir da linguagem RDF, que são registradas em um arquivo nesse formato e precisam ser representadas seguindo uma configuração de serialização.

### 2.2.3 Sintaxes de serialização para os documentos RDF

Serializar as declarações RDF implica em utilizar uma notação específica para registrar as triplas em um arquivo, de maneira a facilitar tanto a sua leitura por máquinas como por humanos. Essas notações são conhecidas também como sintaxes, termo utilizado na Ciência da Computação para se referir à tradução de estruturas de dados para um formato específico, tornando-o apto para o armazenamento ou a transmissão entre sistemas.

A notação RDF/XML foi a primeira sintaxe desenvolvida para o RDF, no entanto há outras sintaxes alternativas, como o *Notation 3 (N3)*, *Turtle*, *N-Triples*, e o *JavaScript Object Notation for Linked Data (JSON-LD)* (HEATH; BIZER, 2011). As figuras abaixo (Figuras 5 a 9) apresentam exemplos dos diversos tipos de sintaxes de serialização, tendo como base o caso do recurso representado em Grafo RDF na Figura 4.

Na sintaxe RDF/XML<sup>13</sup> (Figura 5), as triplas são representadas em termos XML. Os vocabulários empregados são descritos a partir da *tag* <rdf:RDF>, enquanto que o elemento rdf:Description descreve o recurso, propriedades e seus valores.

**Figura 5** - Exemplo de serialização em RDF/XML

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:foaf="http://xmlns.com/foaf/0.1/"
5   xmlns:gn="http://sws.geonames.org/ontology#">
6 <rdf:Description rdf:about="http://dbtune.org/jamendo/artist/5">
7   <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/name">
8   <rdf:type rdf:resource="http://sws.geonames.org/2991627/">
9   <foaf:name>Both</foaf:name>
10  <gn:countryCode>FR<gn:countryCode>
11 </rdf:Description>
12 </rdf:RDF>

```

**Fonte:** Elaborado pelo autor (2021).

A *Notation 3*<sup>14</sup>, ou apenas N3, é uma serialização de notação não-XML projetada para a rápida e fácil compreensão por humanos. Sua notação é compacta, composta apenas pela *tag* @prefix para vincular um prefixo a uma URI. O sujeito é especificado entre os sinais de maior (>) e menor (<), e as *strings* de objetos e predicados são descritos entre aspas duplas (“ ”).

**Figura 6** - Exemplo de serialização em Notation 3 (N3)

```

1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
3 @prefix gn: <http://sws.geonames.org/ontology#> .
4
5 < http://dbtune.org/jamendo/artist/5>
6   foaf:name:title "Both";
7   gn:countryCode "FR".

```

**Fonte:** Elaborado pelo autor (2021).

<sup>13</sup> <https://www.w3.org/TR/rdf-syntax-grammar/>

<sup>14</sup> <https://www.w3.org/TeamSubmission/n3/>

O *Turtle*<sup>15</sup> é considerado um subconjunto da N3. O exemplo da Figura 7 também é válido para o *Turtle*. A diferença entre as duas sintaxes é que o *Turtle* é usado apenas para serializar gráficos RDF válidos, não permitindo nodos em branco, como no N3. O formato utiliza abreviações para especificar os padrões utilizados e os tipos de dados.

**Figura 7** - Exemplo de serialização em *Turtle*

```

1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
3 @prefix gn: <http://sws.geonames.org/ontology#> .
4
5 < http://dbtune.org/jamendo/artist/5>
6     foaf:name:title "Both";
7     gn:countryCode "FR".

```

**Fonte:** Elaborado pelo autor (2021).

A notação *N-triples*<sup>16</sup> é um subconjunto da notação *Turtle*, todavia é considerada mais simples porque cada linha equivale a uma tripla. As triplas são finalizadas por um ponto, e as *strings* também são descritas usando aspas duplas (“ ”) e delimitados a partir dos sinais maior (>) e menor (<).

**Figura 8** - Exemplo de serialização em N-Triples

```

1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#> <http://xmlns.com/foaf/0.1/>
  <http://sws.geonames.org/ontology#> .
2 < http://dbtune.org/jamendo/artist/5> <http://xmlns.com/foaf/0.1/name>
  <http://sws.geonames.org/2991627/"> .

```

**Fonte:** Elaborado pelo autor (2021).

O formato JSON-LD<sup>17</sup> é uma sintaxe criada mais recentemente e surgiu com o objetivo de transformar o código JSON – formato para a troca de dados entre sistemas que utilizam a linguagem de programação Java – para o RDF. Na Figura 9 pode-se observar que os vocabulários usados são representados a partir da *tag* “@context”, descritos no intervalo iniciado e finalizado como o símbolo de chaves ({ }), cujos relacionamentos entre o recurso e os valores são evidenciados pela *tag* @graph, das quais as declarações são introduzidas e conclusas com o símbolo colchetes ([ ]).

<sup>15</sup> <https://www.w3.org/TR/turtle/>

<sup>16</sup> <https://www.w3.org/TR/n-triples/>

<sup>17</sup> <https://www.w3.org/TR/json-ld11/>

**Figura 9** - Exemplo de serialização em JSON-LD

```

1  {
2    "@context": {
3      "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
4      "foaf": "http://xmlns.com/foaf/0.1/",
5      "gn": "http://sws.geonames.org/ontology#"
6    },
7    "@graph": [
8      {
9        "@id": "http://dbtune.org/jamendo/artist/5"
10       "foaf:name": "Both"
11     },
12     {
13       "@id": "http://sws.geonames.org/2991627/"
14       "gn:CountryCode": "FR",
15     }
16   ]
17 }

```

**Fonte:** Elaborado pelo autor (2021).

Em razão das declarações serem registradas em arquivos RDF e armazenadas em servidores *Triple store*, o acesso e a consulta desses dados devem ser sucedidos mediante técnicas e linguagem específicas. Desse modo, para executar buscas, recuperar e manipular dados nos *triple stores* o W3C recomenda o uso do *SPARQL Protocol and RDF Query Language* (SPARQL), sendo este um conjunto de especificações que oportunizam a consulta de dados representados em RDF (W3C, 2008b, 2014b, 2016a).

#### 2.2.4 SPARQL: protocolo e linguagem de consulta RDF

O SPARQL é o acrônimo operado para se referir à linguagem e ao protocolo para a consulta de dados RDF. Foi desenvolvido pelo grupo de trabalho do W3C objetivando viabilizar a transmissão de consultas e a manipulação de informações contidas em grafos RDF. Com o avanço das tecnologias semânticas, houve a necessidade de desenvolver uma linguagem para as demandas de processamento de dados semânticos, superando as limitações da linguagem de consulta estrutura SQL destinada à consulta em bancos de dados relacionais. Para essa ação, o protocolo apresenta duas operações específicas:

O Protocolo SPARQL consiste em duas operações HTTP: uma operação de consulta para a realização de consultas de linguagem SPARQL e uma operação de atualização para a realização de solicitações de atualização

do SPARQL. Os clientes do SPARQL Protocol enviam solicitações HTTP aos serviços do Protocolo SPARQL que lidam com a solicitação e enviam respostas HTTP de volta ao cliente de origem (W3C, 2013a, tradução nossa).

A consulta SPARQL pode ser feita em editores de ontologias ou por meio de um SPARQL Endpoint *web* – exemplificado na Figura 10. Um Endpoint é um ponto de contato no qual os usuários ou clientes podem se comunicar com um serviço na *web*. Por intermédio de diferentes formulários de consulta é possível descobrir novos relacionamentos, realizar inferências e identificar relações existentes entre os dados RDF (SEGUNDO; CONEGLIAN; LUCAS, 2017).

**Figura 10** - Serviço Endpoint do DBpedia

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)  
http://dbpedia.org

Query Text  
select distinct ?Concept where {[ ] a ?Concept} LIMIT 100

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: HTML

Execution timeout: 30000 milliseconds (values less than 1000 are ignored)

Options:  
 Strict checking of void variables  
 Log debug info at the end of output (has no effect on some queries and output formats)  
 Generate SPARQL compilation report (instead of executing the query)

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

**Fonte:** DBpedia<sup>18</sup> (2019).

O SPARQL apresenta uma estrutura de consulta baseada no comando SELECT-FROM-WHERE, que segundo Rautenberg et al. (2018) consiste basicamente em:

- a) SELECT: apresenta os resultados da consulta, especificando quais dados (atributos ou instâncias) serão retornados;

<sup>18</sup> <https://dbpedia.org/sparql>

- b) FROM: identifica as fontes de dados que serão consultadas;
- c) WHERE: especifica as restrições da consulta, ou seja, são definidos os padrões de triplas ou de grafos a serem localizados.

A ideia fundamental do SPARQL, no contexto do *Linked Data*, é estender a definição de um terminal de um triplo para um objeto de primeira classe referenciado por um URI. A consulta pode ser distribuída para vários outros serviços de consultas SPARQL Endpoint, denominada consulta federada. O crescente número de serviços de consulta SPARQL oferece aos consumidores de dados uma oportunidade de mesclar dados distribuídos pela *web*. Desse modo, os serviços de SPARQL Endpoint têm sido amplamente aplicados nas iniciativas de *Linked Data* (W3C, 2013b).

### 2.2.5 Atribuição de licenças para o reuso dos dados

Dada a proposta de publicar dados conectados na *web*, inúmeros projetos de *Linked Data* disponibilizam dados ligados sob uma licença aberta, fazendo surgir o conceito de *Linked Open Data* (LOD). A ideia do LOD implica em disponibilizar conjuntos de dados sem restrições para acesso, permitindo que os recursos na *web* sejam explorados e reutilizados sob licenças livres.

Para inferir qualidade aos projetos de LOD, Berners-Lee (2009) desenvolveu um sistema de classificação de estrelas. Quanto maior a quantidade de estrelas, mais conectados e abertos os dados são apresentados:

- ★ Disponível na *web* (em qualquer formato), mas como uma licença aberta, para ser **Open Data**;
- ★★ Disponível como dados estruturados e legíveis por máquina (por exemplo, uma planilha do Excel ao invés de uma imagem digitalizada de uma tabela);
- ★★★ Assim como no (2), mas em um formato não proprietário (por exemplo, no formato CSV ao invés do formato do Excel);
- ★★★★ Todos os anteriores, use padrões abertos do W3C (RDF e SPARQL) para identificar coisas, assim as pessoas poderão direcionar as suas coisas;
- ★★★★★ Todos os itens anteriores e: relacionar seus dados aos dados de outras pessoas para fornecer contexto (BERNERS-LEE, 2009, não paginado, tradução nossa).

Seguindo a recomendação da primeira estrela, a disponibilização de dados abertos na *web* deve ser acompanhada de um tipo de licença aberta que precisa estar alinhada com a

proposta desses dados nos quais preconizam a reutilização, o compartilhamento e a atualidade. Esta última é concebida a partir da prerrogativa de que a licença deve ser amplamente utilizada, atual e considerada “[...] a melhor prática por um amplo espectro de projetos e atores dentro dos domínios de aplicabilidade da licença.” (OPEN KNOWLEDGE INTERNATIONAL, 2015, tradução nossa).

A atribuição de licenças varia conforme o grau de restrição para o reuso dos dados e são classificadas em quatro tipos: atribuição, compartilhamento pela mesma licença, com restrições e fechadas (RODRIGUEZ DONCEL; GÓMEZ PÉREZ; MIHINDUKULASOORIYA, 2013).

As licenças de dados abertos mais usadas nos diferentes conjuntos de dados são o *Open Knowledge Definition*, *Open Database License* e várias modalidades do *Creative Commons* (PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012). Elas devem ser escolhidas consoante com a política de acesso à informação adotada pela instituição promotora, levando em conta a natureza dos dados e a legislação em vigor. Dentre os diversos tipos, Rodríguez-Doncel, Gómez-Pérez e Mihindikulasooriya (2013) destacam algumas das mais conhecidas:

- a) *Creative Commons*<sup>19</sup>: fornece licenças e ferramentas para conceder permissões de direitos autorais de trabalhos criativos e acadêmicos. Suas licenças permitem que outras pessoas copiem, distribuam e façam uso de obras por meio de diversas modalidades de atribuições;
- b) *Open Data Commons*<sup>20</sup>: fornece um conjunto de ferramentas legais para auxiliar os autores a fornecer e utilizar os dados abertos. Suas atribuições permitem copiar, distribuir, usar ou modificar dados para a criação de novos conjuntos de dados abertos na *web*;
- c) *French Open License*<sup>21</sup>: criado para a atribuição de licenças para dados governamentais da França. Suas licenças são baseadas no *Creative Commons* e aplicadas por diversas instituições governamentais do país, inclusive pela Biblioteca Nacional da França;
- d) Reino Unido - *Open Government License*<sup>22</sup>: licença de direitos autorais elaborada e mantida pelos Arquivos Nacionais. É compatível com a licença

<sup>19</sup> <http://creativecommons.org/>

<sup>20</sup> <https://www.opendatacommons.org/>

<sup>21</sup> <https://www.etalab.gouv.fr/licence-ouverte-open-licence>

<sup>22</sup> <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>



*Creative Commons* e empregada para documentos publicados pelo governo do Reino Unido;

- e) *GNU Free Documentation License*<sup>23</sup>: é uma licença para documentos e textos gratuitos publicados pela *Free Software Foundation*. Permite que textos, apresentações e conteúdos de páginas na *web* sejam distribuídos e lidos de maneira apropriada mediante diversos tipos de concessões.

Algumas dessas licenças – *Licença Open Government* do Reino Unido, a *Open License Francesa* e a *GNU Free Documentation License* – foram criadas para instituições específicas ou em razão de demandas do governo. Contudo, Korn e Oppenheim (2011) argumentam que o uso de licenças padrão pode levar a uma maior interoperabilidade e conscientização do usuário. Muitos projetos de dados abertos conectados escolhem as licenças do *Creative Commons*, pois apresentam termos mais flexíveis, são familiares, disponíveis em formato legível por máquina e podem ser facilmente compreendidas por humanos.

Em geral, uma licença de dados abertos pode não ser útil se houver dados não disponíveis para a reutilização ou se o conjunto de dados não puder ser pesquisado através de um terminal SPARQL. Caso contrário, o nível de integração do ecossistema de dados abertos pode ser muito baixo (PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012).

Uma vez estruturados nos padrões do *Linked Data* e disponibilizados sob licença aberta, o conjunto de dados passa a fazer parte de uma grande rede de dados abertos conectados. Pode haver diagramas que ilustram a vinculação de domínios e conjuntos de dados específicos como, por exemplo, DBpedia<sup>24</sup>, WikiData<sup>25</sup> e o *Global Research Identifier Database* (GRID)<sup>26</sup>. A maioria dos conjuntos de dados compõe a *Linked Open Data Cloud* (LOD cloud). No Brasil, é conhecido como Nuvem LOD (Figura 11). Esse diagrama é mantido pelo *Insight Center for Data Analytics* e mostra todos os conjuntos de dados interligados publicados por membros da comunidade LOD, organizações ou pessoas (INSIGHT CENTRE FOR DATA ANALYTICS, 2019).

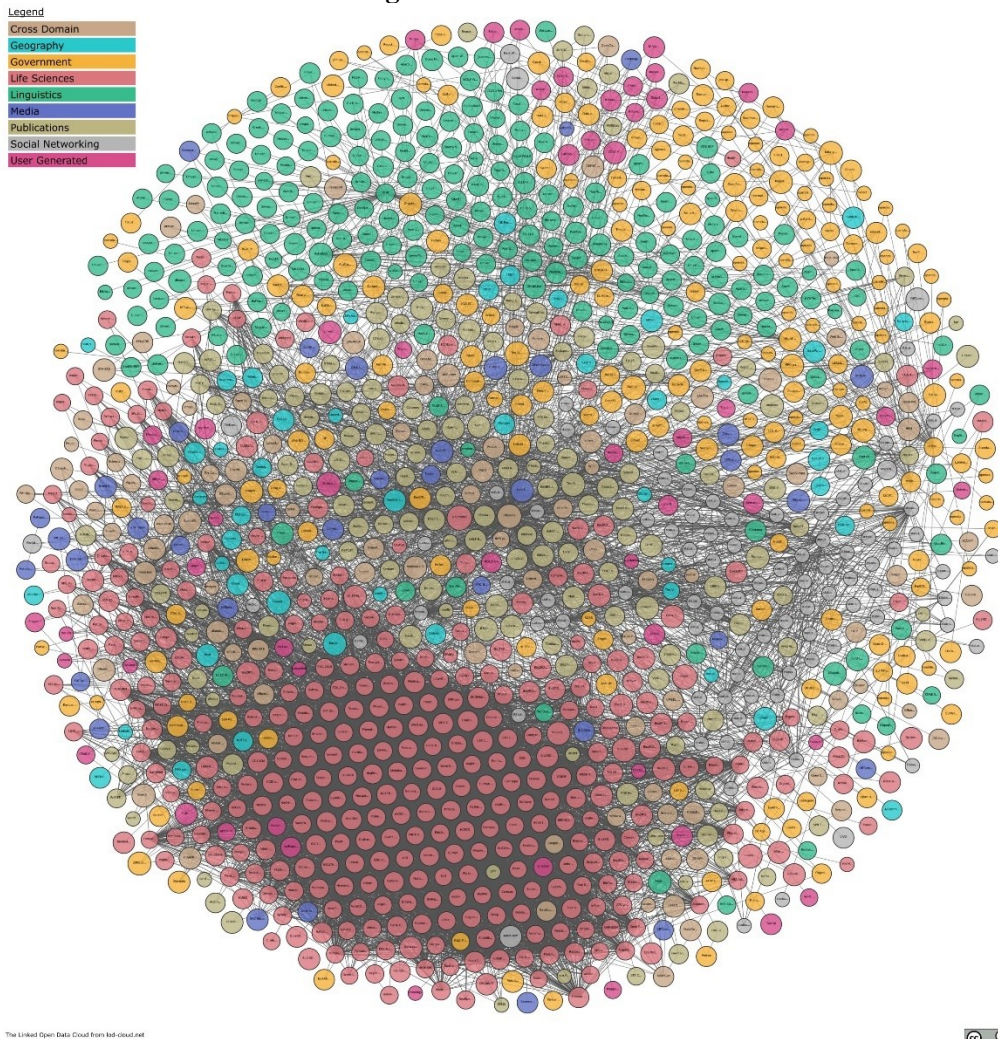
<sup>23</sup> <https://www.gnu.org/licenses/fdl-1.3.html>

<sup>24</sup> Conjunto de dados contendo dados extraídos da Wikipedia; contém cerca de 3,4 milhões de conceitos descritos por 1 bilhão de triplos, incluindo resumos em 11 idiomas diferentes (<https://wiki.dbpedia.org/>).

<sup>25</sup> Conjunto de dados vinculado criado em colaboração que atua como armazenamento central para os dados estruturados de seus projetos irmãos da Wikimedia Foundation (<https://www.wikidata.org/>).

<sup>26</sup> Um banco de dados internacional de 89.506 instituições envolvidas em pesquisa acadêmica, com 14.401 relacionamentos, modela dois tipos de relacionamento: um relacionamento pai-filho que define uma associação subordinada e um relacionamento relacionado que descreve outras associações (<https://www.grid.ac/>).

**Figura 11 - Nuvem LOD**



**Fonte:** *The Linked Open Data Cloud* (mar. 2021).

O desenvolvimento de uma *web* de dados conectados fez com que inúmeras aplicações na rede incorporassem as tecnologias da WS, aderindo especialmente ao conjunto de práticas do LOD. Porém, para que os projetos contribuam com a construção e evolução de uma *web* de dados conectados é preciso que os princípios e tecnologias sejam aplicados adequadamente. Para isso, precisa-se seguir com rigor as recomendações de boas práticas do W3C.

### 2.2.6 Melhores práticas para a publicação de dados na *web*

Com o propósito de apoiar o ecossistema de dados conectados, o W3C apresenta um conjunto de diretrizes reconhecidas como melhores práticas<sup>27</sup> para que desenvolvedores e

<sup>27</sup> <https://www.w3.org/TR/dwbp/#intro>

consumidores de dados possam aproveitar todos os recursos e vantagens da publicação de dados conectados. Esse conjunto é formado por 35 diretrizes, apresentadas de modo condensadas no Quadro 4.

**Quadro 4 - Melhores práticas para a publicação de dados na web**

<b>Melhores práticas</b>	<b>Benefícios</b>
1. Forneça metadados	Permite que os dados sejam compreensíveis tanto por humanos quanto por máquinas.
2. Forneça metadados descritivos	Possibilita conhecer as características gerais dos conjuntos de dados e suas distribuições.
3. Forneça metadados estruturais	Oportuniza reconhecer o esquema e a estrutura interna dos conjuntos de dados.
4. Forneça informações sobre a licença dos dados	Fator determinante para o reuso dos dados, pois oferece detalhes sobre a licença que controla o seu uso.
5. Forneça informações sobre a proveniência dos dados	Detalhes sobre a origem ou alterações realizadas nos dados aumentam a confiança na credibilidade dos dados.
6. Forneça informações sobre a qualidade dos dados	Pontos sobre a qualidade e adequação de uso dos dados facilita a seleção dos dados para reuso.
7. Forneça indicadores de versões	Informações sobre versões dos dados.
8. Forneça informações sobre o histórico de versões	Minúcias acerca das várias versões de um conjunto de dados permite que os consumidores verifiquem como os dados têm sido alterados ao longo do tempo.
9. Use URI persistentes para identificar o conjunto de dados	Utiliza um sistema de identificação comum para que os dados possam ser percebidos e comparados de forma confiável.
10. Use URI persistentes como identificadores dentro do conjunto de dados	Aplica o sistema de identificação persistente para viabilizar a conexão ou os relacionamentos entre os recursos do conjunto de dados.
11. Declare URI para versões dos conjuntos de dados e séries	Usa o sistema de identificação persistente para se referir à situação corrente dos dados ou a uma série específica no tempo.
12. Use formatos de dados padrão legível por máquina	Emprega padrões de dados de maneira que eles possam ser extraídos, transformados e processados por sistemas de computador.
13. Use representações de dados neutros de localidade	Opera dados em um formato neutro de localidade para que não haja a necessidade de estabelecer regras específicas para o intercâmbio dos dados, como idioma e localização.
14. Forneça dados em múltiplos formatos	Facilita a transformações dos dados.
15. Reuse vocabulários, preferencialmente os padronizados	Reutiliza ontologias e vocabulários existentes para a inclusão e expansão da <i>web</i> de dados.
16. Escolha o nível de formalização apropriado	Opera o vocabulário de acordo com o nível de formalização exigido para os dados. Ajuda a transmitir o significado dos dados de maneira precisa e facilita sua exploração por motores de inferência.
17. Forneça <i>download</i> em massa	Indica que os consumidores possam descarregar grandes conjuntos de dados.
18. Forneça subconjuntos de dados	Facilita que os consumidores possam acessar e descarregar partes do conjunto de dados.
19. Use negociação de conteúdo para fornecer dados em múltiplos formatos	Utiliza a negociação de conteúdo para o acesso aos dados em diversos formatos de representação.

20. Forneça acesso em tempo real	Disponibiliza dados à medida que são produzidos para a rápida utilização por outras aplicações na <i>web</i> .
21. Forneça dados atualizados	Dados atualizados com frequência incentivam a confiança dos consumidores acerca dos dados e asseguram sua reutilização.
22. Forneça uma explicação para os dados que não estão disponíveis	Favorece que os publicadores de dados identifiquem lacunas e que os consumidores de dados utilizem apenas os dados disponíveis.
23. Torne os dados disponíveis usando interfaces de programação de aplicações (API)	As API permitem que os dados sejam usados em tempo real, oferecendo mais flexibilidade e capacidade de processamento para os consumidores de dados.
24. Use padrões <i>web</i> com base das API	Uma vez construídas sob padrões <i>web</i> , aproveitam os pontos fortes da infraestrutura da Internet para mapear recursos e conectar os dados.
25. Forneça uma documentação completa para API	Aponta que os consumidores compreendam e codifiquem aplicações de forma eficiente.
26. Evite alterações significativas na API	Minimiza a ruptura do código da aplicação do cliente. Permite que as características dos conjuntos de dados sejam protegidas e que os desenvolvedores continuem utilizando o conjunto de dados.
27. Preserve os identificadores	Viabiliza que o URI original continue identificando o conjunto de dados, mesmo que não estejam diretamente disponíveis.
28. Avalie a cobertura do conjunto de dados	Garante que os dados arquivados preservem seus recursos para que os usuários possam usar os dados novamente no futuro.
29. Obtenha <i>feedback</i> dos consumidores de dados	Proporciona que os publicadores de dados entendam melhor as necessidades dos consumidores de dados.
30. Torne o <i>feedback</i> disponível	Demonstra que as demandas dos consumidores de dados estão sendo observadas ou atendidas.
31. Enriqueça dados gerando novos dados	Melhora o processamento, principalmente de dados não estruturados, e ajuda dos desenvolvedores a publicar dados mais completos.
32. Forneça apresentações complementares	Enseja que os dados possam ser melhor visualizados ou explorados na <i>web</i> .
33. Forneça <i>feedback</i> para o publicador original	Permite que os publicadores de dados saibam como os seus dados estão sendo utilizados.
34. Siga os termos de licença	Minimiza preocupações sobre uma ação legal do publicador original dos dados.
35. Cite o publicador original	Assente que o publicador dos dados seja creditado e reconhecido. Mantém a proveniência e incentiva a confiança dos usuários que irão consumir os dados.

**Fonte:** W3C (2017, tradução nossa).

Observa-se que para publicar dados conectados na *web* não basta utilizar padrões ou tecnologias da *Web Semântica*. É possível perceber a importância de promover os dados junto à comunidade que os utiliza. Logo, todos os atores sociais estão envolvidos com a produção e a publicação dos dados, seja no enriquecimento, seja na avaliação da qualidade. Nesse sentido, vale ressaltar os benefícios do *feedback* dos consumidores ou usuários dos dados, considerado

um dos principais componentes de retroalimentação dessa cadeia de produção, pois tem o potencial de identificar debilidades nos conjuntos de dados, promover melhorias e qualificar como os dados estão sendo reutilizados pela comunidade.

A disponibilização de dados de forma livre tem se tornado um benefício para a organização da informação na *web*, principalmente em projetos que publicam dados governamentais, instituições de memória e patrimônio cultural. Esses projetos oferecem mecanismos de acesso e compartilhamento de conjuntos de dados tratados sob o mesmo padrão, o que os tornam conectados em uma teia global de recursos na *web*.

Entre os instrumentos de representação da informação que aderiram as tecnologias da WS e do *Linked Data* estão os tesouros, elaborados para atender as demandas de informação, especificamente para a recuperação de conteúdo na *web*. Na próxima seção serão discutidos os modelos de dados utilizados para representar os tesouros na *web*.

### 2.3 TESAUROS NA *WEB*

Além de conduzir a indexação, os tesouros têm assumido a função de auxiliar na RI em ambientes digitais, tendo em vista que esta se configura em um ambiente de hipermídias e de grande volume de dados voláteis (BAEZA-YATES; RIBEIRO-NETO, 2013). Desse modo, instrumentos tradicionalmente utilizados para facilitar as buscas em SRI, como os tesouros, necessitam ser adaptados para o contexto tecnológico atual. Assim, para representá-los nesses espaços são empregados modelos de dados capazes de expressar formalmente sua estrutura de conceitos, tornando-os interoperáveis na *web*.

Diante disso, representá-los em um esquema formal permite que os relacionamentos conceituais expressos em sua estrutura rígida e textual se adequem aos padrões da WS. (W3C, 2005a; CATARINO, 2014; RAMALHO, 2015). A *Web* tem impactado o uso dos tesouros, historicamente eficazes em ambientes controlados, com estruturas claramente definidas e úteis para acessar informações, porém

[...] a heterogeneidade de formatos e estruturas, combinada com o enorme crescimento de recursos e conteúdo, torna impossível aplicar esse modelo de trabalho. Os caminhos de acesso que levam a elementos de informação não são delimitados ou estabelecidos com antecedência e a atualização constante do conteúdo torna praticamente impossível aplicar tesouros diretamente à *Web*. (PASTOR-SANCHEZ; MARTÍNEZ-MENDEZ; RODRIGUEZ-MUÑOZ, 2009, não paginado, tradução nossa).

Desde o início dos anos 2000, o W3C tem trabalhado no desenvolvimento e na implementação na área de publicação de vocabulários o uso de linguagem para especificar ontologias e integração de RDF com documentos HTML. Entre as iniciativas está a representação dos SOC como dados na *web*, visando sua integração com diferentes agentes, serviços e aplicativos (W3C, 2006; ZENG; MAYR, 2018).

Assim como a maioria dos projetos da WS, os grupos de trabalho do W3C recomendam o uso de linguagens e modelos para que os tesouros possam ser incorporados aos ambientes digitais. Inúmeros esforços para publicá-los na *web* foram registrados ao longo dos anos 1990 e 2000 pelo O *SWAD-Europe Thesaurus Activity* do W3C, destacando os formatos XML e RDF como os mais utilizados no período aludido. Mais adiante, observou-se iniciativas para representar os vocabulários na ontologia OWL.

### 2.3.1 Tesouros representados em XML, RDF e OWL

No princípio, percebe-se que a XML foi uma das tecnologias utilizadas para migrar a estrutura textual dos tesouros para a *web*. Lançada pelo W3C em 1998 – embora não tenha sido desenvolvida especificamente para a migração dos SOC – auxiliou no intercâmbio de diversos tipos de dados, não só na *web* como em outros ambientes digitais (W3C, 2016a), incluindo os dados bibliográficos. Os benefícios da linguagem foram incorporados pela *Library of Congress* na década 1990, quando produziram ferramentas de base *Standard Generalized Markup Language* (SGML)<sup>28</sup> para a conversão dos registros bibliográficos no formato MARC 21 (*Format for Bibliographic Data*) para a linguagem XML (LIBRARY OF CONGRESS, 2008).

As implementações em XML tinham como propósito fornecer uma estrutura representacional comum para a codificação de tesouros na *web* a partir de sintaxe uniforme. Os elementos de marcação utilizados expressavam tanto o conteúdo quanto a estrutura dos tesouros e, de modo igual, favoreciam a edição e manutenção eletrônica (LEE; BAILLIE; ORO, 1999).

Entre as implementações mais conhecidas está o esquema *Zthes* (*Z39.50 Profile for Thesaurus Navigation*), proposto por Mike Taylor em 1999 para a representação e navegação de tesouros na *web*, tendo como base o protocolo Z39.50. Na proposta, Taylor (2000) apresenta o tesouro eletrônico como um banco de dados de termos interligados, em que cada termo,

---

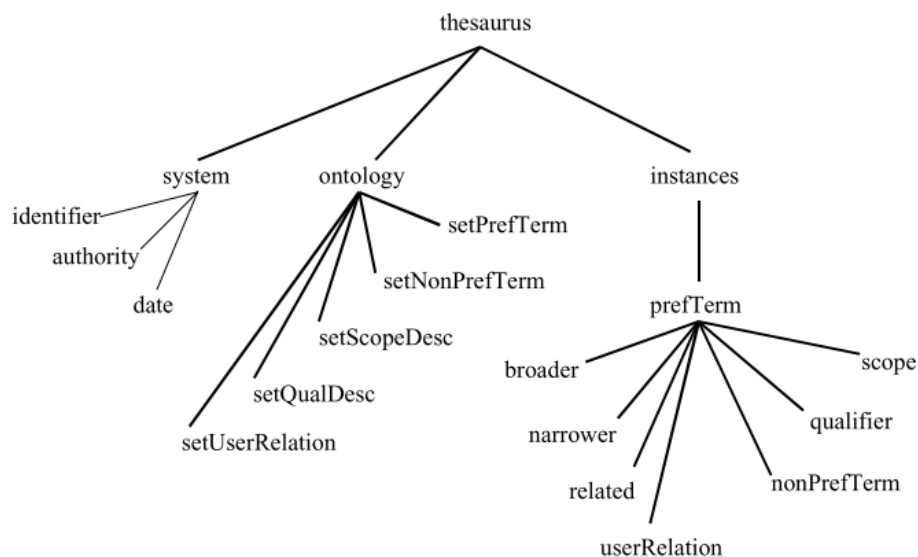
<sup>28</sup> Padrão não proprietário de código aberto aplicado para a troca de dados, permitindo que documentos inscritos nessa linguagem sejam transportados de um ambiente para o outro sem que informações se percam. É um padrão ISO (ISO 8879) que especifica as regras para a criação de linguagens de marcação independentemente da plataforma. Foi elaborada no final dos anos 60 pelo grupo de pesquisadores da IBM e aplicado inicialmente para a partilha de documentos legíveis por máquinas em projetos governamentais de grande dimensão e na indústria aeroespacial (BAEZA-YATES; RIBEIRO-NETO, 2013).

inclusive os não preferenciais, possui um registro individual. O esquema Zthes foi apresentado na linguagem XML e descreve os elementos do tesouro a partir de três elementos:

- a) **registros de nível superior**, em que são descritos os termos, seus qualificadores, notas de escopo, idioma, data de criação e modificação e o tipo de termo (termo preferido, não descritor, indicador de faceta);
- b) **sub-registros de lançamentos**, que descrevem os elementos do termo em um banco de dados de destino, pois se vários tesouros podem ser suportados por um único servidor estes devem ser apresentados como bancos de dados diferentes;
- c) **sub-registros de relacionamentos**, os quais especificam os tipos de relacionamentos entre os termos do tesouro (termo geral, termo específico, termo relacionado, use para e equivalente linguístico).

No mesmo ano de 1999, uma ontologia genérica para representar tesouros (*Generic Thesaural Ontology - GTO*) foi proposta por Lee, Baillie e Oro (1999) para expressar diferentes tesouros em uma linguagem comum e oferecer alto nível de abstração na descrição de suas classes e relacionamentos. A proposta inicial buscou contemplar as normas para a construção de tesouros monolíngues ISO 2788 e não pretendia cobrir o mapeamento de termos entre diferentes tesouros. Para representar a estrutura conceitual do GTO foi proposto um esquema chamado *Thesaural Markup Language (TML)*, usado para descrever os tipos de elementos de marcação permitidos e expressar o conteúdo e a ordenação dos tesouros a partir da linguagem XML. O gráfico de elementos TML é ilustrado na Figura 12.

**Figura 12** - Gráfico de elementos do *Thesaural Markup Language (TML)*



**Fonte:** Lee, Baillie e Oro (1999).



A estrutura gráfica TML for desenvolvida para reorganizar as classes e as relações semânticas do GTO em uma estrutura de dados orientada ao processo, refletindo como eles são dispostos. Os tesouros no esquema TML subordinam três tipos de elementos: sistema, ontologia e instâncias. O elemento *system* representa metadados sobre o tesouro, o elemento *ontology* caracteriza uma base específica do tesouro e o elemento *instances* estende o GTO até uma instância específica do tesouro simbolizado pelo suplemento *prefTerm*, que aponta os relacionamentos entre os termos.

**Figura 13** - Exemplo de marcação utilizando os elementos TML

```

<system>
  <identifier version="CSIRO KeywordAAA 0.1" language="English"/>
  <authority createdBy="CMIS OMT Project"/>
  <date modifiedDate="981022"/>
</system>
<ontology>
  <setPrefTerm      type="KW"      name="Keyword"/>
  <setPrefTerm      type="AD"      name="Activity Descriptor"/>
  <setPrefTerm      type="SD"      name="Subject Descriptor"/>
  <setNonPrefTerm   type="PA"      name="Permitted Acronym"/>
  <setNonPrefTerm   type="FB"      name="Forbidden Term"/>
  <setScopeDesc    type="SN"      name="Supplementary Note"/>
  <setUserRel       type="TOP"     name="Top"/>
</ontology>
<instances>
  <prefTerm         type="KW"      value="Occupational Health & Safety">
    <narrower       type="AD"      value="Policy"/>
    <nonPrefTerm    type="PA"      value="OHS"/>
    <nonPrefTerm    type="FB"      value="OHS&S"/>
  </prefTerm>
  <prefTerm         type="AD"      value="Policy">
    <broader        type="KW"      value="Occupational Health & Safety"/>
    <broader        type="KW"      value="EEQ"/>
    <broader        type="KW"      value="Establishment"/>
    <narrower       type="SD"      value="Access & Equity"/>
    <narrower       type="SD"      value="Code of Ethics"/>
    <narrower       type="SD"      value="Copyright"/>
    <related        type="AD"      value="Employment Condition"/>
  </prefTerm>
</instances>

```

**Fonte:** Lee, Baillie e Oro (1999).

Outras iniciativas usando a linguagem XML se apresentaram no início dos anos 2000 como: o MeSH *in* XML, para representar os dados do *Medical Subject Headings* (MeSH); GEMET XML format, para representar o *General Multilingual Environmental Thesaurus*



(GEMET); o ADL *Thesaurus Protocol XML formats*, desenvolvido pela *Alexandria Digital Library*; os elementos propostos por Soergel (2002) para especificar os dados dos tesouros em XML; e o próprio MARC 21 XML, criado pela LC a fim de representar os registros bibliográficos, até os arquivos de autoridades em esquema padrão e interoperável para a *web*.

Apesar dos ganhos com a adoção da XML para a publicação e compartilhamento dos tesouros pode-se encontrar dificuldades em expressar formalmente sua estrutura, pois “[...] a *Web* requer uma visão mais ampla que vai além da ideia de um termo como elemento central do tesouro, expressos em número, tipo e significado dos relacionamentos existentes.” (PASTOR-SANCHEZ; MARTÍNEZ-MENDEZ; RODRIGUEZ-MUÑOZ, 2009, não paginado, tradução nossa).

Nessa oportunidade, o modelo RDF pode oferecer uma distinção clara entre significado, relações conceituais e informação sobre os termos. Enquanto as representações em XML dos tesouros estão focadas em expressar os relacionamentos entre os termos, o esquema RDF introduz a noção de conceitos e instâncias. Um conceito descreve uma entidade em um nível abstrato com propriedades genéricas, enquanto que uma instância é uma representação real desse conceito com valores específicos para propriedades.

Um dos primeiros projetos para modelar tesouros no esquema RDF foi o CERES/NBII *Thesaurus Partnership Project*, criado pelo *California Environmental Resources Evaluation System* (CERES) em cooperação com o *United States Geological Survey Biological Resources Division* (USGS/BRD). O CERES era composto por classes e subclasses que especificavam o termo e por propriedades que definiam os tipos de relacionamentos existentes entre eles. Ainda que utilizando o modelo RDF, o CERES pode ser apontado como um esquema baseado em termos, visto que as propriedades eram aplicadas para orientar o uso de um termo preferido em relação aos demais da estrutura do tesouro. Igualmente, um esboço baseado em conceitos considera os termos como unidades independentes, apesar de também indicar seus relacionamentos (KLIE, 2003).

Para a orientação ontológica no esquema RDF, os conceitos devem ser ordenados em uma hierarquia de subconceito/ superconceito, em que a parte superior da hierarquia é fornecida por um nó raiz. Assim, um subconceito herda todas as propriedades e relacionamentos de seu superconceito. Desse modo, uma ontologia RDF oferece princípios de modelagem consideravelmente mais poderosos que os esquemas baseado em termos como: a noção de conceitos versus instâncias; a separação de entidades e termos; a especificação de conceitos por propriedades e atributos; a capacidade de criar propriedades infinitas; mais específicas e semanticamente mais ricas; e a noção de paradigma da herança (LAUSER, 2004).

Outras implementações aplicando o modelo RDF foram desenvolvidas ao longo do período. Os formatos e os padrões mais comuns entre os anos de 1999 e 2004 – fase anterior as implementações do *Linked Data* para os SOC – são apresentados no Quadro 5, no qual são sumarizadas as principais iniciativas para representar, intercambiar, especificar e promover a navegação *web* de tesouros.

**Quadro 5** - Formatos e iniciativas para representação dos tesouros na *web*

Formatos	Iniciativas e datas
RDF	<p><i>TIF RDF Thesaurus Interchange Format, SWAD-E Project</i> (2003).  <i>ILRT RDF thesaurus draft specification</i> (2001).  <i>Limber Project RDF schema for ISO compliant multi-lingual thesauri</i> (2001).  <i>CERES/NBII Project RDF thesaurus descriptor standard</i> (2000).  <i>DRC DAML+OIL ontology for the CALL thesaurus</i> (2002).  <i>ETB RDF schema for the multilingual educational thesaurus version 0.4</i> (2001).  <i>GEM Consortium RDF schema for monolingual thesauri</i> (2002).  <i>Agrovoc/Kaon RDF ontology/thesaurus schema</i> (2001).  <i>Wordnet RDF schema by Sergey Melnik</i>.</p>
XML	<p><i>MARC-21 XMLSchema</i>.  <i>Zthes Z39.50 profile for thesaurus navigation</i> (2001).  <i>TML thesaurus markup language</i> (1999).  <i>ADL Thesaurus Protocol XML formats</i> (2002).  <i>MeSH XML format</i> (2001).  <i>GEMET XML format</i> (2003).  <i>APAIS XML thesaurus format, an extension of Zthes</i> (2000).  <i>Open University thesaurus schemas</i> (2002).  <i>Soergel XML thesaurus specification</i> (2001).</p>
Formatos relacionados	<p><i>OWL Web Ontology Language</i> (2003).  <i>XTM XML Topic Maps 1.0</i> (2001).  <i>XFML eXchangeable Faceted Metadata Language</i> (2002).  <i>MARC 21 format for authority data</i> (2002).  <i>MARC 21 format for classification data</i> (2002).  <i>MARTIF terminology interchange format</i> (1999).  <i>bcXML taxonomy format</i> (2003).  <i>KIF Knowledge Interchange Format</i> (1998).  <i>OML Ontology Markup Language version 0.3</i> (1999).  <i>CKML Conceptual Knowledge Markup Language</i> (1999).  <i>DAML Darpa Agent Markup Language</i> (2001).  <i>OIL Ontology Inference Layer</i> (2000).  <i>RDF Resource Description Framework</i> (1999).  <i>VocML Vocabulary Markup Language</i> (2001).  <i>GlossXML XML format for glossaries</i>.</p>

**Fonte:** W3C (2004).

Os tesouros podem estar armazenados em banco de dados relacionais ou até representados na forma de cabeçalhos de assuntos em bases bibliográficas, em que a saída padrão do sistema de gerenciamento pode ser em um formato de texto XML ou estruturado (W3C, 2005b; HARPER, 2006; SUMMERS et al., 2008). Quando vinculadas as bases bibliográficas estas dispõem de padrões de metadados específicos para modelar os Registros de Autoridades<sup>29</sup> e podem ser expressas tanto em linguagem XML quanto em vocabulários RDF. Entre os modelos de dados mais comuns utilizados atualmente para representar registros bibliográficos estão o MARC 21 *Format for Authority Data*, o *Metadata Authority Description* (MADS), o RDA *Elements Sets* e o *Bibliographic Framework* (BIBFRAME).

**Quadro 6 – Padrões para a representação de registros bibliográficos**

<b>Formatos de registros de autoridade</b>	<b>Descrição</b>
MARC 21 <i>Format for Authority Data</i>	Integra o conjunto de formatos MARC 21 <i>Format for Bibliographic Data</i> . Foi projetado para fornecer informações sobre as formas autorizadas de nome, assuntos e subdivisões de assunto. Apesar da possibilidade de conversão de registros MARC 21 em XML, o formato apresentou limitações quanto aos registros dos dados de autoridades, fazendo surgir o <i>Metadata Object Description Schema</i> (MODS) e o <i>Metadata Authority Description</i> (MADS).
<i>Metadata Authority Description</i> (MADS)	Esquema de metadados para dados de autoridade que tem como base em sua origem o XML, mais especificamente definido para representar as complexidades dos dados de autoridade, de vocabulários e de tesouros. Foi publicado pela LC em 2005 como acompanhamento para o MODS.
RDA <i>Element Sets</i>	Conjunto de elementos do <i>Resource Description and Access</i> (RDA) que apresenta vocabulários de classes, propriedades e valores para auxiliar no registro de atributos e relacionamentos de entidades do domínio bibliográfico como: obra, expressão, manifestação, item, pessoa, família, entidade coletiva, conceito, objeto, evento e lugar.
<i>Bibliographic Framework</i> BIBFRAME	Modelo de dados feito pelo LC para a descrição bibliográfica no âmbito dos dados conectados tentando tornar os dados bibliográficos amplamente utilizados não só pela comunidade bibliotecária, mas também por outras aplicações na <i>web</i> . No BIBFRAME, a descrição de um recurso apresenta elementos de informação que são organizados em três níveis de abstração: Obra, Instancia e item. O vocabulário do BIBFRAME é formado por classes e propriedades RDF.

**Fonte:** Elaborado pelo autor (2021) a partir de (LIBRARY OF CONGRESS, 1999, 2008, 2016, 2018; ASSUMPÇÃO; SANTOS, 2013; RDA STEERING COMMITTEE, 2016).

<sup>29</sup> Do inglês *Authority record*, é considerado “um conjunto de elementos de dados que identifica uma entidade e pode ser utilizado para facilitar o acesso ao ponto de acesso autorizado para uma entidade [...]” representada por um nome, um assunto ou subdivisões de assuntos utilizados nos registros bibliográficos (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2009, p. 9, tradução nossa) Contextualizando, o tesouro é um tipo de vocabulário controlado que pode ser expresso por meio de registros de autoridades, principalmente em bases de dados bibliográficas.

Os registros bibliográficos têm procurado contemplar formatos alternativos visando sua interoperabilidade na *web*. Por outro lado, os vocabulários RDF têm ganhado força nessa área devido a necessidade de representação desses recursos para as tecnologias da WS, em especial para o contexto do *Linked Data*. Nessa oportunidade, as ontologias OWL foram utilizadas para expandir a capacidade dos relacionamentos em tesouros, tornando-os mais adaptáveis e reutilizáveis em diferentes domínios. Essas ontologias são capazes de formalizar sua estrutura em um conjunto de axiomas e têm maior capacidade expressiva de modelagem de sistemas de informação (GARCÍA-TORRES; PAREJA-LORA; PRADANA-LÓPEZ, 2008).

Corroborando com essa assertiva, o W3C (2009a, não paginado, tradução nossa) entende que

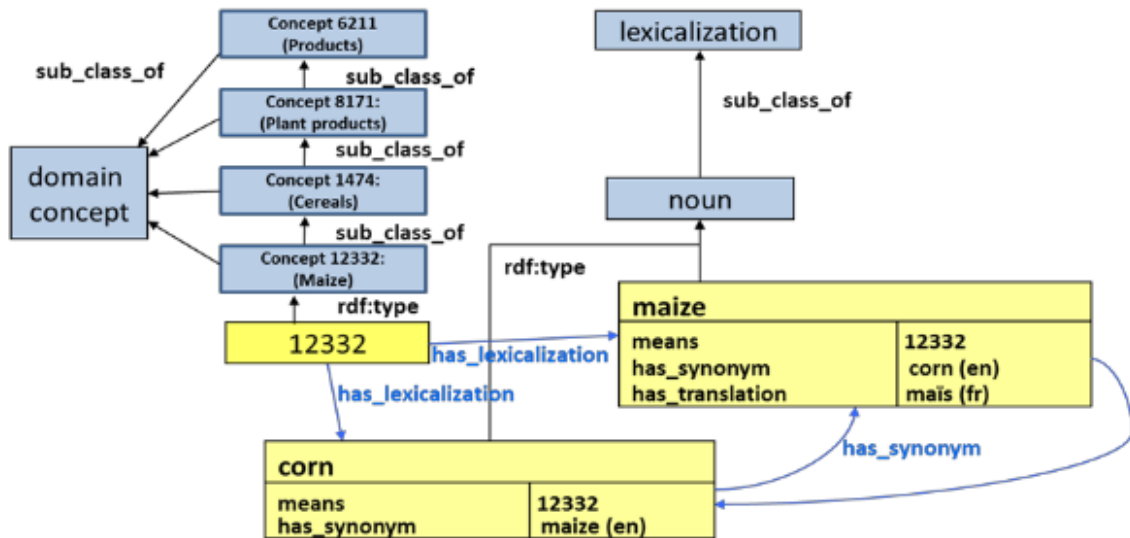
[...] pode-se obter muito dos tesouros se utilizado em conjunto com uma linguagem OWL (*Ontology Web Language*), visto que a conversão desses SOC para uma representação baseada em lógica formal pode, na prática, contribuir com o desenvolvimento de aplicação mais avançadas para expressar e trocar conhecimentos sobre um domínio.

Tesouros e outros tipos de vocabulários controlados são cada vez mais reprojatados em ontologias descritas usando a OWL (KLESS et al., 2014). Não obstante, Matthews, Milles e Wilson (2009) concordam que projetar o tesouro na WS envolve necessariamente a permeabilidade com outras propostas, como ontologias e mapas de conhecimento, ou seu uso combinado para estabelecer uma mudança de paradigma na criação de tesouros, enfocando os tesouros conceituais.

Como exemplo dessas iniciativas destaca-se o AGROVOC – um tesouro tradicional criado e mantido pela Organização das Nações Unidas para Agricultura e Alimentação (FAO) – que em 2004 serviu como um estudo de caso para explorar a reengenharia de um tesouro tradicional em uma ontologia (SOERGEL et al., 2004). Esse tesouro tradicional buscou modelar seu vocabulário a partir da linguagem OWL, escolhida devido a possibilidade de especificar um domínio a partir de objetos e classes de objetos.

Os conceitos do tesouro foram organizados em uma hierarquia e representados a partir das propriedades *rdfs:subClassof*, em que cada conceito pode ser caracterizado por uma classe e sua instância única associada, preenchida com valores de propriedade, como ilustra a Figura 14. No entanto, a OWL foi considerada rígida para modelar os recursos do tesouro e, ao mesmo tempo, simplista para afeiçoar recursos multilíngues, além de prejudicar a visualização hierárquica original dos termos pelos editores (CARACCILO et al., 2011).

**Figura 14 - Modelagem do AGROVOC baseado no OWL**



Fonte: Caracciolo et al. (2011).

Apesar dos desafios encontrados na modelagem do AGROVOC, o uso da OWL para expressar e enriquecer os relacionamentos em tesouros é uma realidade especialmente observada em domínios específicos, para os quais há redes de vocabulários semânticos. Um exemplo disso é o *National Cancer Institute Thesaurus* (NCIt), que contém 34.000 termos de referência no domínio do câncer, foi elaborado como uma ontologia OWL, com axiomas anatômicos completos vinculados a conceitos de doenças de vários domínios relacionados à pesquisa do câncer, e tem como um de seus objetivos aumentar a integração do tesouro às ontologias da *Open Biomedical Ontologies* (OBO)<sup>30</sup>, fornecendo uma maneira de integrar ou vincular esses tipos de informações mediante relacionamentos semânticos (THE OBO FOUNDRY, 2019).

O tesouro publicado em LOD segue as diretrizes do W3C para os projetos da WS e incorpora tecnologias e padrões que interligam os dados do tesouro com outros recursos na *web*. Desde sua proposta inicial até os dias atuais, o LOD tem admitido a permeabilidade dos diferentes tipos de SOC na *web* de dados, inclusive os tesouros. Diante das limitações encontradas nos modelos anteriores, o W3C elaborou o modelo *Simple Knowledge Organization System* (SKOS) para a representação dos SOC na *web* de dados conectados.

<sup>30</sup> Família de ontologias interoperáveis para o uso compartilhado em diferentes domínios biológicos e médicos. Faz parte dos recursos do Centro Nacional de Ontologia Biomédica dos EUA (THE OBO FOUNDRY, 2019).

### 2.3.2 Tesouros representados em SKOS

O modelo de dados SKOS foi desenvolvido pelo W3C com o objetivo de oferecer suporte efetivo às ações de organização da informação na *web*, permitindo que os tradicionais SOC possam ser expressos e compartilhados como dados legíveis por computador, tornando-os interoperáveis com outras aplicações na *web* (W3C, 2009a). Esse modelo começou a ser elaborado em 2002 pelo grupo de trabalho *Semantic Web Advanced Development for Europe* (SWAD-E) e, em 2005, foi difundido por meio da publicação do *SKOS Core Guide*. Porém, foi apenas em 2009 que ele veio a se tornar uma recomendação oficial do W3C (PASTOR-SANCHEZ; MARTÍNEZ-MENDEZ; RODRIGUEZ-MUÑOZ, 2009).

O vocabulário SKOS é formado por um conjunto de propriedades utilizadas para expressar o conteúdo e a estrutura de um esquema de conceitos, admitindo a representação de um conceito enquanto unidade de pensamento, bem como apresenta uma distinção clara entre rótulos lexicais preferenciais, alternativos e ocultos (W3C, 2005a; NOVIANTI; NUGROHO, 2014). O Quadro 7 mostra as principais características utilizadas para significar os SOC conforme o *SKOS Core Vocabulary*, recomendação oficial do W3C.

**Quadro 7** - Principais propriedades do *SKOS Core Vocabulary*

<b>Propriedades</b>	<b>Descrição</b>
<i>skos:prefLabel</i>	Termo Preferido ou Termo Adotado. O termo lexical preferido para um recurso em uma dada linguagem.
<i>skos:altLabel</i>	Termo Alternativo: Acrônimos, abreviações, variações de ortografia e formas de plural/singular irregulares podem ser incluídas entre os termos alternativos para um conceito
<i>skos:hiddenLabel</i>	Termo oculto. Um termo lexical para um recurso que deve ser ocultado na geração de apresentações visuais do recurso, mas deve ser acessível para operações de busca livre em texto. Erros de ortografia e digitação são normalmente incluídos como termos ocultos.
<i>skos:broader</i>	Termo Genérico. Um conceito de significação mais genérica.
<i>skos:narrower</i>	Termo Específico. Um conceito de significação mais específica.
<i>skos:related</i>	Termo Relacionado. Um conceito com o qual existe um relacionamento semântico associativo.
<i>skos:scopeNote</i>	Nota de Escopo. Uma nota que ajuda a clarificar o significado de um conceito.
<i>skos:definition</i>	Definição. Explicação formal do significado de um conceito.
<i>skos:example</i>	Exemplo. Um exemplo de utilização de um conceito.
<i>skos:historyNote</i>	Nota Histórica. Uma nota sobre a significação ou a utilização passada de um conceito.

**Fonte:** W3C (2005a).

A codificação do SKOS também é baseada no modelo RDF e pode ser expressa em diferentes sintaxes, como *Turtle*, *N-Triples*, *Notation 3* e a tradicional linguagem XML como sintaxe para o intercâmbio e processamento de metadados. Esse modelo contribui com a interoperabilidade entre os vocabulários semânticos encontrados na *web*, visto ser capaz de expressar o relacionamento de um objeto com suas propriedades e valores. Uma vez identificados por URI, os conceitos do tesouro podem ser evidentes mediante modelo de dados SKOS encapsulados em um documento RDF.

A Figura 15 mostra o fragmento da representação textual do conceito “*Musical Styles*” no UNESCO Thesaurus, em que os termos são estruturados conforme os tipos de relacionamento. A partir de um conceito preferido (*Preferred Term*), de maior abrangência, são apresentadas suas subdivisões em conceitos específicos (*Narrower Concepts*) e as associações com outros conceitos, em conceitos relacionados (*Related concepts*).

**Figura 15** - Fragmento da representação textual no UNESCO Thesaurus

PREFERRED TERM	<b>Musical styles</b> <a href="#">Search in UNESDOC</a>						
<hr style="border-top: 1px dashed #000;"/>							
NARROWER CONCEPTS	<a href="#">Classical music</a> <a href="#">Concrete music</a> <a href="#">Contemporary music</a> <a href="#">Jazz</a> <a href="#">Pop music</a> <a href="#">Popular music</a> <a href="#">Religious music</a> <a href="#">Traditional music</a>						
RELATED CONCEPTS	<a href="#">Musicology</a>						
BELONGS TO GROUP	<a href="#">Culture &gt; Performing arts</a>						
IN OTHER LANGUAGES	<table border="0"> <tr> <td><a href="#">Style musical</a></td> <td>French</td> </tr> <tr> <td><a href="#">Музыкальные стили</a></td> <td>Russian</td> </tr> <tr> <td><a href="#">Estilo musical</a></td> <td>Spanish</td> </tr> </table>	<a href="#">Style musical</a>	French	<a href="#">Музыкальные стили</a>	Russian	<a href="#">Estilo musical</a>	Spanish
<a href="#">Style musical</a>	French						
<a href="#">Музыкальные стили</a>	Russian						
<a href="#">Estilo musical</a>	Spanish						
URI	<a href="http://vocabularies.unesco.org/thesaurus/concept354">http://vocabularies.unesco.org/thesaurus/concept354</a>						
Download this concept:	<a href="#">RDF/XML</a> <a href="#">TURTLE</a> <a href="#">JSON-LD</a> <span style="float: right;">Last modified 5/23/06</span>						

**Fonte:** UNESCO Thesaurus (2019).

A Figura 16 exhibe o extrato da representação textual do conceito “*Classical music*” do UNESCO Thesaurus em RDF/XML. O tesouro utiliza as peculiaridades do vocabulário SKOS para mostrar os relacionamentos existentes entre o recurso principal e o termo “*Classical music*” com os demais termos do tesouro, também identificados por URI.

Figura 16 - Representação do SKOS em RDF/XML

```

-<rdf:RDF>
  -<skos:ConceptScheme rdf:about="http://vocabularies.unesco.org/thesaurus">
    <skos:prefLabel xml:lang="en">UNESCO Thesaurus</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Thésaurus de l'UNESCO</skos:prefLabel>
    <skos:prefLabel xml:lang="ru">Тезаурус ЮНЕСКО</skos:prefLabel>
    <skos:prefLabel xml:lang="es">Tesouro de la UNESCO</skos:prefLabel>
  </skos:ConceptScheme>
  -<skos:Concept rdf:about="http://vocabularies.unesco.org/thesaurus/concept5394">
    <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2006-05-23T00:00:00</dc:modified>
    <skos:inScheme rdf:resource="http://vocabularies.unesco.org/thesaurus"/>
    <skos:prefLabel xml:lang="ru">Классическая музыка</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Musique classique</skos:prefLabel>
    <skos:prefLabel xml:lang="en">Classical music</skos:prefLabel>
    <skos:prefLabel xml:lang="es">Música clásica</skos:prefLabel>
    <skos:hiddenLabel xml:lang="es">Musica clasica</skos:hiddenLabel>
    -<skos:broader>
      -<skos:Concept rdf:about="http://vocabularies.unesco.org/thesaurus/concept354">
        <skos:prefLabel xml:lang="es">Estilo musical</skos:prefLabel>
        <skos:prefLabel xml:lang="en">Musical styles</skos:prefLabel>
        <skos:prefLabel xml:lang="ru">Музыкальные стили</skos:prefLabel>
        <skos:prefLabel xml:lang="fr">Style musical</skos:prefLabel>
        <skos:narrower rdf:resource="http://vocabularies.unesco.org/thesaurus/concept5394"/>
      </skos:Concept>
    </skos:broader>
    -<skos:related>
      -<skos:Concept rdf:about="http://vocabularies.unesco.org/thesaurus/concept5396">
        <skos:prefLabel xml:lang="ru">Опера</skos:prefLabel>
        <skos:prefLabel xml:lang="es">Ópera</skos:prefLabel>
        <skos:prefLabel xml:lang="en">Opera</skos:prefLabel>
        <skos:prefLabel xml:lang="fr">Opéra</skos:prefLabel>
        <skos:related rdf:resource="http://vocabularies.unesco.org/thesaurus/concept5394"/>
      </skos:Concept>
    </skos:related>
  </skos:Concept>
  -<isothes:ConceptGroup rdf:about="http://vocabularies.unesco.org/thesaurus/domain3">
    <rdf:type rdf:resource="http://vocabularies.unesco.org/ontology#Domain"/>
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Collection"/>
    <skos:prefLabel xml:lang="en">Culture</skos:prefLabel>
    <skos:prefLabel xml:lang="fr">Culture</skos:prefLabel>
    <skos:prefLabel xml:lang="ru">Культура</skos:prefLabel>
    <skos:prefLabel xml:lang="es">Cultura</skos:prefLabel>
    -<skos:member>
      -<isothes:ConceptGroup rdf:about="http://vocabularies.unesco.org/thesaurus/mt3.55">
        <rdf:type rdf:resource="http://vocabularies.unesco.org/ontology#MicroThesaurus"/>
        <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Collection"/>
        <skos:prefLabel xml:lang="es">Artes escénicas</skos:prefLabel>
        <skos:prefLabel xml:lang="ru">Исполнительские виды искусства</skos:prefLabel>
        <skos:prefLabel xml:lang="fr">Arts du spectacle</skos:prefLabel>
        <skos:prefLabel xml:lang="en">Performing arts</skos:prefLabel>
        <skos:member rdf:resource="http://vocabularies.unesco.org/thesaurus/concept5394"/>
      </isothes:ConceptGroup>
    </skos:member>
  </isothes:ConceptGroup>
</rdf:RDF>

```

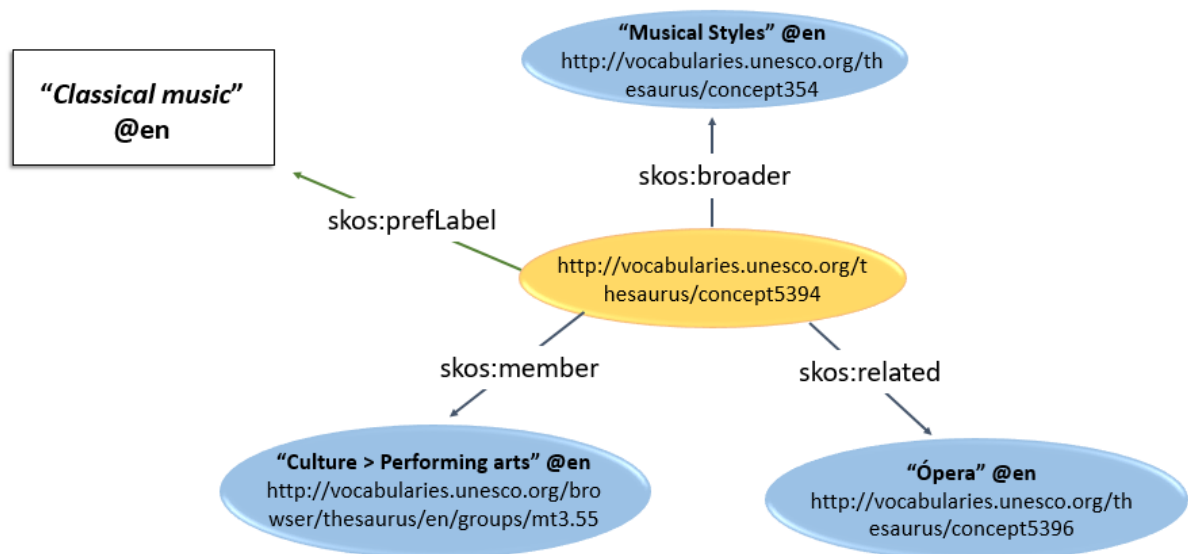
Fonte: UNESCO Thesaurus (2019).

No fragmento acima, o RDF também pode ser usado para expressar metapropriedades de um tesouro que não são contempladas pelo vocabulário SKOS, como o título da página, descrição, data de modificação e direitos de reprodução. Por conseguinte, os termos de metadados Dublin Core são aplicados como vocabulários de suporte, pois incluem várias domínios complementares para a publicação do esquema conceitual na *web* (W3C, 2005b; HARPER, 2006).



Para melhor visualizar os relacionamentos do termo “*Classical music*” com os demais termos do tesauro a Figura 17 ilustra a representação gráfica do RDF, em que é demonstrado como os termos são ligados a partir das classes e propriedades do vocabulário SKOS. O conceito “*Classical music*” é simbolizado pela propriedade *skos:prefLabel*, ligado ao conceito mais geral (*Broader Concept*) “*Musical Styles*” e ao conceito relacionado (*Related Concept*) “*Ópera*”. É possível definir o grupo ou o domínio mais abrangente que o conceito principal está ligado (*Belongs to Group*) atendendo à propriedade *skos:member*.

**Figura 17** - Extrato do UNESCO Thesaurus representado em RDF *Graph*



**Fonte:** Elaborado pelo autor (2021).

Como os conceitos são identificados por URI e o conjunto de dados declarados em RDF, é possível vinculá-los ou combiná-los com outros documentos em RDF. Com isso, a adoção desses modelos de dados permite que a representação textual de um tesauro seja publicada como LOD, assim como os termos ligados a outros tesouros na *web* igualmente publicados no modelo SKOS.

O *Library of Congress Subject Headings* (LCSH) – importante vocabulário controlado para a indexação de assuntos que conta com termos de diversos domínios – é apresentado como um exemplo de tesauro que utiliza o modelo SKOS e publica seus dados como LOD. A Figura 18 mostra a URI do termo “*Jazz*” representado no LCSH<sup>31</sup>, que por meio da propriedade *skos:closeMatch* se vincula à URI do mesmo termo disponível no serviço finlandês para a publicação e utilização de vocabulários, ontologias e classificações, denominado Finto<sup>32</sup>.

<sup>31</sup> <http://id.loc.gov/authorities/subjects/sh85069833>

<sup>32</sup> <https://finto.fi/yso/en/page/p4484>

Igualmente, os termos identificados por URI são interligados com outros recursos semelhantes na *web*.

**Figura 18 - Conceitos publicados como LOD**

The screenshot displays the Finto web interface for the concept 'Jazz'. At the top, there are two tabs: 'URI do conceito Jazz no LCSH' and 'URI do conceito Jazz no Finto (Finlândia)'. The main content area shows the 'Jazz' concept within the 'YSO - General Finnish ontology'. The interface includes a search bar, navigation tabs (A-Z, Hierarchy, Groups, New), and a detailed view of the 'Jazz' concept. The detailed view shows the concept's type as 'General concept', its broader concept as 'Afro-American music', and its narrower concepts as 'acid jazz', 'bebop', 'free jazz', 'fusion music', 'hard bop', 'mainstream jazz', 'swing (music)', 'third stream (jazz)', and 'traditional jazz'. Related concepts include 'barrelhouse', 'blues (Afro-American music)', 'boogie woogie', and 'Finnish accordion jazz'.

**Fonte:** Elaborado pelo autor (2021) a partir de LCSH (2019) e Finto (2019).

O conjunto de dados RDF pode ser liberado em um arquivo único de endereço *web* para que desenvolvedores possam vincular seus dados às URI de conceitos. Assim, Summers et al. (2008) destacam que devido ao volume de dados um serviço de Endpoint SPARQL permite que os usuários recuperem conceitos de forma programada, sem ser necessário baixar e indexar esse conjunto por completo para depois explorá-lo. Dessa forma, um serviço SPARQL torna a pesquisa mais prática, pois sua linguagem fornece um conjunto de operações de consulta analítica capaz de unir todos os triplos com um assunto correspondente.

Desde o seu lançamento, inúmeros tesouros na *web* aderiram ao modelo SKOS devido a capacidade de superar as limitações dos padrões utilizados anteriormente. O Quadro 8 lança as principais iniciativas identificadas na literatura, sendo a maioria ligada aos setores governamentais e de patrimônio cultural, em especial centros de pesquisa, arquivos, museus e bibliotecas. Em geral, elas são criadas por países da América do Norte e Europa e se concentram nos domínios da Arte, Arquitetura, Economia, Meio ambiente, Agricultura e Alimentos e Geociências.

**Quadro 8** - Principais tesouros publicados em LOD que utilizam o SKOS

<b>País/Região</b>	<b>Sector ou domínio</b>	<b>Tesouros</b>
Alemanha	Patrimônio Cultural: Bibliotecas	DE - German National Library <a href="http://dnb.de/EN/lds">http://dnb.de/EN/lds</a>
Alemanha	Governamental: Economia	STW Thesaurus for Economics <a href="http://zbw.eu/stw/versions/latest/about">http://zbw.eu/stw/versions/latest/about</a>
Alemanha	Centro de Pesquisa: Ciência Sociais	Instituto GESIS - Leibniz de Ciências Sociais Thesaurus Sozialwissenschaften (TheSoz) <a href="http://lod.gesis.org/thesoz/de.html">http://lod.gesis.org/thesoz/de.html</a>
Áustria	Governamental: Geociências	AT - Austrian Geological Survey (GBA) <a href="http://www.geologie.ac.at/services/thesaurus/">http://www.geologie.ac.at/services/thesaurus/</a>
Bélgica	Patrimônio Cultural: Bibliotecas	BE - Vlaams Theater Instituut – Travelogue <a href="http://vti.be/nl/linked-data">http://vti.be/nl/linked-data</a>
Dinamarca	Governamental: Agricultura, Recursos Naturais, Terra e meio ambiente	The Cadastre and Land Administration Thesaurus <a href="http://cadastralvocabulary.org/">http://cadastralvocabulary.org/</a>
Estados Unidos	Centro de Pesquisa: Arte e Arquitetura	The Art & Architecture Thesaurus (AAT) <a href="http://www.getty.edu/research/tools/vocabularies/aat/">http://www.getty.edu/research/tools/vocabularies/aat/</a>
Estados Unidos	Centro de Pesquisa: Terra, História e Arqueologia	The Getty Thesaurus of Geographic Names (TGN) <a href="http://www.getty.edu/research/tools/vocabularies/tgn/">http://www.getty.edu/research/tools/vocabularies/tgn/</a>
Estados Unidos	Governamental/ Patrimônio Cultural: Agricultura e alimentação	Thesaurus of the US National Agricultural Library <a href="https://agclass.nal.usda.gov/agt/agt.shtml">https://agclass.nal.usda.gov/agt/agt.shtml</a>
Estados Unidos	Patrimônio Cultural: Bibliotecas	US - Library of Congress: <a href="http://id.loc.gov/">http://id.loc.gov/</a>
Finlândia	Geral	Finto – Finnish Thesaurus and Ontology Service <a href="https://finto.fi/en/about">https://finto.fi/en/about</a>
França	Patrimônio Cultural: Bibliotecas	Bibliothèque Nationale de France - Répertoire d'autorité-matière encyclopédique et alphabétique unifié : <a href="http://rameau.bnf.fr/">http://rameau.bnf.fr/</a>
Nações Unidas	Governamental: Agricultura e alimentação	UN - Food and Agriculture Organisation of the United Nations (FAO) - AGROVOC thesaurus <a href="http://aims.fao.org/standards/agrovoc/about">http://aims.fao.org/standards/agrovoc/about</a>
Reino Unido	Governamental/ Patrimônio Cultural: Arquivos	UK - National Archives - UKAT Archival Thesaurus <a href="https://ukat.aim25.com/">https://ukat.aim25.com/</a>
União Europeia	Geral	EuroVoc: Multilingual Thesaurus of the European Union <a href="https://publications.europa.eu/en/web/eu-vocabularies">https://publications.europa.eu/en/web/eu-vocabularies</a>
União Europeia	Patrimônio Cultural: Museus	MIMO - Musical Instruments Museums Online <a href="http://www.mimo-db.eu/InstrumentsKeywords/">http://www.mimo-db.eu/InstrumentsKeywords/</a>
União Europeia	Governamental: Meio ambiente	General Multilingual Environmental Thesaurus <a href="https://www.eionet.europa.eu/gemet/en/themes/">https://www.eionet.europa.eu/gemet/en/themes/</a>

**Fonte:** Elaborado pelo autor (2021).

As experiências de aplicação do SKOS levaram os desenvolvedores de tesouros a perceberem algumas lacunas no modelo de dados. À vista disso, foram propostas novas ontologias para estender a capacidade de representação, tanto no que concerne à estrutura semântica – à medida que foi possível estabelecer relacionamentos mais refinados – como no tocante à capacidade de manipulação e processamento de arquivos RDF.

### 2.3.3 Extensões do vocabulário SKOS

O SKOS é uma recomendação oficial do W3C e a ontologia mais usada para representar sistemas interoperáveis de organização do conhecimento na WS. Entretanto, a versão inicial foi elaborada considerando os padrões ISO-5964:1985 e ISO-2788:1986. Assim, na tentativa de oferecer soluções às limitações do SKOS Core original algumas extensões do modelo foram desenvolvidas ao longo do tempo, como o SKOS *eXtension for Labels* (SKOS-XL) e, mais recentemente, o ISO-THES e o UNESKOS.

#### a) *SKOS eXtension for Labels* (SKOS-XL)

O SKOS-XL foi lançado em 2009 com o propósito de fornecer suporte adicional para enriquecer a modelagem de entidades lexicais. Isso se aplica nos casos em que os recursos de rótulo lexical apresentam uma forma literal única, porém podem representar rótulos diferentes nos esquemas conceituais. Em outras palavras, o SKOS-XL permite que cada entidade lexical diferente seja configurada como um recurso.

No SKOS-XL, uma classe especial de entidades lexicais, chamada *skosxl:Label*, fornecida e cada instância desta classe dispõe uma única forma literal simples. Com a extensão é possível definir três propriedades de rotulagem para caracterizar conceitos com instâncias de *skosxl:Label*, quais sejam: *skosxl:prefLabel*, *skosxl:altLabel* e *skosxl:hiddenLabel*. Essas propriedades apresentam funções semelhantes às propriedades *skos:prefLabel*, *skos:altLabel* e *skos:hiddenLabel*, definidas no modelo SKOS convencional (W3C, 2009).

**Quadro 9** - O vocabulário SKOS-XL

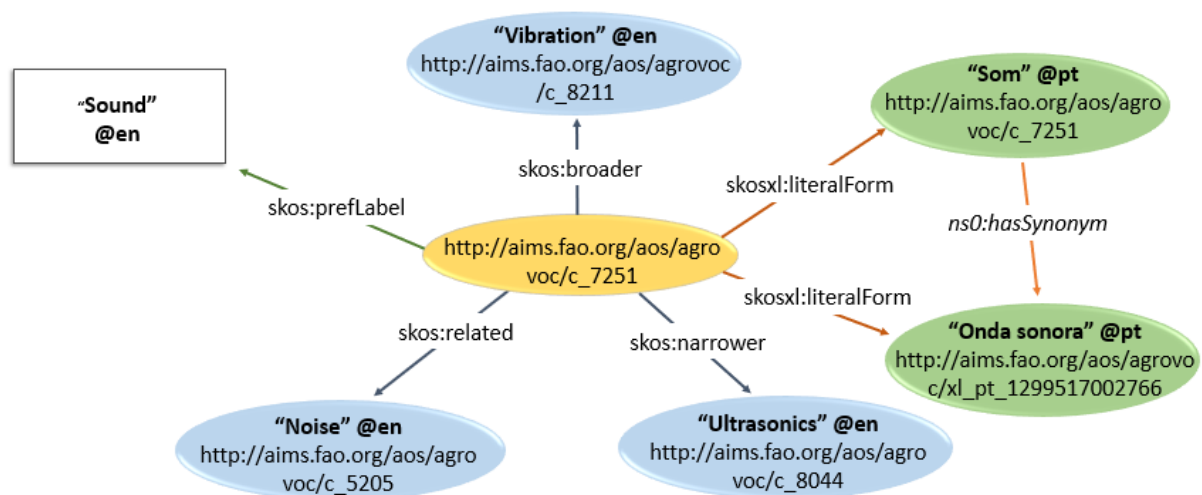
Propriedades	Descrição
<i>skosxl:Label</i>	Classe especial de entidades lexicais
<i>skosxl:literalForm</i>	Forma literal de uma entidade lexical
<i>skosxl:prefLabel</i>	Etiquetas preferidas

<i>skosxl:altLabel</i>	Etiquetas alternativas
<i>skosxl:hiddenLabel</i>	Etiquetas ocultas
<i>skosxl:labelRelation</i>	Serve como um ponto de extensão, que pode ser refinado para cenários de rotulagem mais específicos.

Fonte: W3C (2009).

Para exemplificar a aplicação do modelo, a Figura 19 ilustra o conceito “Sound” representado no AGROVOC no idioma inglês. A propriedade *skosxl:literalForm* é usada para expressar duas variantes lexicais no idioma português, “Som” e “Onda sonora”. A relação sinonímia entre esses dois termos é especificada mediante propriedade *ns0:hasSynonym* da Agrontology<sup>33</sup>.

Figura 19 - Utilização das propriedades do SKOS-XL no AGROVOC



Fonte: Elaborado pelo autor (2021).

Enquanto os rótulos SKOS descrevem propriedades simples, os rótulos SKOS-XL são classes que podem ser vinculadas a outros conceitos por propriedades de objetos. No SKOS Core apenas os conceitos podem ser operados como recursos, as entidades alternativas se apresentam vinculadas somente aos conceitos e não é possível serem descritas com mais detalhes. Com a aplicação do SKOS-XL, entidades lexicais relacionadas a um conceito

<sup>33</sup> *Vocabulário Web Ontology Language (OWL)* que fornece um conjunto de propriedades de domínio para o AGROVOC Thesaurus (<http://aims.fao.org/agrovoc/agrontology>).

representado em SKOS também puderam ser tratadas como rótulos semânticos nos esquemas de conceitos.

Apesar dos benefícios do SKOS-XL para representar entidades lexicais complexas, a extensão não cobre alguns relacionamentos de equivalência compostos. Segundo Pastor-Sánchez (2015), uma equivalência simples se estabelece entre dois conceitos do tesauro, porém quando um conceito de um tesauro está relacionado a vários outros essa relação é considerada como um tipo de equivalência composta. Nesse caso, se tornou ainda mais necessário adaptar o SKOS, tendo em vista a atualização das normas ISO-25964:2011 e a ISO-25964:2013.

Essas normas discorrem acerca da necessidade de representar os conceitos considerando tipos de relacionamentos mais complexos, que podem exigir níveis de correspondências sintáticas e semânticas mais difíceis de realizar, especialmente no que toca à interoperabilidade semântica entre vocabulários controlados diferentes (MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019a). A partir disso, outras experiências foram realizadas tentando melhorar a modelagem dos vocabulários. Uma delas gerou a extensão ISO-THES.

## b) ISO-THES

O ISO-THES foi criado a fim de sanar as dificuldades encontradas durante a “skosificação” do tesauro da UNESCO. O principal desafio se concentra na representação de diferentes níveis de relacionamento e na interoperabilidade com outros SOC. A norma ISO-25964-1:2011 trata da interoperabilidade dos tesouros e define uma tipologia de equivalências entre os conceitos de vocabulários diferentes, os quais são apresentados na sequência trazendo alguns exemplos de uso de termos do domínio da Música:

- i. **equivalência exata:** ambos os conceitos, nos vocabulários de origem e de destino, representam a mesma ideia e ambos têm escopo idêntico. Exemplo: “Educação musical” e “Ensino da música”;
- ii. **equivalência inexata:** conceitos que têm escopos sobrepostos ou uma pequena diferença de conotação ou, ainda, contexto de uso. Exemplo: “Canções” e “Árias”. “Canções” é um tipo de composição musical para música vocal, enquanto que “Árias” é uma peça musical ou uma composição musical escrita para um cantor solista;
- iii. **equivalência parcial:** conceitos com significado próximo, embora um seja um pouco mais amplo que o outro. Exemplo: “Cameratas” e “Orquestra de câmara”. A camerata concerne a um grupo de poucos instrumentistas especializados em

executar um tipo de gênero musical específico. Já a Orquestra de câmara é o termo mais usual para se referir a pequenos grupos de instrumentistas;

- iv. **mapeamentos hierárquicos:** utilizado quando um conceito é claramente mais amplo que o outro. Exemplo: “Teoria musical” e “Harmonia”. Este último é uma área da teoria musical;
- v. **mapeamentos associativos:** empregado quando não é possível definir qualquer equivalência ou mapeamento hierárquico entre dois conceitos, mas é semanticamente relacionado. Exemplo: “Composição musical” e “Arranjo musical”. Este último se refere a uma técnica de preparação da composição musical para a execução de vozes ou instrumentos.

Foram criadas classes e propriedades para ajudar na declaração desses relacionamentos e para estruturar termos conforme as áreas ou disciplinas do tesouro. O ISO-THES não só permite uma melhor visualização dos conceitos à medida que os agrupa por áreas de conhecimento, como também promove a organização e a vinculação dos triplos RDF de maneira mais adequada, facilitando a manutenção do conjunto de dados do tesouro.

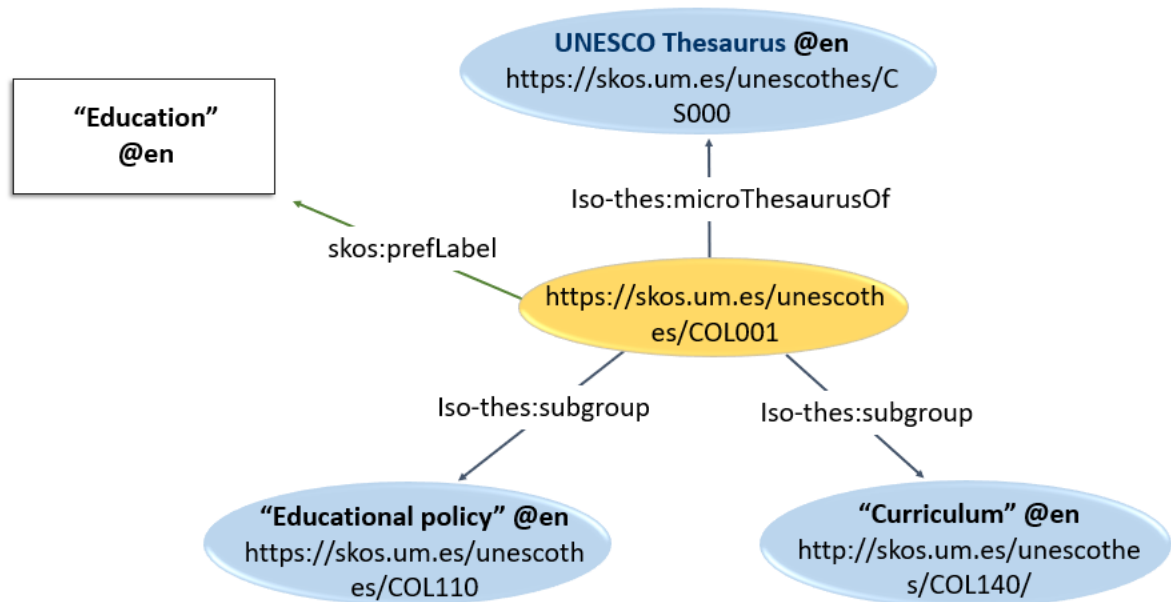
**Quadro 10** – Exemplos de classes do vocabulário ISO-THES

<b>Principais classes</b>	<b>Descrição</b>
iso-thes:CompoundEquivalence	Para representar a relação de equivalência composta entre um termo preferido e um termo não preferido.
Iso-thes:ConceptGroup	Para evidenciar um grupo de conceitos que constituem um subconjunto do tesouro.
iso-thes:microThesaurusof	Para vincular um termo a um grupo de conceitos do tesouro, reconhecida como um microtesouro.
Iso-thes:supergroup	Para associar um termo a um grupo superior de conceitos do microtesouro.
iso-thes:subgroup	Para relacionar um termo a um grupo subordinado de conceitos do microtesouro.
iso-thes:ThesaurusArray	Para relacionar grupos de conceitos irmãos.

**Fonte:** Isaac e Tudhope (2015).

Como exemplo, temos a estruturação dos grupos e subgrupos de conceitos no Tesouro da UNESCO da versão espanhola disponibilizada em SKOS (Figura 20). O termo em realce é versado como um rótulo principal de esquema de conceitos. Subordinado a ele há diversos grupos de conceitos tratados como microtesauros. Um desses grupos é representado pelo conceito “*Education*”, visto estar relacionado a outros conceitos tratados como subgrupos de conceitos, como “*Education policy*” e “*Curriculum*”.

**Figura 20** - Utilização das propriedades do ISO-THES no UNESCO Thesaurus



**Fonte:** Elaborado pelo autor (2021).

Os microtesauros, supergrupos e subgrupos de conceitos não são estruturados como termos equivalentes, abrangentes ou específicos, mas propriedades empregadas apenas para organizar as áreas do tesouro em termos macroestruturais. Embora os conceitos “*Education policy*” e “*Curriculum*” constituam um subgrupo do conceito “*Education*” eles são representados como domínios a parte, para os quais podem haver outros conceitos vinculados que assumem tipos de relacionamentos convencionais. A partir disso surgiu a necessidade de desenvolver o modelo de extensão UNESKOS.

### c) UNESKOS

A extensão UNESKOS, assim como a ontologia ISO-THES, foi produzida pelo grupo de trabalho ligado ao UNESCO Thesaurus. Proposta para mitigar a limitação de manutenção dos diferentes tipos de equivalências e agrupamento de múltiplos domínios em um único esquema de conceitos. Um dos gatilhos para o seguimento dessa extensão foi a dificuldade em representar áreas de conhecimento e os microtesauros, uma vez que o SKOS não fornece nenhum elemento nativo para uma representação direta dessas estruturas. Por essa razão, o UNESKOS busca complementar o SKOS cumprindo as especificações da ISO-25964, à medida que facilita a representação de qualquer tipo de tesouro, estruturando-os segundo grupos,



subgrupos e supergrupos de classes de um domínio específico (ISAAC; TUDHOPE, 2015; PASTOR-SÁNCHEZ, 2015).

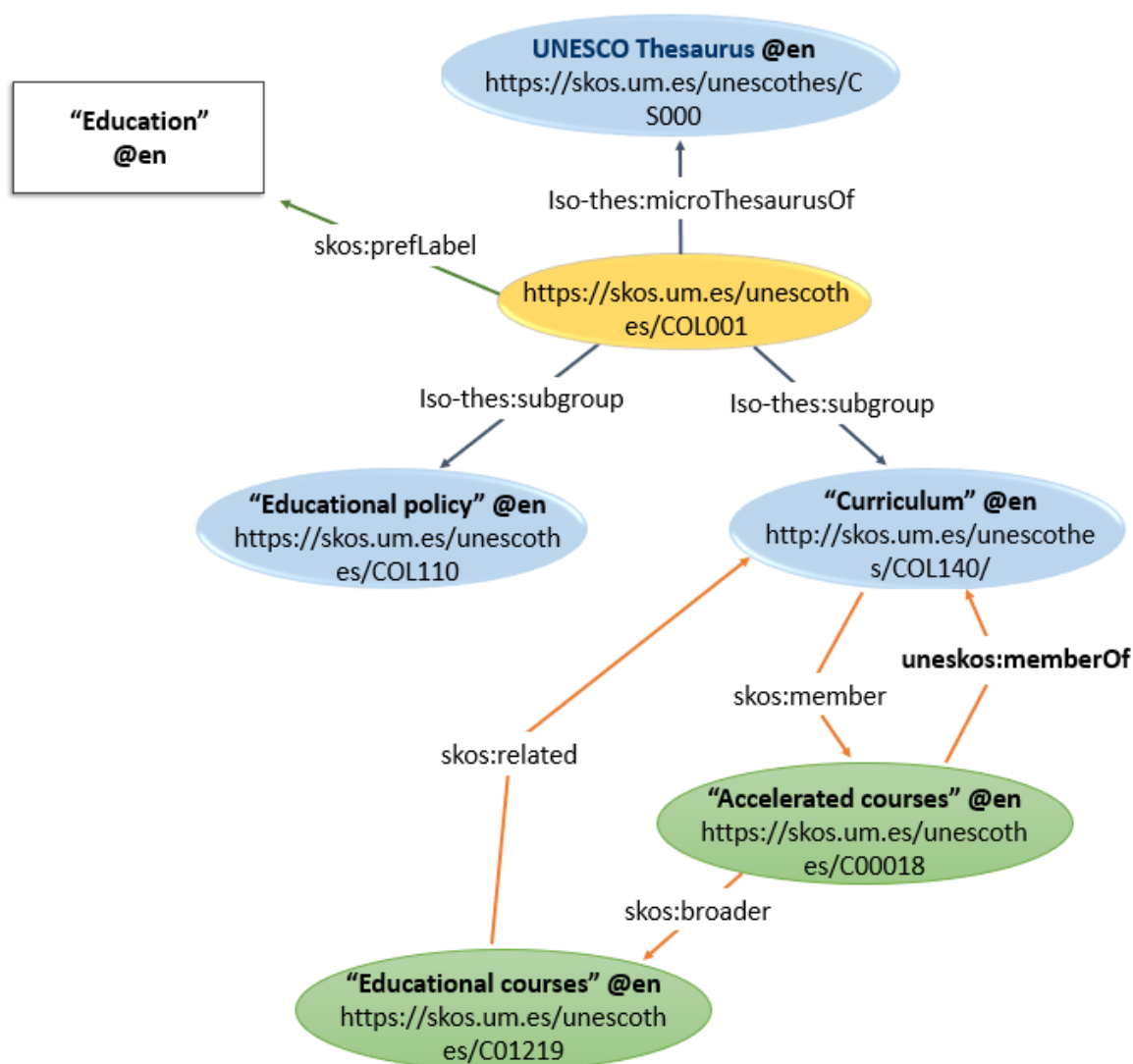
**Quadro 11 - UNESKOS Vocabulary Specification**

<b>Propriedades</b>	<b>Descrição</b>
uneskos:contains	Propriedade inversa do <i>skos:inScheme</i> que permite definir os elementos que estão contidos em um esquema de conceito.
uneskos:hasMainConcept	Propriedade que permite definir certos conceitos como pontos de acesso às coleções de conceito.
uneskos:hasMicroThesaurus	Propriedade proposta como um mecanismo para descobrir o(s) grupo(s) de conceito(s) que conformam um esquema de conceito. Define o caminho para o grupo de conceitos que o conceito pertence.
uneskos:mainConceptOf	Conecta um conceito a uma coleção.
uneskos:memberOf	Propriedade inversa do <i>skos:member</i> para identificar a(s) Coleção (ões) às quais um conceito pertence. Esta propriedade permite definir links dos conceitos para suas coleções.

**Fonte:** Pastor-Sánchez (2015).

Para melhor ilustrar a aplicação da extensão UNESKOS, a Figura 21 apresenta um exemplo de uso da propriedade *uneskos:memberOf* no UNESCO Thesaurus. O conceito “Curriculum” é um grupo de conceitos o qual “Accelerated courses” veicula como um conceito membro. Este último, por sua vez, está vinculado a outros conceitos mais abrangentes, relacionados ou mais específicos.

**Figura 21** - Exemplo de aplicação do UNESKOS no UNESCO Thesaurus



**Fonte:** Elaborado pelo autor (2021).

A aplicação do UNESKOS amplia as possibilidades de estruturação dos conceitos na macroestrutura do tesouro, na perspectiva de que um termo pode estar relacionado a diversas classes de conceitos. Assim, os tesouros passam de disposições convencionais planas a uma rede semântica de conceitos mais complexas que podem haver outros níveis de relacionamentos, além dos tipos tradicionalmente aplicados.

A utilização do SKOS e dos modelos de extensão contribuem com a permeabilidade e a evolução dos tesouros no ecossistema de dados conectados. Desse modo, as tecnologias do LOD modificaram a forma como os tesouros são construídos e publicados, e por isso considera-se o aspecto colaborativo como um dos requisitos para tornar a prática de construção mais ágil

e ampliar suas possibilidades de consumo diante da comunidade do LOD. Nesse sentido, na seção seguinte serão apresentadas as iniciativas do *Linked Data* para o domínio da Música, sendo esta a área de aplicação tratada neste estudo.

## 2.4 INICIATIVAS DO *LINKED DATA* NO DOMÍNIO DA MÚSICA

No cenário atual em que as tecnologias de *Linked Data* são utilizadas para expandir os relacionamentos semânticos entre os recursos na *web*, inúmeras plataformas e repositórios digitais têm empregue uma ampla variedade de modelos de dados, tecnologias e padrões da WS. Esses recursos contribuem para enriquecer suas bases de dados, vinculando obras e artistas a outros recursos digitais que têm a IM como foco principal (DE ROURE et al., 2011; CRAWFORD et al., 2014; KELLY, 2015).

De igual modo, o campo da RIM tem incorporado modelos criados pela comunidade do *Linked Data* em seus projetos como uma forma de otimizar e integrar recursos musicais na *web*. A iniciativa busca trazer soluções para a quantidade de dados de música disponíveis na rede atualmente, sejam em plataformas digitais, sejam em coleções de bibliotecas, arquivos e museus. Apesar de aproveitar todos os benefícios da vinculação de dados, a RIM no âmbito do *Linked Data* ainda enfrenta alguns desafios que serão abordados adiante.

A crescente disponibilidade de IM em formatos digitais tem ocasionado uma mudança de comportamento dos usuários que interagem com uma vasta quantidade de bases de dados de música na *web* (KELLY, 2010, 2015; ALLIK; FAZEKAS; SANDLER, 2016). Para aproveitar todo o potencial dos recursos informacionais frente a esse paradigma de acesso e uso da informação, o campo da RIM tem produzido diversos meios que ampliam as possibilidades de pesquisa (SCHEDL; GÓMEZ; URBANO, 2014).

A maioria dos métodos de pesquisa em RIM são computacionais, mas as técnicas podem variar não só em relação ao tipo de obra musical, bem como em conformidade com os papéis que ela pode desempenhar na vida dos usuários. Além das tecnologias que fazem recomendações personalizadas, *softwares* que estimam a altura e ritmo de faixas musicais e *scanners* que convertem música impressas em partituras digitais editáveis, o campo da RIM procura gerar meios ou interfaces de busca amigáveis para que seus usuários possam recuperar e acessar de forma rápida e prática a IM (BURGOYNE; FUJINAGA; DOWNIE, 2015).

No contexto da RIM com foco na percepção do usuário, estudos mostram uma forte correlação entre as atividades profissionais ou sociais dos consumidores de música. Alguns aspectos são fundamentais, como a importância do humor/sentimento da música e a capacidade

desses seres de interagir com os recursos de busca em plataformas digitais – serviços de *streaming*. Assim, a RIM não está estritamente relacionada às tecnologias computacionais, mas também ao comportamento desses sujeitos, especialistas ou não, ao acessar e usar o conteúdo musical (KAMALZADEH; BAUR, 2012; WOELFER; LEE, 2012)

Algumas abordagens computacionais na RIM criam modelos e utilizam recursos que descrevem a música mediante uma ou mais categorias de percepção. Essas consideram os elementos de baixo e alto nível da Música. Por conseguinte, os tipos de percepção da música são apresentadas por Schedl, Flexer e Urbano (2013, p. 526), ilustrados na Figura 22, nas quais variam entre conteúdo da música, contexto da música, propriedades do usuário e contexto do usuário.

De modo geral, o **conteúdo da música** está relacionado aos “[...] aspectos que são codificados no sinal de áudio, enquanto o **contexto da música** compreende fatores que não podem ser extraídos diretamente do áudio, mas estão relacionados ao item da música, artista ou intérprete.” (SCHEDL; GÓMEZ; URBANO, 2014, p. 5).

**Figura 22** - Fatores que influenciam a percepção humana da música



**Fonte:** Adaptado de Schedl, Flexer e Urbano (2013, p. 526).

Em relação ao **conteúdo da música**, características como ritmo, melodia e timbre são considerados elementos essenciais para a representação e recuperação da informação musical. Por outro lado, o contexto histórico, político e cultural do artista, rótulos semânticos (na maioria

das vezes atribuídos de forma colaborativa pelos usuários) e informações complementares sobre o álbum são itens do **contexto da música**.

Na extremidade desse ator social, aspectos do **contexto do usuário** representam fatores desenvolvidos que podem se alterar dependendo da atividade social desse sujeito. Quanto às **propriedades do usuário** são consideradas as particularidades constantes ou levemente dinâmicas (mais difíceis de se alterar), como a preferência musical e o nível de especialidade, a exemplo a habilidade de tocar instrumentos (SCHEDL; FLEXER; URBANO, 2013; SCHEDL; GÓMEZ; URBANO, 2014).

Diante da complexidade em torno da percepção e uso da Música pelos diversos públicos, esforços devem ser dispensados para expandir a capacidade de vinculação de dados musicais entre os recursos de busca na *web*. Tais iniciativas representam um benefício tanto para o campo científico da Música, na figura de pesquisadores e interpretes que utilizam a IM, como objeto de pesquisa, e também para que usuários não especialistas usufruam do rico acervo cultural expresso por meio de obras musicais (PAGE et al., 2010; KELLY, 2015; NURMIKKO-FULLER et al., 2016; ABDALLAH et al., 2017).

Diante disso, vincular dados de obras musicais se tornou uma necessidade vigente como forma de maximizar o potencial dos mecanismos da RIM.

Explorar conexões entre peças, pessoas e lugares e relacioná-los à cultura como um todo é uma atividade central da musicologia. À medida que as bibliotecas aumentam a disponibilidade de informações musicais em formato digital, os dados disponíveis para essa pesquisa também se expandem, mas para reunir esses recursos e combiná-los com outros relevantes, é necessário um passo adicional de alinhamento e vinculação (CRAWFORD et al., 2014, p. 309, tradução nossa)

Incorporar dados de música é especialmente útil na RIM, visto proporcionar uma maior integração entre os dados de alto nível (representados pelo contexto da música e do usuário) e de baixo nível (representados pelo conteúdo da música). De igual modo, tornar esses dados interoperáveis torna-se um desafio à medida que as características acústicas da música estão profundamente incorporadas à mídia, enquanto que o significado de música pode ser ambíguo, subjetivo e depender do contexto do usuário (ALLIK; FAZEKAS; SANDLER, 2016; KEIL; WARD, 2017).

Nessa perspectiva, algumas iniciativas de LOD no âmbito da IM disponibilizam um conjunto de dados que podem auxiliar na RIM, porque compreendem tanto dados contextuais como dados profundamente incorporados à mídia, geralmente associados ao conteúdo dos recursos musicais. Esses projetos podem ser de instituições de patrimônio cultural (como

bibliotecas, museus e arquivos) ou demais organizações que mantêm um conjunto de dados para uso nas plataformas digitais de música (KELLY, 2015; MORA-MCGINITY et al., 2016).

#### 2.4.1 Projetos de *Linked Data* no âmbito da Informação Musical

Como forma de demonstrar seu potencial serão apresentados alguns projetos de *Linked Data* no âmbito da IM, com ressaltos para os objetivos, instituições mantenedoras, características do público que se destinam, vocabulários, recursos tecnológicos e modelos de metadados utilizados. Ao investigar esses elementos se percebe a amplitude desses recursos e os benefícios que podem proporcionar tanto à comunidade científica especializada como aos usuários gerais.

a) **DBTune**<sup>34</sup>:

Plataforma mantida pelo *Centre for Digital Music*, da *Queen Mary University of London*, que interliga e publica mais de 14 bilhões de triplos de diferentes repositórios de música sob licenças do CC. Fornece URI para os recursos e permite que qualquer projeto no âmbito LOD compartilhem anotações sobre recursos musicais na *web*. Seus dados são vinculados com os dados do DBpedia, Geonames e *MusicBrainz* e são descritos em termos de conceitos e relacionamentos definidos pelo *Music Ontology* (RAIMOND; ABDALLAH; MARK SANDLER, 2007; CENTRE FOR DIGITAL MUSIC, 2017; ETCHEVERRY; VAISMAN, 2012; KELLY, 2015).



A proposta do DBTune é mapear os registros de gravações musicais e álbuns das coleções de diversas plataformas relacionadas à música, como MySpace, *MusicBrainz* e a rádio da BBC (SANTIS, 2016).

---

<sup>34</sup> <http://dbtune.org/>

**Figura 23** - Página principal do DBTune.org

**DBTune.org**  
 DBTune - Serving music-related RDF since 2007

**0. Disclaimer**  
 This is a non-commercial site. The information contained in this site has been collected from several sources and its accuracy cannot be guaranteed. No copyright has been intentionally violated. If you feel a copyright has been violated please contact [me](#) immediately and the material will be removed

**1. Introduction**  
 DBTune hosts a number of servers, providing access to music-related structured data, in a [Linked Data](#) fashion. All the services hosted here are based on open Web standards such as [RDF](#) and [SPARQL](#).  
 This effort is part of the [Linking Open Data on the Semantic Web](#) community project, hosted by the [Semantic Web Education and Outreach](#) interest group.  
*November 2017 update:* This server has recently been revived at the Centre for Digital Music following a period of downtime. Many of the datasets have been restored, mostly using a default ClíoPatria server viewing static data (i.e. datasets that were originally dynamic are now old static snapshots). Some datasets have been officially retired. More details are below.

**2. Quick overview of underlying concepts (or the "Semantic Web" in 10 lines... Don't slap me, please)**  
 Resources on the Web can be far more than web pages. They can identify anything: [me](#), [a French band](#), [an audio signal](#), etc.  
 Such resources have associated *representations* (accessed, in our case, through [HTTP](#)), which may be either human-readable (an [XHTML document](#), for example) or machine-processable ([Microformats](#) or [RDF](#)). These representations may hold links to further resources, allowing to jump from one resource to another, which may be actually hosted in different places. This is were

**Fonte:** *Centre for Digital Music* (2017).

b) **BBC Music**<sup>35</sup>:

A *British Broadcasting Corporation* (BBC) é a maior corporação de rádio difusão no mundo que publica grande quantidade de conteúdo *online*, como texto, áudio e vídeo. A BBC fornecia apenas páginas no formato *HyperText Markup Language* (HTML), separadas e independentes, projetadas para o acesso a partir de um navegador *web*. A ausência de um *cross linking* limitou o tipo de interação que a BBC é capaz de fornecer ao usuário devido à falta de integração em nível de dados.

Assim, intencionando sanar esse problema a corporação lançou em 2007 o *BBC Programs* objetivando fornecer um identificador *web* para seus recursos, incluindo páginas HTML associadas e *feeds* legíveis por máquina para cada programa transmitido. A iniciativa possibilitou que outras equipes dentro da empresa incorporassem seus conteúdos em páginas de diversos programas, seja por meio de canais de TV, seja por meio de estações de rádio.

A *BBC Music* segue os mesmos princípios que os programas da BBC e fornece URI para todos os artistas que têm interesse (destaque em programas de música, em eventos da BBC e outros). O projeto é apoiado pelo conjunto de dados sobre música do *MusicBrainz* e pela Wikipédia. Igualmente, seus editores também contribuem com o *MusicBrainz* e Wikipédia e oferecem uma visão agregada dessas informações inseridas no contexto da BBC.

<sup>35</sup> <http://www.bbc.com/culture/music>

Figura 24 - Página da artista no BBC Music

Fonte: BBC Music (2019).

**c) RISM (*Répertoire International des Sources Musicales*)<sup>36</sup>:**

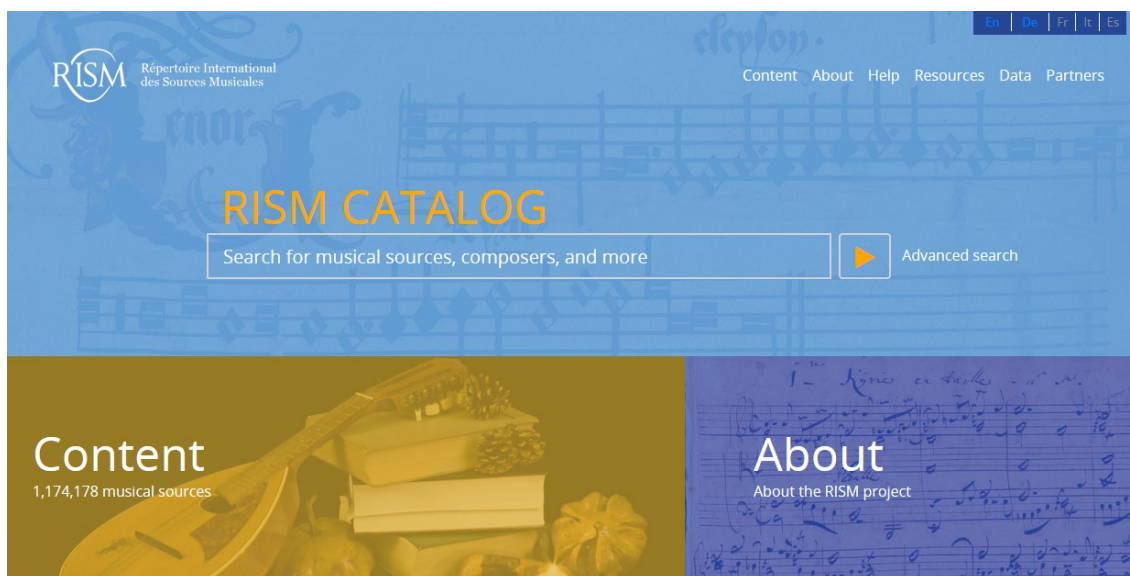
Projeto internacional que registra e descreve fontes musicais (manuscritos ou música impressa, escritos sobre teoria musical e libretos) em acervos espalhados por todo o mundo. O projeto, iniciado em 1952, é fruto de uma parceria entre a *International Musicological Society* (IMS) e a *International Association of Music Libraries, Archives and Documentation Centres* (IAML). Atualmente, a RISM tem parcerias com instituições, grupos de trabalho e indivíduos em mais de 35 países nos quais colaboradores detalham as fontes musicais preservadas em seus acervos e enviam dados ao RISM *Zentralredaktion* (Escritório Central) em Frankfurt, Alemanha.

O Escritório central edita e unifica as informações antes da publicação. Por conseguinte, seu conjunto de dados descreve fontes musicais que datam de 1600 a 1850, porém os dados publicados antes ou depois desse intervalo podem ser encontrados. As informações são recuperadas gratuitamente em um Catálogo de Acesso Público *online* (OPAC). Desde 2013, esses dados estão disponíveis como dados abertos, e a partir de 2014 como dados abertos conectados. Esse serviço é direcionado às bibliotecas que desejam importar os registros disponíveis na RISM em seus catálogos locais ou projetos musicológicos. O conjunto de dados pode ser recuperado por meio de um terminal de consultas SPARQL (KEIL; WARD, 2017).

<sup>36</sup> [www.rism.info](http://www.rism.info)



**Figura 25** - Catálogo online da RISM



Fonte: RISM (2019).

**d) *Musical Instrument Museums Online (MIMO)*<sup>37</sup>:**

O projeto iniciou com a colaboração de onze instituições europeias, sendo financiado, em parte, pelo programa *eContentplus* da *European Commission*. A comissão tem como objetivo tornar o conteúdo digital europeu acessível, utilizável e explorável, facilitando a criação e difusão de informações em áreas de interesse público. Inicialmente, o MIMO ocorreu entre 2009 e 2011, período em que os museus colaboradores produziram cerca de 35.000 novos registros digitais de instrumentos em seus acervos, adicionados a 10.000 registros já existentes.

A Europeana, agregadora de conteúdo para a coleção digital europeia, promoveu uma exposição em 2011 (*Explore the World of Musical Instruments*) para mostrar a amplitude do MIMO. No início do projeto, os acervos das instituições contribuintes representavam aproximadamente 40% dos instrumentos mantidos em museus públicos europeus (DOLAN, 2017). O projeto disponibilizou conteúdo multilíngue, possibilitando a pesquisa em seis idiomas diferentes e associando o vocabulário não especializado a termos e sistemas de classificação usados pelos profissionais.

Outros resultados importantes incluíram o desenvolvimento de uma documentação que estabelece os padrões para fotografar instrumentos musicais, na perspectiva de que outros museus possam usá-los para digitalizar suas coleções e diretrizes detalhadas sobre como configurar um repositório para permitir a coleta de conteúdo digital.

<sup>37</sup> <http://www.mimo-international.com/MIMO/>

O propósito do consórcio é tornar o único ponto de acesso de informações sobre coleções públicas de instrumentos musicais em todo o mundo e, para tanto, continua buscando parcerias com outros museus. Assim, o MIMO pretende ampliar o projeto não apenas restringindo a acervos de museus da Europa, mas no globo para que outros museus aproveitem seus recursos e contribuam com uma iniciativa global de vinculação de dados de coleções de instrumentos musicais.

O MIMO tem um vocabulário controlado próprio de domínio dos instrumentos musicais (MIMO Thesaurus), no qual os usuários podem explorar os conceitos que representam as coleções de instrumentos. O tesouro *online*, construído e publicado como LOD, dispõem de recursos para a visualização dos conceitos, além de oferecer acesso ao conjunto de dados (MIMO-Data Base), facilitando a pesquisa e a gestão dos dados de sua coleção (MIMO, 2019).

**Figura 26 - Musical Instrument Museums Online (MIMO)**



**e) MIDI *Linked Data*<sup>38</sup>:**

Projeto que objetiva conectar todos os arquivos no formato *Musical Instrument Digital Interface* (MIDI) do mundo em um gráfico de instruções MIDI interconectadas utilizando tecnologia da WS como o modelo de dados RDF para expressar informações MIDI na forma de triplos e a linguagem de consulta SPARQL. Esses triplos permitem que os dados musicais sejam interpretados por humanos, visto utilizarem a linguagem natural, assim como possibilitam que as máquinas possam processar seu conjunto de dados. Parte do pressuposto de

<sup>38</sup> <https://midi-ld.github.io/>

que os dados conectados são potencialmente úteis para a pesquisa e análise da notação musical, contribuindo com a interoperabilidade de notações à medida que os formatos de notação atuais usam o RDF para codificar informações musicais e, igualmente, têm o potencial de vincular entidades relacionadas musicalmente a outros projetos da WS.

O propósito do conjunto de dados MIDI *Linked Data* é apoiar pesquisadores que investigam a interoperabilidade de várias notações musicais usando padrões da *web*, além de fornecer um conjunto de dados heterogêneos para o incremento de aplicações na RIM. O projeto fornece código-fonte e complexo de dados por meio do GitHub (plataforma de hospedagem de códigos-fonte e controle de versões) que hospeda todos os repositórios do projeto, incluindo tutoriais, coleções MIDI de origem e o código de geração de conjunto de dados. Esses dados são acessíveis para *download* na plataforma de hospedagem mediante um Endpoint SPARQL ou por meio de uma *Application Programming Interface* (API) (MEROÑO-PEÑUELA et al., 2017).

**Figura 27** - Página principal do MIDI *Linked Data*



**Fonte:** MIDI *Linked Data* (2019).

**f) DOREMUS<sup>39</sup>:**

O *Doing Reusable Musical Data* (DOREMUS) é um projeto de pesquisa baseado nas tecnologias da WS que busca desenvolver ferramentas e métodos para descrever, publicar,

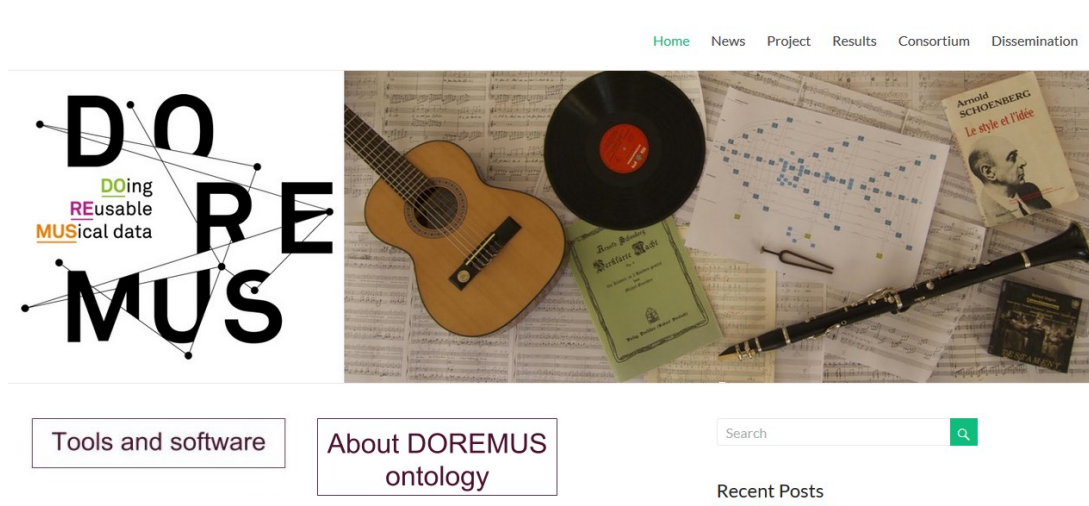
<sup>39</sup> <http://www.doremus.org/>

conectar e contextualizar catálogos de música na rede de dados. Mediante modelos de dados e vocabulários controlados compartilhados multilíngues, esse projeto oferece novas formas de explorar a IM de acordo com as necessidades dos usuários.

À vista disso, foi criado a fim de vincular os dados de três instituições musicais francesas: A Biblioteca Nacional da França (BnF), a *Rádio France* (RF) e a *Philharmonie de Paris* (PP). O projeto reúne profissionais de várias especialidades e suas parcerias vão além das três instituições provedoras de dados, visto reunir colaboradores do *Laboratory of Informatics, Robotics and Microelectronics of Montpellier* (LIRMM), do laboratório do Centro de Pesquisa em Sistemas de Comunicação (EURECOM), do laboratório de Ciência da Informação e Comunicação (GERiiCO) da Universidade de Lille e duas empresas privadas (Ourouk e Motores Meaning).

O DOREMUS foi desenvolvido como uma extensão do modelo *Functional Requirements for Bibliographic Records* (FRBR) e do *Conceptual Reference Model* (CRM), do *Committee for Documentation of the International Council of Museums* (CIDOC), criado para descrever o patrimônio cultural. A extensão desses modelos visa suprir a lacuna existente nos vocabulários amplamente utilizados, como o *Music Ontology*, que não contempla a descrição de concertos, festivais e registros de gravações, pois são parte da atividade da RF e da PP (DOREMUS, 2019).

**Figura 28** - Página principal do DOREMUS



**Fonte:** DOREMUS (2019).

### g) *MusciBrainz*<sup>40</sup>:

Funciona como uma enciclopédia aberta mantida de forma colaborativa por uma comunidade global de usuários que coleta dados de música e os disponibiliza a partir de uma licença livre na *web*. É mantido pela *MetaBrainz Foundation*, corporação sem fins lucrativos com sede na Califórnia.

Os dados podem ser adicionados usando sua interface *web*, na qual os usuários cadastrados adicionam informações por meio de *softwares aplicativos* que identificam automaticamente dados sobre artistas, álbuns, faixas de gravações musicais mantidos em seus computadores pessoais – um tipo de colaboração coletiva conhecida como *Crowdsourcing*.

O conjunto de dados disponível utiliza um esquema proprietário que pode ser mapeado diretamente com diversas ontologias e aplicado em vários projetos de LOD no âmbito da IM, como é o caso do *BBC Music*, que usa o *MusicBrainz* para agregar informações sobre faixas sonoras e artistas (METABRAINZ FOUNDATION, 2019).

Figura 29 - Página principal do *MusicBrainz*

The screenshot shows the MusicBrainz homepage layout. At the top, there is a navigation bar with 'Log In' and 'Create Account' links, a search bar, and a language selector set to 'English'. Below the navigation bar, there is a 'Welcome to MusicBrainz!' section with a brief description and a list of goals. To the right, there are sections for 'MusicBrainz Blog' with 'Latest posts' and 'Tag Your Music' with 'Quick Start'. Further down, there are sections for 'Community', 'MusicBrainz Database', and 'Developers'. At the bottom, there is a 'Recent Additions' section displaying a row of album covers.

Fonte: *MusicBrainz* (2019).

### h) *AcousticBrainz*<sup>41</sup>:

O projeto visa compilar informações acústicas de músicas e disponibilizá-las ao público de forma aberta, com vistas a caracterizar espectrais de baixo nível (informações tonais

<sup>40</sup> <https://musicbrainz.org/>

<sup>41</sup> <https://acousticbrainz.org/>



e ritmicas como, por exemplo, notação, dancibilidade e batidas por minuto) e informações de alto nível, como gêneros, humor/sentimento e timbre. O público-alvo são pesquisadores da área de Informática Musical, pois fornece um banco de dados complexo para o suporte de pesquisas em tecnologia musical, especialmente no contexto dos mecanismos de recomendação.

O *AcousticBrainz* é fruto da parceria entre o *Music Technology Group*, da Universidade Pompeu Fabra, em Barcelona, e do projeto *MusicBrainz*. Por meio de um kit tecnológico conhecido como *Essentia*, disponibiliza ferramentas de código aberto que permitem a análise automática de registros sonoros. Os dados de saída são coletados pelo projeto *MusicBrainz* e disponibilizado ao público por meio de um identificador único (ID) que representa uma gravação musical. Esse ID, conhecido no *MusicBrainz*, pode ser facilmente recuperado na base de dados do *AcousticBrainz*, os quais são licenciados sob a *Creative Commons* de Domínio público (MUSIC TECHNOLOGY GROUP, 2019).

**Figura 30** - Página principal do *AcousticBrainz*

**AcousticBrainz** About ▾ Downloads API/Data Datasets ▾ Sign in

Welcome to AcousticBrainz!

The AcousticBrainz project aims to crowd source acoustic information for all music in the world and to make it available to the public. This acoustic information describes the acoustic characteristics of music and includes low-level spectral information and information for genres, moods, keys, scales and much more. The goal of AcousticBrainz is to provide music technology researchers and open source hackers with a massive database of information about music. We hope that this database will spur the development of new music technology research and allow music hackers to create new and interesting recommendation engines.

AcousticBrainz is a joint effort between **Music Technology Group** at **Universitat Pompeu Fabra** in Barcelona and the **MusicBrainz** project. AcousticBrainz was originally envisioned by Xavier Serra, the founder and head of the MTG. At the heart of this project lies the **Essentia toolkit** from the MTG – this open source toolkit enables the automatic analysis of music. The output from Essentia is collected by the AcousticBrainz project and made available to the public.

AcousticBrainz organizes the data on a recording basis, indexed by the **MusicBrainz ID** for recordings. If you know the MBID for a recording, you can easily fetch from AcousticBrainz. For details on how to do this, visit our **API documentation**.

All of the data contained in AcousticBrainz is licensed under the **CC0 license** (public domain).

Examples

If you're wondering what this collected data actually looks like, have a look at the last 5 recordings that have been submitted:

1. Proem / Scapula
2. Lighthouse family / Raincloud

**Data Statistics**

	Lossy	Lossless	Total
<b>Unique</b>	3,091,437	1,480,098	<b>3,865,451</b>
<b>All</b>	8,436,911	2,957,382	<b>11,394,293</b>

updated every 10 minutes  
last updated: 2018-12-24 04:00  
[more on the detailed statistics page](#)

**Tweets** by @AcousticBrainz

**AcousticBrainz** @AcousticBrainz  
We just found a dependency that wasn't the best idea in the world, so we're going to fix that now. In any case, the service is back! -rk  
Jan 29, 2019

**Fonte:** *AcousticBrainz* (2019).

### i) *Music Ontology*<sup>42</sup>:

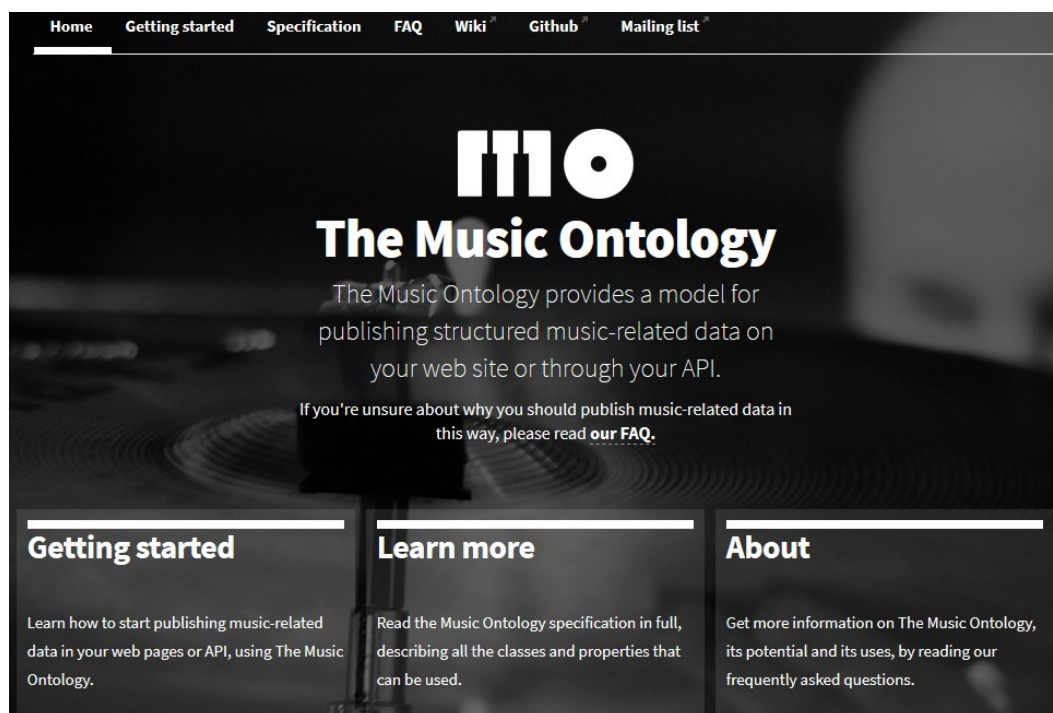
É uma estrutura formal para aplicações que lida com informações relacionadas à música na WS, incluindo materiais editoriais, culturais e acadêmicas. A ontologia fornece um vocabulário para publicar dados conectados na *web* e foi desenvolvida para representar

<sup>42</sup> <http://musicontology.com/>

informações temporais e eventos relativos à música. Logo, foi pensada para expressar tal informação temporal mediante uma Ontologia da Linha de Tempo (OWL-Time) construída sob dois conceitos: *Interval* (representado intervalos de tempo) e *Instant* (representando instantes).

Do ponto de vista dos eventos (*event ontology*), parte da prerrogativa de que a produção musical envolve eventos físicos que ocorrem em dado local e tempo e pode envolver a participação de inúmeros objetos físicos, animados ou inanimados, como *performances* abrangendo alguns músicos e seus instrumentos (RAIMOND; ABDALLAH; SANDLER, 2007). Usa padrões recomendados pelo W3C, como o RDF, e seus formatos de serialização JSON-LD e *Turtle*. As várias serializações RDF disponíveis permitem que a ontologia seja incorporada a dados de música em páginas da *web* como parte de uma API ou compartilhada facilmente por grandes bancos de dados musicais (PICKERING, 2012).

**Figura 31** - Página principal do *The Music Ontology*



**Fonte:** *The Music Ontology* (2019).

#### j) *Linked Jazz*<sup>43</sup>:

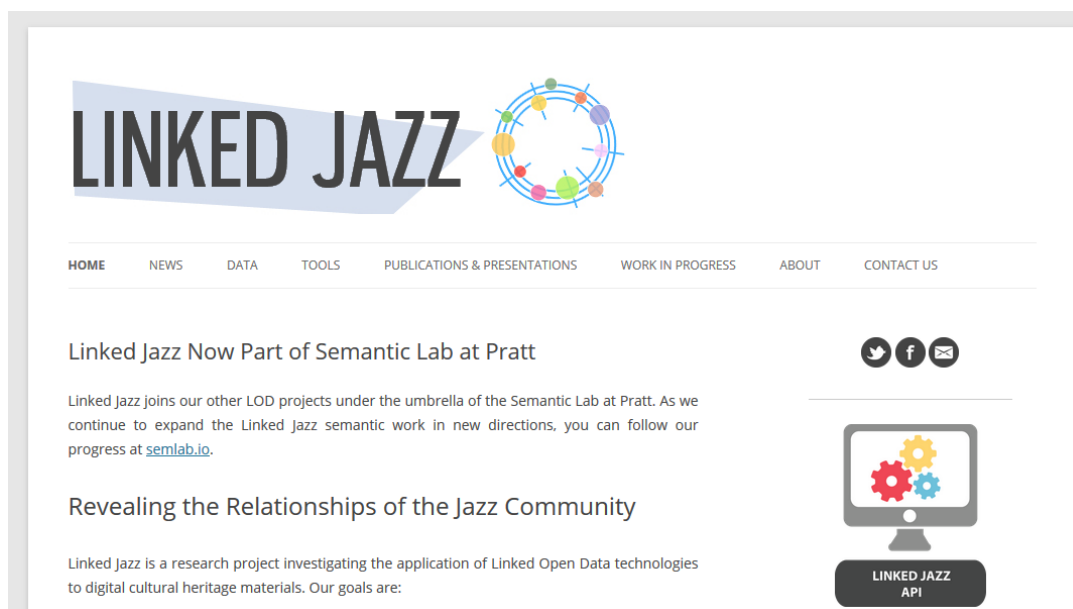
O projeto investiga a aplicação de tecnologias do LOD em materiais de patrimônio cultural digital. Objetiva descobrir conexões significativas entre documentos e dados sobre a

<sup>43</sup> <https://linkedjazz.org/>

vida pessoal e profissional de artistas de jazz. Desenvolve ferramentas e métodos relacionados às iniciativas de dados abertos conectados a fim de revelar as relações entre os músicos a partir de arquivos digitais da história do jazz. É uma iniciativa do *Online Computer Library Center (OCLC)* e *Association for Library and Information Science Education (ALISE) Library and Information Science Research*, que iniciaram o trabalho com base em transcrições de entrevistas vindas de diversas fontes, nas quais incluem: *Jazz Studies Archives*, *Smithsonian Jazz Oral Histories*, *Hamilton College Jazz Archive*, *Central Avenue Sounds Series* da *University of California, Los Angeles (UCLA)* e da *Nathaniel C. Standifer Video Archive of Oral History* da *University of Michigan*.

As entrevistas foram transcritas e os triplos gerados com o auxílio de ferramentas de transcrição e *crowdsourcing*. Nessa última, os excertos foram apresentados aos voluntários da comunidade musical do Jazz com vistas a solicitar a definição de termos mais detalhados para descrever as relações entre os entrevistados e as pessoas mencionadas nos trechos. As ferramentas desenvolvidas demonstraram o potencial para explorar e vincular recursos como agentes e eventos no âmbito da música (PATTUELLI et al., 2013; 2016).

**Figura 32** - Página principal do *Linked Jazz*



**Fonte:** *Linked Jazz* (2019).



**k) *Sheet Music Consortium (SMS)*<sup>44</sup>:**

Consórcio iniciado em 2002 em colaboração com várias universidades para criar uma coleção aberta de partituras digitalizadas usando o protocolo *Open Archives Initiative for Metadata Harvesting* (OAI-PMH). Devido o aumento do número de colaboradores, em 2013 foram criados uma maior variedade de sistemas, esquemas e padrões para facilitar a integração diversificada de metadados de partituras, o que resultou em um conjunto de dados não normalizados. Foi então que parte do consórcio da Universidade da Califórnia decidiu experimentar a publicação de metadados no SMS usando tecnologias do LOD.

O consórcio, apoiado pelo *Institute of Museum and Library Services*, fornece acesso a 228.463 itens de 27 instituições, que incluem 24 universidades, uma biblioteca pública, biblioteca acadêmica e duas bibliotecas nacionais – a Biblioteca do Congresso Americano (*Library of Congress*) e a Biblioteca Nacional da Austrália. (DAVISON et al., 2013). Os metadados das organizações participantes são coletados e hospedados pelo *UCLA Digital Library Program*, que fornece um serviço de acesso aos registros extraídos das bibliotecas cooperantes.

Embora os provedores de dados cataloguem partituras de maneiras diferentes, um número significativo de partituras originais se encontra digitalizado, permitindo que os usuários acessem-no integralmente. Na atualidade, o consórcio de partituras está expondo informações dos editores de música extraídas de metadados agregados como dados abertos conectados. Acredita-se que adicionar uma camada de LOD, que expõe as relações entre compositores, intérpretes, letristas e editores, tem o potencial de criar novos recursos valiosos de acesso e uso da IM para os usuários e pesquisadores (UCLA DIGITAL LIBRARY PROGRAM, 2011).

---

<sup>44</sup> <http://digital2.library.ucla.edu/sheetmusic/>

**Figura 33** - Página principal do *Sheet Music Consortium*



**Fonte:** *UCLA Digital Library Program* (2011).

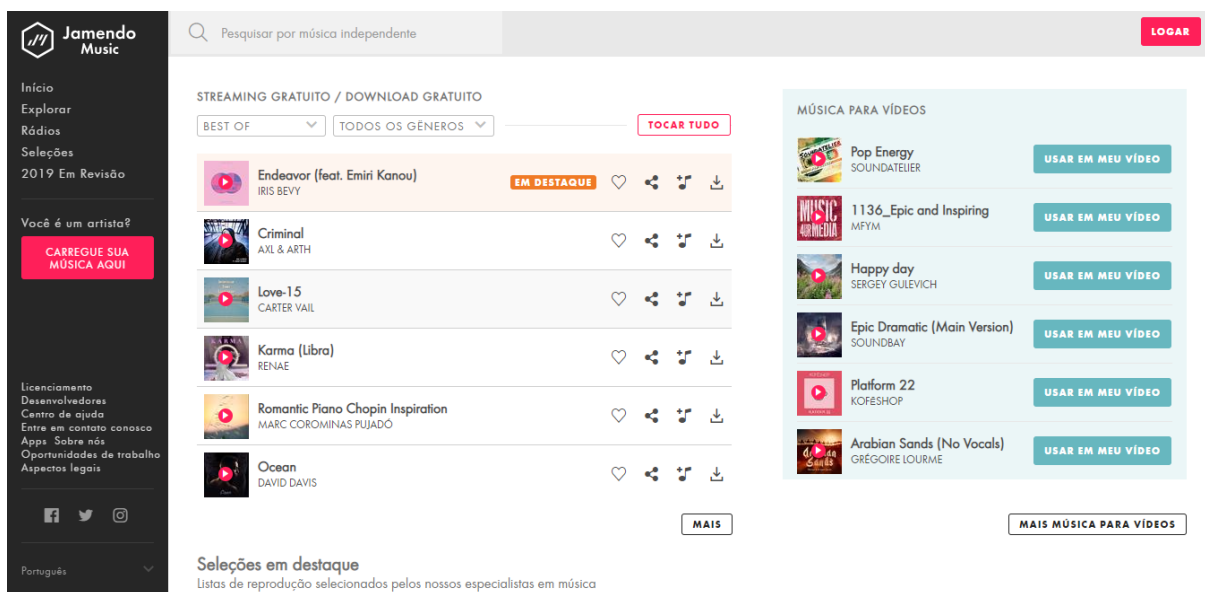
#### **I) Jamendo<sup>45</sup>:**

Repositório de música criado em 2005, possibilita aos artistas disponibilizarem suas composições licenciadas pelo *Creative Commons*. Os usuários podem obter acesso legal às obras sem custos, no entanto alguns artistas podem optar por permissões que impeçam o uso comercial. Quando utilizado para fins comerciais, eles são recompensados pela plataforma.

O Jamendo é um dos pioneiros no movimento de música livre na França e atualmente seu conjunto de dados é estruturado e disponibilizado como *Linked Data*. Esse conjunto de dados está conectado a outros, como o *MusicBrainz*, por exemplo. Os dados do Jamendo podem ser acessados por intermédio de um Endpoint mantido pelo DBTune que funciona como um servidor, no qual pode-se obter acesso aos URI e suas representações em RDF com *links* para conjuntos de dados externos. Os documentos RDF também estão ligados aos termos do *Music Ontology*.

<sup>45</sup> <https://www.jamendo.com/start>

**Figura 34 - Página inicial do Jamendo**



Fonte: Jamendo (2020).

Parte das iniciativas elencadas estão ligadas às instituições de patrimônio cultural e se concentram na vinculação de dados a partir de tesouros, ontologias ou padrões de metadados para associar recursos musicais na *web*. Porém, apesar dos benefícios já listados, o campo da RIM para o contexto do LOD ainda enfrenta alguns desafios, que serão apresentados na sequência.

#### 2.4.2 Desafios e oportunidades para os dados conectados no domínio da Música

Uma das principais dificuldades enfrentadas nos projetos de dados musicais conectados é apontada por Abdallah et al. (2017) e tem relação com o tamanho das coleções, em especial as não textuais, como gravações de áudio. Estas consistem em milhões de dados com necessidades específicas de análise e transcrição de conteúdo musical. Como os dados são ricamente estruturados, são necessários vários tipos de avaliações que podem variar conforme a melodia, harmonia, ritmo, timbre e estrutura de similaridade da música. Assim, relações dentro e entre as obras e coleções podem ser criadas.

Para pensar em uma *web* de dados musicais conectados é preciso conhecer as características básicas da representação simbólica da música (notas, melodias, acordes), bem como os conceitos centrais do padrão MIDI para melhor representá-los em RDF, uma vez que estes se configuram como novos recursos da WS para o contexto da IM.

Segundo Meroño-Peñuela e Hoekstra (2016), a tradução desses conceitos para o RDF é um desafio, principalmente com relação à preservação da ordem dos eventos MIDI. Alguns padrões atuais, como o *MusicXML*, podem facilitar a representação de notação musical usando URI e permitir uma maneira mais padronizada de comparação e combinação da música, o que representaria um benefício para a vinculação de dados de gravações musicais.

Nesse aspecto, duas oportunidades são postas para o contexto do LOD:

[...] manter dados conectados pode ser útil, na medida em que usuários podem executar buscas mais específicas que exigem uma vinculação extensiva a fontes externas de conhecimento, incluindo arquivos de autoridade, DBpedia e outras fontes de dados. [...]. Os musicólogos podem se beneficiar muito de um sistema distribuído que permita análises em dados mantidos por diferentes bibliotecas e arquivos (ABDALLAH et al., 2017, p. 17-18, tradução nossa).<sup>46</sup>

Conforme Meroño-Peñuela et al. (2017), deve-se considerar a tarefa de encontrar automaticamente novos *links* entre as entidades. Assim, o trabalho com arquivos de gravações musicais na perspectiva do LOD tem o potencial para novos desafios em métodos para vinculação de entidades e descoberta de *links*. Nessa oportunidade, o conjunto de dados do MIDI *Linked Data Cloud* permite a combinação potencial de metadados de música com conteúdo musical da seguinte forma:

A primeira tarefa importante é gerar links de qualidade entre notação e metadados. Por exemplo: como encontrar links relevantes entre a partitura de Hey Jude dos Beatles e outros recursos Linked Data que descrevem essa música (por exemplo, [http://dbpedia.org/resource/Hey\\_Jude](http://dbpedia.org/resource/Hey_Jude))? Metadados relevantes para esta tarefa, como o nome do artista e o título da música, podem frequentemente ser extraídos do texto incorporado [ao padrão] MIDI8, mais harmonizado com o *Music Ontology*, e usados para gerar links para serviços do MusicBrainz e o gráfico de informações musicais da DBpedia (MEROÑO-PEÑUELA et al., 2017, p. 5-6, tradução nossa).<sup>47</sup>

Além das bases de dados de gravações musicais tem-se a necessidade de vincular dados em outros tipos de registros como, por exemplo, os registros bibliográficos que guardam informações sobre obras em bibliotecas, arquivos e museus. Uma das iniciativas dessa área foi

<sup>46</sup> “Having metadata fully linked to the Semantic Web would be extremely useful, as musicologists did ask for very specific queries that require extensive linking to outside sources of knowledge, including authority files, DBpedia and other data sources. [...] Musicologists could greatly benefit from a distributed system that enables analyses on data held by different libraries and archives” (ABDALLAH et al., 2017, p. 17-18).

<sup>47</sup> “A first important task is to generate quality links between notation and metadata. For example: how to find relevant links between the score of the Beatles’ Hey Jude and other Linked Data resources describing this song (e.g. [http://dbpedia.org/resource/Hey\\_Jude](http://dbpedia.org/resource/Hey_Jude))? Relevant metadata for this task, such as the artist name and the song title, can often be extracted from MIDI embedded text<sup>17</sup>, further harmonised with the Music Ontology to generate links to e.g. MusicBrainz services and the DBpedia musical information graph” (MEROÑO-PEÑUELA et al., 2017, p. 5-6).

o projeto *Metadata Vocabulary Junction*, caracterizado como um esforço de pesquisa da Escola de Biblioteconomia e Ciência da Informação da *Kent State University*, que busca auxiliar bibliotecas e arquivos a enriquecer seus dados vinculando-os a outras fontes na *web* a partir de tecnologias e padrões do LOD.

O projeto *Metadata Vocabulary Junction* se concentrou em como as bibliotecas podem aumentar o acesso a fontes de informação por meio de conjuntos de dados conectados relevantes utilizando vocabulários RDF. Uma das principais tarefas do projeto foi encontrar elementos que pudessem ser vinculados a partir de dados bibliográficos com fontes de informações musicais disponíveis. Assim, foram mapeados os campos, subcampos e códigos do formato *Machine Readable Cataloging* (MARC) para as classes e propriedades do *Music Ontology*. Por conseguinte, elementos relevantes nos registros bibliográficos foram identificados e testados usando uma amostra representativa de registros MARC que engloba vários gêneros da música (ZENG; GRACY; SKIRVIN, 2013).

O mapeamento dos dados bibliográficos com outros vocabulários RDF é uma oportunidade para as bibliotecas ampliarem a possibilidade de acesso e uso da IM. Seus registros podem se integrar às fontes de dados relevantes para a RIM, como o *BBC Music*, *DBTune* ou *MusicBrainz* e assim por diante, permitindo que os usuários estendam os acessos a partir de pontos de dados úteis.

Do ponto de vista do vocabulário, os conceitos aplicados para representar a IM nessas fontes de dados representam uma oportunidade, porque uma vez utilizados para indexar conteúdo nas bibliotecas os tesouros construídos e publicados como *Linked Data* podem ser associados a outras fontes externas, ampliando a capacidade de busca e recuperação da informação (CARACCILO et al., 2011). Uma terminologia eficiente para o propósito da recuperação é aquela que pode enriquecer semanticamente recursos e, conseqüentemente, otimizar consultas nesses ambientes digitais.

O tesouro parece ser uma alternativa para vincular dados bibliográficos às fontes externas, pois é usado para descrever itens de uma coleção de forma unívoca, permitindo a recuperação da informação semanticamente precisa e sem ambigüidade (JUAN; IGLESIAS, 2010). O Quadro 12 apresenta de maneira resumida as dificuldades que os projetos de LOD no domínio da Música precisam mitigar.

**Quadro 12** - Principais desafios e oportunidades para os dados conectados do domínio da Música

<b>Autores</b>	<b>Desafios</b>	<b>Oportunidades</b>
Abdallah et al. (2017)	Volume de dados; Coleções não textuais (gravação de áudio).	Vinculação entre as obras e coleções.
Meroño-Peñuela e Hoekstra (2016) Abdallah et al. (2017)	Tradução dos conceitos/características básicas da música (notas, melodias, acordes); Codificação no padrão MIDI.	Uso de padrões de metadados, como o MusicXML, para representação da notação musical por meio de URI; Associação de dados de gravações musicais a fontes externas: arquivos de autoridades e DBpedia; Executar buscas e análises de dados musicais a partir de diferentes conjuntos de dados.
Meroño-Peñuela et al. (2017)	Encontrar automaticamente novos <i>links</i> entre as entidades.	Integração de entidades de descoberta de <i>links</i> a partir do conjunto de dados MIDI <i>Linked Data Cloud</i> .
Zeng, Gray e Skirvin (2013)	Vinculação de dados a outros tipos de registros, como os registros bibliográficos.	Mapeamento de registros bibliográficos em padrões MARC para as classes e propriedades do <i>Music Ontology</i> ou demais vocabulários RDF utilizados para representar informação musical na <i>web</i> de dados.
Zeng, Gray e Skirvin (2013) Juan e Iglesias (2010)	Mapeamento de conceitos utilizados para representar a informação musical em bibliotecas.	Mapeamento e publicação de vocabulários representados em SKOS; Vinculação com outras fontes externas para favorecer as buscas contextuais.

**Fonte:** Elaborado pelo autor (2021).

No âmbito do *Linked Data*, os tesouros e ontologias estão sendo construídos de forma colaborativa, sendo capazes de promover o controle compartilhado em projetos que desenvolvem e publicam conteúdo aberto de domínios específicos. Como são elaborados por instituições governamentais, de ensino, pesquisa e cultura – reconhecidas pela sua legitimidade e fidedignidade – os tesouros funcionam como instrumentos para padronizar as representações conceituais e interligar recursos digitais no contexto do *Linked Data*, possibilitando o enriquecimento semântico das informações nessas fontes de dados (TUDORACHE et al., 2010).

Na próxima seção pretende-se apresentar os caminhos para o alcance dos objetivos deste estudo. Serão lançadas as opções metodológicas e estabelecidas as etapas e instrumentos para a coleta, processamento dos dados, análise e elaboração do modelo.

### 3 PERCURSO METODOLÓGICO

O caminho metodológico de uma pesquisa se caracteriza pelo conjunto de procedimentos racionais e sistemáticos definidos a fim de alcançar os objetivos iniciais. O êxito do estudo, em certa medida, depende do rigor com o qual esse conjunto de procedimentos são estruturados e seguidos, cabendo ao pesquisador apresentar delineamentos de maneira ordenada (SEVERINO, 2016).

Nesta seção serão apresentados os traçados da investigação, ou seja, os procedimentos e etapas escolhidos para o atingimento dos objetivos. A Figura 35 caracteriza o estudo, estabelecido a partir da natureza da pesquisa, do objeto, do tipo de abordagem, do universo da amostra, das fontes de referência, dos instrumentos de coleta de dados e da forma como esses dados serão processados, analisados e apresentados na discussão.

**Figura 35** - Caracterização da pesquisa

Natureza da pesquisa	• Pesquisa aplicada.
Natureza do objeto	• Exploratória e descritiva.
Abordagem	• Qualitativa.
Universo e amostra	• Unidades de informação especializadas em Música no Brasil.
Fontes de referência	• Profissionais responsáveis pelo vocabulário controlado do domínio da Música; Manuais técnicos e políticas de indexação; Normas para a elaboração e manutenção de tesouros ANSI/NISO (2005) e ISO (2011; 2013).
Coleta de dados	• Pesquisa bibliográfica (Busca Sistemática); Pesquisa documental; Pesquisa de campo.
Processamento, análise e apresentação dos dados	• Análise de Conteúdo; • Teoria Fundamenta em Dados; • Tabelas, quadros e figuras.

**Fonte:** Elaborado pelo autor (2021).

### 3.1 CARACTERIZAÇÃO DA PESQUISA

Por se tratar de uma pesquisa qualitativa, o método para a execução da proposta de pesquisa caracteriza-se como indutivo, pois parte da compreensão dos fenômenos particulares de processos sociais básicos para a inferência de uma verdade geral a partir da percepção de padrões (MARCONI; LAKATOS, 2003). Para fins deste estudo compreende-se as práticas colaborativas como um processo social, logo buscar-se-á analisar como essas práticas ocorrem na elaboração de tesouros para a partir dessas evidências difundir um modelo aplicado ao contexto do LOD.

Quanto à natureza do estudo é caracterizado pesquisa aplicada. Segundo Silva e Menezes (2005), esse tipo de pesquisa tem como foco a solução de um problema prático encontrado nas atividades de instituições, organizações, grupos ou atores sociais. A solução para esses problemas perpassa pelo diagnóstico inicial, que busca identificar inadequações e posteriormente mostrar soluções que respondam às demandas desses “atores sociais ou instituições” (THIOLENT, 2009).

No caso da proposta, a morosidade na construção e manutenção dos tesouros e a representação do conhecimento de domínio complexos visando atender a públicos heterogêneos são os problemas práticos encontrados. Assim, uma das etapas da pesquisa consiste em realizar um levantamento da situação atual e buscar soluções, especificamente para o contexto da publicação em LOD, tendo a área da Música como o domínio escolhido para a investigação. O conjunto de padrões, processos e tecnologias são demandas potenciais e serão sistematizados por um modelo fundamentado nas práticas colaborativas evidenciadas.

Por conseguinte, optou-se pela abordagem qualitativa visto o estudo exigir uma compreensão da relação dinâmica entre a prática colaborativa e a elaboração e manutenção de tesouro do domínio da Música – objetividade do fazer e a subjetividade do sujeito que o faz. Esse tipo de abordagem apresenta algumas particularidades básicas que se alinham ao objetivo deste estudo, entende que:

[...] um fenômeno pode ser melhor compreendido no contexto em que ocorre e do qual é parte, devendo ser analisado numa perspectiva integrada. Para tanto, o pesquisador vai a campo buscando “captar” o fenômeno em estudo a partir da perspectiva das pessoas nele envolvidas, considerando todos os pontos de vista relevantes. Vários tipos de dados são coletados e analisados para que se entenda a dinâmica do fenômeno (GODOY, 1995, p. 21).

O processo colaborativo na construção de tesouros consiste no fenômeno a ser investigado. Dessa maneira, para que os padrões, procedimentos e tecnologias relacionados a



essas práticas sejam analisadas sistematicamente é necessário compreendê-las a partir da visão dos responsáveis pela sua elaboração, os quais são qualificados como fontes de referência para a coleta de dados.

Em relação à natureza do objeto de estudo pesquisado, Marconi e Lakatos (2003) consideram que há basicamente três tipos de pesquisa: pesquisa exploratória, descritiva e experimental. Nesse sentido, realizou-se um estudo exploratório descritivo intentando levantar fatos e fenômenos abordados para posteriormente apresentá-los mediante observações sistemáticas.

A pesquisa exploratória objetiva esclarecer conceitos e ideias, a fim de formular problemas mais precisos que, por sua vez, poderão auxiliar no desenvolvimento de estudos posteriores. À vista disso, são menos rígidas visto proporcionar uma visão geral dos fenômenos estudados. Já a pesquisa descritiva busca detalhar um fenômeno, como as características de um indivíduo ou grupos de indivíduos entorno de uma situação, permitindo explorar sistematicamente a relação entre os eventos estudados (GIL, 2008).

### 3.2 UNIVERSO E AMOSTRA

O universo da pesquisa concerne às unidades de informação especializadas em Música no Brasil. A amostra é composta pelas unidades que constroem e mantêm seus próprios vocabulários controlados. Para a seleção das unidades empregou-se como fonte de informação o Guia de Acervos Musicais Brasileiros (2018), elaborado pelo extinto Comitê de Bibliotecários Catalogadores de Acervos Musicais (CBCAM) e hoje mantido pela Biblioteca da Escola de Comunicação e Artes da Universidade de São Paulo (ECA/USP). O Quadro 13 apresenta a lista de instituições que cooperam com o guia de acervos.

**Quadro 13** - Unidades de Informação especializadas em Música no Brasil

No.	Instituição	Tipo de Unidade
1	Arquivo Histórico do Museu Imperial. (Instituto Brasileiro de Museus – Ibram): <a href="http://www.museuimperial.gov.br">http://www.museuimperial.gov.br</a>	Arquivo
2	Arquivo Histórico Municipal de Salvador. (Secretaria de Cultura e Turismo de Salvador): <a href="http://www.cultura.salvador.ba.gov.br/">http://www.cultura.salvador.ba.gov.br/</a>	Arquivo
3	Arquivo Nirez <a href="http://nirez.ims.com.br">http://nirez.ims.com.br</a>	Acervo particular
4	Biblioteca Alberto Nepomuceno da Escola de Música da Universidade Federal do Rio de Janeiro: <a href="http://www.musica.ufrj.br">http://www.musica.ufrj.br</a>	Biblioteca Universitária
5	Biblioteca Carlos Barbosa do Instituto de Artes da Universidade Federal do Rio Grande do Sul: <a href="http://http://www.ufrgs.br/bibart/">http://http://www.ufrgs.br/bibart/</a>	Biblioteca Universitária

6	Biblioteca Central da Universidade de Brasília: <a href="http://www.bce.unb.br/">http://www.bce.unb.br/</a>	Biblioteca Universitária
7	Biblioteca Central Santa Mônica da Universidade Federal de Uberlândia – UFU: <a href="http://babao.dr.ufu.br:8080/">http://babao.dr.ufu.br:8080/</a>	Biblioteca Universitária
8	Biblioteca da Escola de Comunicações e Artes da Universidade de São Paulo: <a href="http://www3.eca.usp.br/biblioteca">http://www3.eca.usp.br/biblioteca</a>	Biblioteca Universitária
9	Biblioteca da Escola de Música da Universidade do Estado de Minas Gerais: <a href="http://www.uemg.br/biblioteca">http://www.uemg.br/biblioteca</a>	Biblioteca Universitária
10	Biblioteca da Escola de Música Villa-Lobos Espaço Maestro Alceo Bocchino. (Fundação de Artes do Estado do Rio de Janeiro): <a href="http://www.villa-lobos.org.br">http://www.villa-lobos.org.br</a>	Biblioteca Pública
11	Biblioteca DEMUS/DEART da Universidade Federal de Ouro Preto – Departamento de Música e Artes Cênicas: <a href="http://www.sisbin.ufop.br/">http://www.sisbin.ufop.br/</a>	Biblioteca Universitária
12	Biblioteca do Instituto de Artes, Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp): <a href="http://www.ia.unesp.br/">http://www.ia.unesp.br/</a>	Biblioteca Universitária
13	Biblioteca do Museu Imperial. (Museu Imperial/ Ibram / MinC): <a href="http://www.museuimperial.gov.br/palacio/biblioteca">http://www.museuimperial.gov.br/palacio/biblioteca</a>	Biblioteca Pública
14	Biblioteca Flausino Vale da Escola de Música da Universidade Federal de Minas Gerais – UFMG: <a href="http://www.musica.ufmg.br/biblioteca">http://www.musica.ufmg.br/biblioteca</a>	Biblioteca Universitária
15	Biblioteca Joaquim Cardozo do Centro de Artes e Comunicação da Universidade Federal de Pernambuco (UFPE): <a href="http://www.biblioteca.ufpe.br">http://www.biblioteca.ufpe.br</a>	Biblioteca Universitária
16	Biblioteca Mercedes Reis Pequeno da Academia Brasileira de Música: <a href="http://www.abmusica.org.br">http://www.abmusica.org.br</a>	Acervo particular
17	Biblioteca Pe. Jaime Diniz da Escola de Música da Universidade Federal do Rio Grande do Norte: <a href="http://bibliotecaemufrn.blogspot.com.br/">http://bibliotecaemufrn.blogspot.com.br/</a>	Biblioteca Universitária
18	Biblioteca Universitária Dr. Enoch de Oliveira da UNASP - Centro Universitário Adventista de São Paulo - campus Engenheiro Coelho: <a href="http://www.unasp-ec.edu.br">www.unasp-ec.edu.br</a>	Biblioteca Universitária
19	Centro de Documentação e Informação- Cedoc / Funarte: <a href="http://www.funarte.gov.br/cedoc/">http://www.funarte.gov.br/cedoc/</a>	Centro de documentação
20	Coordenação de Documentação de Música Contemporânea do Centro de Integração, Documentação e Difusão Cultural (CIDDIC) da Universidade Estadual de Campinas – Unicamp <a href="http://www.ciddic.unicamp.br/">http://www.ciddic.unicamp.br/</a>	Biblioteca Universitária
21	Discoteca Oneyda Alvarenga. (Prefeitura do Município de São Paulo): <a href="http://centrocultural.cc/discoteca">http://centrocultural.cc/discoteca</a>	Biblioteca Pública
22	Divisão de Música e Arquivo Sonoro da Fundação Biblioteca Nacional <a href="http://www.bn.br">http://www.bn.br</a>	Biblioteca Pública
23	Memorial Denis Bernardes da Biblioteca Central da Universidade Federal de Pernambuco: <a href="https://www.ufpe.br/sib/memorial">https://www.ufpe.br/sib/memorial</a>	Biblioteca Universitária

Fonte: Guia de Acervos Musicais Brasileiros (2021).

As unidades de informação que mantêm e tratam tecnicamente coleções musicais no Brasil variam conforme as características dos acervos e seus públicos. Estão distribuídas entre acervos particulares, arquivos, centros de documentação, bibliotecas públicas e bibliotecas universitárias públicas ou privadas. A maioria delas está localizada na região Sudeste do Brasil (17 instituições), e as demais situadas nas regiões Nordeste (4 instituições), Centro-oeste (1 instituição) e Sul (1 instituição).

Em um primeiro momento, os coordenadores dessas unidades foram contactados por meio de mensagens eletrônicas (*e-mail*) tencionando identificar quais delas constroem seus próprios vocabulários controlados. Essa etapa foi fundamental para selecionar a amostra. Nesse contato, os responsáveis foram indagados quanto ao uso de vocabulários controlados na indexação de seus acervos, buscando verificar se as unidades de informação:

- a) Utilizam estritamente vocabulários controlados elaborados e mantidos por outras instituições (por exemplo: Terminologias de Assuntos da Biblioteca Nacional, Thesaurus Brasileiro da Educação (Brased) ou quaisquer outras terminologias de assuntos que cubram o domínio da Música;
- b) Usam vocabulário controlado elaborado e mantido pela própria instituição;
- c) Não lidam com vocabulários controlados, a indexação é realizada a partir de termos livres selecionados pelo bibliotecário catalogador.

As respostas são apresentadas no Quadro 14 e permitiram perceber que a maioria das unidades empregam vocabulários controlados mantidos por outras instituições. Entre as terminologias se destacaram: a Terminologia de Assuntos da Biblioteca Nacional; o *Library of Congress Subject Headings*; o Thesaurus Brased; e o Vocabulário Controlado da USP.

**Quadro 14** - Respostas das UI em relação ao uso de vocabulários controlados na indexação

<b>Tipos de Unidades de Informação</b>	<b>Respostas</b>
<b>Arquivo</b>	<i>Sim [utilizam vocabulários mantidos por outras instituições], seguimos a Terminologia de Assuntos da Biblioteca Nacional.</i>
<b>Arquivo</b>	<i>Usamos alguns campos controlados criados por nós. Não é um vocabulário super estruturado como num tesouro, mas é controlado.</i>
<b>Biblioteca pública universitária</b>	<i>Sim [utilizam vocabulários mantidos por outras instituições]. Sempre pesquisamos os vocabulários controlado na Biblioteca Nacional, ou Library of Congress e instituições da área como referência. Ainda não possuímos [vocabulário próprio], mas pretendemos elaborar, em fase de construção.</i>
<b>Biblioteca pública universitária</b>	<i>No Sistema de Bibliotecas da [nome da instituição] os itens são indexados com base no vocabulário controlado da Biblioteca Nacional e da Rede</i>

	<i>[nome do sistema de catalogação]. Ainda não utilizamos vocabulários controlados mais específicos para os documentos musicais.</i>
<b>Biblioteca pública universitária</b>	<p>a) <i>Quando é importado o registro da OCLA, aproveita-se o assunto, traduz o termo e utilizamos o mesmo.</i></p> <p>b) <i>É utilizado o vocabulário controlado, as vezes a pesquisa é feito no PESQ, OCLC e Biblioteca Nacional.</i></p> <p>c) <i>Sim e não. As vezes o termo não é encontrado, então o bibliotecário utiliza-se de termo livre.</i></p>
<b>Biblioteca pública universitária</b>	<p><i>Terminologias e vocabulários utilizados para a indexação da Biblioteca [nome da instituição]:</i></p> <ul style="list-style-type: none"> <li>- <i>Terminologia de Assuntos da Biblioteca Nacional;</i></li> <li>- <i>Library of Congress Medium of Performance Thesaurus for Music PDF Files (Meio de expressão/Performance);</i></li> <li>- <i>Library of Congress Genre/Form Terms PDF Files (Assunto gênero/forma).</i></li> </ul>
<b>Biblioteca pública universitária</b>	<i>Utiliza estritamente vocabulários controlados elaborados e mantidos por outras instituições (por exemplo: Terminologias de Assuntos da Biblioteca Nacional, Thesaurus Brasileiro da Educação (Brased), ou quaisquer outras terminologias de assuntos que cubram o domínio da música).</i>
<b>Biblioteca pública universitária</b>	<i>A biblioteca da [nome da instituição] utiliza vocabulário controlado da Biblioteca Nacional e da Library of Congress na indexação do acervo.</i>
<b>Biblioteca pública universitária</b>	<i>O que temos aqui na [nome da instituição] é uma política de indexação, que nos dá algumas orientações com relação ao uso de descritores, mas não possuímos vocabulário controlado.</i>
<b>Biblioteca pública universitária</b>	<i>[...] possuímos uma base de autoridade (assunto e autoria) para toda a rede de Bibliotecas da Universidade. Este vocabulário foi inicialmente construído pelo Grupo de Linguagem [nome da instituição], composto por bibliotecários de várias áreas do conhecimento que trabalham aqui. Após um período os demais bibliotecários receberam treinamento e hoje todos podem inserir assuntos nessa base de autoridade, SEMPRE readequando registros importados de outras linguagens (BN, LCSH, MeSH, Biblioteca Britânica, Vocabulário USP) e utilizando manual elaborado pelo Grupo. Para assuntos não encontrados em nenhuma destas bases citadas, é feito um pedido para o Grupo, que fará estudos para inclusão ou não do novo termo.</i>
<b>Biblioteca pública universitária</b>	<i>[utilizamos] BN [Biblioteca Nacional], LC [Library of Congress], Rede [nome do sistema de gestão] e Bibliodata que no momento está inoperante. Nenhuma adaptação [é realizada nos termos], conforme consta nas bases usamos igual.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Com base nos *feedbacks*, foi possível conceber que as políticas de indexação desenvolvidas em âmbito local orientam o uso de termos controlados, e que o uso de termos livres também é possível, principalmente quando o indexador não encontra um termo correspondente nos vocabulários mantidos por outras instituições. Para aquelas unidades que

elaboram seus próprios vocabulários, a construção é efetuada por profissionais ou especialistas que se organizam em grupos por área de conhecimento.

Em vista disso, foram selecionadas apenas as UI que afirmaram construir e manter seus próprios vocabulários controlados, resultando em um total de três instituições. Após selecioná-las, os profissionais responsáveis pela manutenção dos vocabulários do domínio da Música foram identificados e convidados a participar da pesquisa.

### 3.3 COLETA DOS DADOS

Para a coleta de dados foram aplicados os métodos de pesquisa bibliográfica, a pesquisa documental e o questionário semiestruturado.

#### 3.3.1 Pesquisa Bibliográfica

A pesquisa bibliográfica pode ser considerada um tipo de revisão da produção científica sobre os principais temas abordados em determinada área do conhecimento. Por meio de levantamento de estudos publicados busca-se analisar as principais contribuições teóricas da área estudada com o propósito de subsidiar a investigação (PIZZANI et al., 2012).

Neste estudo, o método de pesquisa bibliográfica foi utilizado para coletar trabalhos relacionados ao tema. Desse modo, foram executadas buscas sistemáticas em bases de dados nacionais e internacionais das áreas da Ciência da Informação e Ciência da Computação, especificadas no Apêndice A. Com a pesquisa procurou-se identificar padrões e tecnologias empregados para a construção e publicação de tesouros no contexto do LOD. Os resultados dessas buscas seguem são lançados no Quadro 15.

**Quadro 15** - Resultados das buscas nas bases de dados

Expressões de busca	Bases de dados	Resultados	Relevantes	Selecionados
Thesaurus and "Linked Open Data"; Thesaurus and "collaborative management";	<i>Web of Science (WOS)</i>	29	23	17
	Scopus	64	19	12
	<i>Library and Information Science Abstracts (LISA)</i>	115	36	7
	<i>Science Direct</i>	167	31	14

Thesaurus and "collaborative method";	IEEE	14	9	4
Thesaurus and "collaborative construction";	<i>Networked Digital Library of Theses and Dissertations</i> (NDLTD)	177	23	3
Thesaurus and SKOS;	Biblioteca Digital de Teses e Dissertações	3	1	1
Tesauros e "Linked Open Data";	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI)	7	5	2
Tesauros e SKOS;				
Tesauros e "construção colaborativa".				
Busca por título	<i>Google Scholar</i> <sup>48</sup>	-	14	7
<b>Total</b>		568	161	67

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Foram feitas as leituras dos títulos e resumos dos estudos e selecionados apenas aqueles diretamente relacionados à temática. Desse modo, escolheu-se 67 estudos, entre eles artigos publicados em periódicos revisados por pares, artigos publicados em eventos científicos, capítulos de livros e teses. O Apêndice D apresenta a lista completa dos estudos analisados.

### 3.3.2 Questionário semiestruturado

O instrumento para a coleta dos dados foi o questionário semiestruturado. A partir de questões estruturadas em tópicos foram coletadas informações sobre a prática cotidiana profissional, observando os aspectos de cooperação, coordenação e comunicação nas atividades inerentes ao processo de construção dos vocabulários controlados (cf. Apêndice C).

Os dados foram coletados a partir de relatos dos Bibliotecários<sup>49</sup> responsáveis pelo controle terminológico dos vocabulários controlados mantidos pelas unidades de informação selecionadas. Esses profissionais são qualificados nesta pesquisa como fontes de referência, pois a partir de suas experiências com a terminologia da Música puderam oferecer um

<sup>48</sup> A busca por título realizada no *Google Scholar* objetivou recuperar as publicações consideradas relevantes encontradas nas bases de dados ou nas referências citadas da bibliografia selecionada, seja porque não foram identificadas com as buscas sistemáticas, seja porque não puderam ser acessadas integralmente a partir das bases de dados.

<sup>49</sup> Submeteu-se a pesquisa ao Comitê de Ética em Pesquisa (CEP) da Universidade Federal de Santa Catarina, a qual foi aprovada conforme o parecer final de número 3.910.772, de 11 de março de 2020.

diagnóstico de como essas unidades de informação elaboram e mantêm seus vocabulários controlados. Buscou-se, ainda, verificar se as unidades têm incorporado tecnologias do *Linked Data*.

O uso do questionário se justifica pela necessidade de se obter informações de forma rápida sem a incumbência de o pesquisador se deslocar geograficamente, haja vista que o universo e amostra do estudo são formados por instituições situadas em diferentes regiões do país<sup>50</sup>. Contudo, as perguntas abertas e fechadas desse instrumento contribuíram para a coleta exata de dados e, igualmente, favoreceu a liberdade de respostas em alguns quesitos (MARCONI; LAKATOS, 2003).

### 3.3.3 Pesquisa documental

A pesquisa documental utiliza fontes primárias, ou seja, documentos que não foram tratados sistematicamente por nenhum autor (SÁ-SILVA; ALMEIDA; GUINDANI, 2009). Esse tipo de pesquisa na investigação se mostra adequado, porque busca mediante a organização e tratamento do material extrair informações relevantes em normas técnicas, manuais e políticas de indexação.

Assim, a análise foi efetuada nas normas técnicas para a construção de tesouros monolíngues e multilíngues. A saber:

- a) ANI/NISO Z39.19-2005 (R2010): Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies;
- b) ISO 25964-1: Information and Documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval;
- c) ISO 25964-2: Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies.

Foram analisados também manuais e políticas de indexação disponibilizados pelas instituições que participaram desta pesquisa. Os títulos e as fontes não serão especificados a fim de manter o anonimato dessas instituições. Os documentos revelam detalhes de como os vocabulários controlados são construídos em função da tarefa de indexação, bem como foram operados para identificar os procedimentos, métodos utilizados, fluxo de tarefas e a existência de práticas colaborativas entre os membros da equipe de gestão dos vocabulários controlados.

---

<sup>50</sup> Vale ressaltar que os dados foram coletados durante a pandemia da COVID19, entre abril e junho de 2020, inviabilizando outros tipos de coleta, como entrevistas presenciais, por exemplo.

### 3.4 ETAPAS DA PESQUISA

A pesquisa foi distribuída em quatro etapas, estruturadas consoante com os objetivos específicos. Conforme as informações dispostas no Quadro 16, cada objetivo resultou em um produto final, sendo eles: a descrição dos principais atores sociais, suas tarefas, ações ou fluxos de trabalho, a estruturação dos princípios, a representação do modelo e a avaliação com especialistas.

**Quadro 16** – Etapas da pesquisa

<b>Objetivos</b>	<b>Coleta dos dados</b>	<b>Análise dos dados</b>	<b>Resultados</b>
Identificar práticas colaborativas na elaboração de vocabulários controlados do domínio da Música em Unidades de informação especializadas no Brasil	Questionário e Pesquisa documental	Análise de conteúdo	Descrição e relação dos principais atores sociais, procedimentos e fluxos de trabalho
Estruturar, com base na literatura, os princípios que norteiam o processo de construção colaborativa e publicação de tesouros como <i>Linked Open Data</i>	Pesquisa bibliográfica	Análise de conteúdo	Descrição e relação dos princípios
Representar, por meio de um modelo, os procedimentos e tecnologias voltadas para a construção colaborativa e publicação de tesouros do domínio da Música no contexto do <i>Linked Open Data</i>	Questionário, Pesquisa documental e Pesquisa bibliográfica	Teoria Fundamentada em Dados	Sistematização dos elementos (categorias de ações, propriedades e conexões entre os tecnologias e tarefas)
Avaliar o modelo com especialistas da área	Formulário para a avaliação	Análise de conteúdo	Aprimoramento e definição do modelo

**Fonte:** Elaborado pelo autor (2021).

Para cada objetivo aplicou-se um tipo de método de coleta e análise de dados. As opções de análise, fases, etapas e procedimentos implementados durante o tratamento dos dados são especificadas na seção seguinte.



### 3.5 ANÁLISE E TRATAMENTO DOS DADOS

Para a análise e processamento dos dados aplicou-se o método da Teoria Fundamentada em Dados (TFD), o qual contribuiu com a elaboração do modelo detectando semelhanças, diferenças e padrões evidenciados na elaboração de tesouros para o contexto do LOD. Segundo Roncato (2006, p. 2379), a TFD consiste em:

[...] um método indutivo de pesquisa qualitativa, tem como objetivo explicar fenômenos sociais e psicológicos, uma teoria ou modelo conceitual é construído através da análise de dados obtidos sistematicamente a partir de entrevistas sobre vivências concretas ou anotações detalhadas de observações.

O objetivo da TFD é “[...] gerar uma teoria ou explicação teórica unificada para um processo, uma ação ou uma interação moldada por uma visão de um grande número de participantes expressa nos dados coletados.” (SANTOS et al., 2016, não paginado). Assim, justifica-se a escolha do método tendo em vista que a TFD busca gerar um construto, modelo ou matriz teórica para explicar uma ação no contexto social.

Neste estudo considera-se a prática colaborativa como um comportamento social, pois lida com a cooperação, comunicação e coordenação de atividades inerentes ao controle terminológico e conta com a participação de diversos atores. Além disso, o método é conhecido pelo seu rigor ao especificar as etapas e procedimentos aplicados na coleta e saturação teórica. Tal abordagem tende a superar limitações metodológicas de estudos qualitativos, na maioria das vezes representado por estudos de caso (PETRINI; POZZEBON, 2009).

O ponto de partida para a construção da TFD são os dados. A reunião desses dados e posterior análise dão origem aos conceitos construídos. Dessa maneira, os pesquisadores reúnem dados para elaborar as análises teóricas e descobrir o que ocorre nos ambientes de pesquisa. Especificamente, são verificadas as formas como os enunciados e as ações são explicados e como podem ser compreendidos analiticamente (CHARMAZ, 2009, p. 15).

Os dados coletados nas etapas anteriores foram interpretados e tratados sistematicamente mediante procedimento chamado de codificação, o qual implica em associar marcadores e segmentos de dados agrupando-os em áreas ou temas de interesse para que posteriormente sejam estabelecidas comparações entre eles (CHARMAZ, 2009). O método da TFD apresenta diretrizes explícitas, especificando como deve ocorrer a coleta e análise.

Isto posto, durante a codificação foi possível comparar eventos relatados pelo sujeito em consonância com os códigos estabelecidos pelo pesquisador ao analisar a amostra teórica. Concomitantemente, fez-se anotações analíticas, chamadas de memorandos, para facilitar a

categorização e análise sistemática dos relatos tratados. (STRAUSS; CORBIN, 2008; CHARMAZ, 2009). As fases, etapas e procedimentos básicos para a aplicação da TFD nesta pesquisa são apresentadas no Quadro 17.

**Quadro 17** –Fases, etapas e procedimentos para a análise e tratamento dos dados conforme a TFD

Fases	Etapas	Procedimentos básicos
Delineamentos da pesquisa	Etapa 1 – Pesquisa bibliográfica (Buscas Sistemáticas)	a) Definição da questão de pesquisa; b) Definição de critérios de inclusão e exclusão para os estudos.
	Etapa 2 - Seleção dos estudos	c) Análise preliminar da amostra teórica.
	Etapa 3 – População e amostra	d) Selecionar as unidades de informação especializadas em Música no Brasil.
	Etapa 4 – Instrumento de coleta	e) Elaboração do questionário semiestruturado.
Coleta de dados	Etapa 5 – Pesquisa de campo e documental	f) Selecionar os atores sociais fundamentais (principais responsáveis); g) Registrar os relatos completos e detalhados sobre ações individuais e coletivas; h) Utilizar fontes de dados auxiliares (manuais de procedimentos, instruções normativas).
Ordenação dos dados	Etapa 6 – Ordenar os dados	i) Preparação dos dados para o tratamento primário.
Análise de dados	Etapa 7 – Sistematização	j) Codificar os dados.
	Etapa 8 – Confrontar com a amostra teórica	k) Sobrepor amostra teórica e os dados obtidos com os questionários (voltar a etapa 2c).
	Etapa 9 – Fechamento	l) Elaboração e saturação do modelo.
Avaliação	Etapa 10 – Avaliação do modelo com especialistas.	m) Refinamento do modelo.

**Fonte:** Elaborado pelo autor (2021).

Quanto aos instrumentos que auxiliaram nos tratamentos dos dados lidou-se com os seguintes *softwares* aplicativos:

- a) **Mendeley:** gerenciador de referências empregue para a organização da bibliografia (amostra teórica) e da documentação (normas, manuais e políticas de indexação);
- b) **NVivo:** programa para análise de informação qualitativa usado na preparação do *corpus* teórico, exploração do material, codificação, análise de conteúdo, elaboração de diagramas e figuras (árvore de palavras, nuvem de palavras, dendograma e gráficos);

- c) **Microsoft office (Power Point e Word):** executados para a edição de texto, para a elaboração de quadros, tabelas, diagramas ou figuras.

### 3.6 AVALIAÇÃO DO MODELO

A última fase da pesquisa consistiu na avaliação do modelo com especialistas. A avaliação é um instrumento de validação considerada uma estratégia metodológica que consiste “[...] em comparar conceitos e determinar o quanto eles são apropriados para a investigação que foi desenvolvida” (SOUZA; SILVA, 2011, p. 780). À vista disso, pretendeu-se verificar a consistência da proposta.

Para a seleção dos especialistas observou-se dois critérios fundamentais: experiência e produção científica relevante na área. Assim, participaram da fase de avaliação seis especialistas da área de Organização do Conhecimento e Representação da Informação, dentre os quais cinco são docentes de universidades públicas – todos com nível de doutorado – lecionam disciplinas relacionadas às áreas mencionadas anteriormente ou têm experiência na construção de vocabulários controlados na *web*, elaboração de tesouros e tecnologias do *Linked Data*. No Quadro 18 são apresentados os perfis de cada participante.

**Quadro 18 - Perfil dos especialistas que avaliaram o modelo**

<b>Especialistas</b>	<b>Formação, atuação profissional e áreas de interesse de pesquisa</b>
Especialista 1	Graduação em Biblioteconomia. Possui mestrado e doutorado em Ciência da Informação. Leciona em Universidade pública federal, atuando nos cursos de graduação em Biblioteconomia e na pós-graduação em Gestão e Organização do Conhecimento. Faz parte de grupos de pesquisa na temática de representação do conhecimento e recuperação da informação. Tem interesse na linha de Arquitetura e Organização do Conhecimento, com ênfase em questões teóricas e metodológicas de desenvolvimento de Sistemas de Organização do Conhecimento (sistemas de classificação, taxonomias, tesouros e redes semânticas); serviços informacionais (bibliotecas tradicionais, digitais, hipertexto, arquivos); integração de dados e compatibilidade de linguagens; segurança de dados e informação.
Especialista 2	Graduação em Arquitetura. Possui mestrado e doutorado em Ciência da Informação. Leciona em universidade pública federal, atuando como professor titular do Departamento de Ciência da Informação. Tem experiência na área de Ciência da Informação, com ênfase em Organização e representação do conhecimento em ambientes digitais, atuando principalmente nos seguintes temas: fundamentos teóricos da Ciência da Informação e Organização do Conhecimento, publicações semânticas, ontologias, tecnologias semânticas e de dados abertos interligados, bibliotecas e museus digitais e interoperabilidade.

Especialista 3	Graduação em biblioteconomia. Possui mestrado e doutorado em Ciência da Informação. Leciona em universidade pública federal. Participa de grupo de pesquisa sobre folksonomia e sistemas híbridos. Membro do comitê de <i>Machine Learning da International Association of Artificial Intelligence</i> (Associação Internacional de Inteligência Artificial). Tem interesse e desenvolve pesquisas nas áreas: Organização e representação da informação e do conhecimento; recuperação da informação; inteligência artificial; modelos colaborativos de indexação; folksonomia; política de indexação híbrida; sistemas de organização do conhecimento; arquitetura da informação; bibliotecas digitais; repositórios digitais; bases de dados; gestão de unidades de informação; produtos e serviços de informação.
Especialista 4	Graduação em Biblioteconomia. Possui mestrado e doutorado em Ciência da Informação. Atua como bibliotecário em universidade pública federal. Realiza estudos na área de catalogação descritiva com ênfase na conversão de registros, controle de autoridade, Formatos MARC 21, MARCXML, RDA, modelos conceituais FRBR e FRAD e <i>Linked Data</i> .
Especialista 5	Graduação em Biblioteconomia e Ciência da Informação. Possui mestrado em Linguística, e mestrado em Ciência, Tecnologia e Sociedade. Atualmente, cursa doutorado em Ciência da Informação. Leciona em universidade pública federal, atuando em Departamento de Ciência da Informação. Leciona disciplinas de linguística documentária, tesauros e ontologias e tratamento temático da informação. Tem experiência com os temas: Mediação Cultural, Estudos métricos em informação, Tecnologias e Inovação, Produtos e serviços de informação e Tratamento da Informação.
Especialista 6	Graduação em Biblioteconomia. Mestre em Ciência da Informação. Atualmente cursa o doutorado em Ciência da Informação. Leciona em Universidade pública federal, no curso de Biblioteconomia e Gestão de Unidades de Informação. Participa de grupo de pesquisa na área de Organização e representação do conhecimento. Desenvolve pesquisas e tem interesse nas áreas de: Organização e Representação da Informação e do Conhecimento; A imagem na Ciência da Informação; Comunicação Científica; Dados de pesquisa; e Estudos métricos.

**Fonte:** Elaborado pelo autor com base no *Currículo Lattes* (2021).

Na próxima seção serão apresentados os resultados das duas primeiras etapas da pesquisa. A saber: pesquisa de campo com a análise dos questionários; e pesquisa bibliográfica com a apreciação dos estudos coletados nas bases de dados.

## 4 ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS

Para o cumprimento dos objetivos específicos a e b desta pesquisa serão apresentados no capítulo em realce as seguintes ações: a) análise dos relatos dos gestores para identificar práticas colaborativas no controle de vocabulários do domínio da Música; b) estruturação dos princípios que norteiam a construção colaborativa e a publicação de tesouros no contexto do LOD.

### 4.1 PRÁTICAS COLABORATIVAS NA CONSTRUÇÃO DE VOCABULÁRIOS DO DOMÍNIO DA MÚSICA

As análises produzidas nesta seção foram realizadas com base nos manuais, políticas de indexação ou metodologias disponibilizadas pelas instituições e questionários aplicados aos gestores dos vocabulários controlados do domínio da Música.

Seguindo os princípios éticos da pesquisa, que visa o sigilo dos responsáveis técnicos e das instituições, os nomes destes não serão revelados. Assim, no Quadro 19, seguem tratados pelas abreviações: Vocabulários controlados (VC1, VC2, VC3); Gestores do vocabulário (G1, G2, G3).

**Quadro 19** – Vocabulários controlados, gestores e os tipos de unidades de informação que atuam

Vocabulários controlados (VC)	Gestores dos vocabulários (G)	Tipos de Unidades de informação
VC1	G1	<i>Biblioteca Pública</i>
VC2	G2	<i>Biblioteca Pública Universitária</i>
VC3	G3	<i>Biblioteca Pública Universitária</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020)

Na sequência serão discutidas as respostas obtidas com as referidas unidades de informação. Os quadros que possuem as menções aos vocabulários controlados (VC) apresentam os resultados da análise documental (manuais e políticas de indexação), e os quadros com as menções aos gestores de vocabulários (G) apresentam as análises dos dados obtidos com os questionários.

#### 4.1.1 Caracterização das Unidades de Informação: acervo e sistemas de gestão

As Unidades de Informação (UI) que compõem a amostra da pesquisa estão localizadas na Região Sudeste do Brasil e atendem a um público majoritariamente universitário, composto por usuários internos – docentes e discentes – e usuários externos às instituições. Suas coleções são tratadas tecnicamente com o objetivo de fornecer acesso à informação musical ou preservar a memória cultural regional ou nacional. Ainda, estão disponíveis para a consulta de usuários locais, pesquisadores externos e público em geral.

No que diz respeito aos acervos, são constituídos de partituras em impressão padrão ou manuscritas, discos de vinil, discos compactos (CD) e gravações de vídeos que cobrem as principais áreas da música popular e erudita, brasileira ou estrangeira, com destaque para as coleções digitais que disponibilizam partituras e áudios em formatos MID e MP3. Todos eles estão disponíveis para a consulta no catálogo *online* e o acesso ao conteúdo digital é gratuito.

Por conseguinte, constatou-se que a gestão dos vocabulários controlados pode ocorrer por meio do mesmo sistema de gestão do acervo. Apenas o VC3 utiliza um sistema específico para gerenciar seu próprio vocabulário, desenvolvido com a ajuda de profissionais da equipe de gestão relacionados ao setor de desenvolvimento de sistemas.

Os *softwares* prioritários apontados pelos Gestores oferecem suporte à manutenção de suas coleções mediante bases administrativas, bibliográficas e de autoridades que realizam a catalogação, controle de aquisições, doações, empréstimos e devolução, além de fornecer um canal de comunicação com os usuários através dos catálogos *online* em um terminal *web*. Todos eles indicam um módulo exclusivo para o registro de autoridades que se baseia em padrões internacionais de catalogação e comunicação dos dados, como o MARC21, ISO 2709<sup>51</sup>, Z39.50<sup>52</sup>, XML e *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH)<sup>53</sup>.

O sistema desenvolvido pela própria instituição, apontado pelo VC3, é aplicado estritamente para as atividades de registro e compatibilização dos termos do vocabulário. O aplicativo se mostra integrado ao sistema de gestão do acervo, no qual é também possível realizar buscas no catálogo *online* da instituição.

---

<sup>51</sup> Norma que especifica o formato padrão para o intercâmbio de registros bibliográficos em meio magnético entre sistemas (INTERNATIONAL STANDARD ORGANIZATION, 2008).

<sup>52</sup> Protocolo cliente-servidor de padrão internacional para a pesquisa e recuperação da informação em redes de computadores distribuídos (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2014).

<sup>53</sup> Protocolo desenvolvido pela *Open Archives Initiative* que define um mecanismo para a coleta de registros de metadados em repositórios (OPEN ARCHIVES INITIATIVE, 2015).

A próxima subseção abordará os aspectos gerenciais e as metodologias empregadas, trazendo informações de quem são os atores sociais envolvidos na construção, como os vocabulários controlados surgiram nessas instituições e são estruturados.

#### 4.1.2 Aspectos gerenciais e metodológicos dos vocabulários controlados

Os vocabulários controlados apresentados nesta pesquisa se apresentam estruturados de diversas maneiras. As instituições os denominaram como listas ou terminologias de assunto e linguagens de indexação. Em suas páginas eletrônicas são disponibilizados manuais e políticas de indexação que descrevem brevemente os procedimentos e as metodologias empregadas em sua construção. Assim, com base nos enunciados e nos documentos oficiais disponíveis na *web*, o Quadro 20 destaca os principais pontos sobre a origem dessas linguagens.

**Quadro 20** - Origem dos vocabulários controlados

<b>Vocabulários controlados</b>	<b>Origem do vocabulário</b>
VC1	Foi idealizada com o objetivo de controlar a terminologia dos pontos de acesso de assunto e só foi efetivada após a adesão da UI à Rede Bibliodata. O vocabulário controlado segue a <i>Library of Congress Subject Headings (LCSH)</i> e contempla uma lista multidisciplinar de assuntos. Os cabeçalhos de assuntos em inglês são acrescentados, o que o torna o catálogo de assuntos bilíngue. Considera o aspecto de manter registros em inglês como uma característica importante da terminologia, pois contribui com a interoperabilidade linguística.
VC2	Utilizava diferentes terminologias até optar por uniformizar as linguagens com a finalidade de compatibilizar o vocabulário empregue na indexação e na estratégia de busca, oferecendo maior precisão na recuperação da informação.
VC3	Desenvolvida para a representação temática dos seus acervos. No início, as unidades utilizavam terminologias especializadas de cada área tentando uniformizar a linguagem. Foi necessário desenvolver uma linguagem de tratamento comum para o sistema de gestão do acervo.

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

É possível perceber que as iniciativas de construção dos vocabulários controlados nas respectivas instituições se originaram da necessidade de desenvolver uma linguagem comum de representação que contemplassem todas as áreas do conhecimento atendidas pelas instituições e servissem a um repertório de indexação mais específico ou mais especializado.

Isto porque as unidades de informação tradicionalmente utilizavam duas ou mais linguagens mantidas por outras instituições, observadas suas limitações ou divergências, optando por adaptar ou construir seus próprios vocabulários.

A participação de diferentes atores sociais é concebida na descrição de suas metodologias, em que há a atuação majoritária dos bibliotecários que compõem os sistemas de bibliotecas das instituições. Esses profissionais parecem executar as funções de forma compartilhada, designados em exercício da área de conhecimento das unidades que atuam. Com base nessas descrições, o Quadro 21 mostra que as equipes técnicas são formadas basicamente por bibliotecários, docentes das instituições, colaboradores externos e analistas de sistema.

**Quadro 21** - Equipe técnica que atua na construção dos vocabulários controlados

<b>Vocabulários controlados</b>	<b>Equipe técnica</b>
VC1	Bibliotecários que atuam no setor de catalogação e classificação das diferentes divisões especializadas da unidade.
VC2	Bibliotecários que realizam a atividade de catalogação e indexação na rede de bibliotecas da instituição; Pesquisadores da área de Indexação que atuam como docentes na instituição; Pesquisadores e colaboradores externos à instituição.
VC3	Bibliotecários das várias unidades especializadas da instituição; Docentes e pesquisadores especializados em Análise Documentária e Terminologia; Docentes colaboradores das áreas de conhecimento do Vocabulário; Analistas de sistemas do departamento técnico da instituição.

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Foi exequível identificar os papéis dos membros da equipe técnica. O que se sabe, basicamente, é que os profissionais bibliotecários atuam na construção de macroestruturas, solicitações e análises das sugestões de termos e no registro destes nos respectivos sistemas de gestão. Quanto aos docentes e pesquisadores podem ser: docentes colaboradores das áreas de conhecimento do vocabulário (especialistas do domínio) nas etapas de construção; e docentes pesquisadores que colaboram dando suporte metodológico na etapa de planejamento. Estes docentes são da área de Análise documentária e Terminologia e auxiliam aqueles profissionais a desenvolver um conhecimento teórico-metodológico para aprimorar suas competências e aplicar os métodos.

Nesses vocabulários controlados os termos são estruturados de diferentes maneiras, sejam por meio de registros de autoridades de assuntos, expressos por indicadores de relacionamentos padronizados, como na estrutura dos tesouros, sejam na forma de



macroestruturas, como uma lista por ordem alfabética, hierárquica ou organizada por gênero e forma. Com base nas instruções normativas de cada instituição, o Quadro 22 evidencia as principais características dos vocabulários em função de sua estrutura de termos.

**Quadro 22** – Estruturação dos termos nos vocabulários controlados

Vocabulários controlados	Estruturação dos vocabulários controlados
VC1	Está estruturado na forma de <b>tesauros</b> , apresentando para cada assunto os termos gerais (TG), os termos específicos (TE) e os termos relacionados (TR). Engloba assuntos, tópicos, remissivas (ver e ver também), além das subdivisões gerais, cronológicas e geográficas.
VC2	Organizado em registros de autoridades em três tipos: Registro de cabeçalho autorizado - contém um cabeçalho autorizado e campos de referência cruzada para cabeçalhos variantes ou relacionados e notas que registra informações sobre o uso e o estabelecimento dos termos; Registro de subdivisão - inclui um cabeçalho parcial não autorizado como cabeçalho tópico e deve ser utilizado somente como subdivisão de um cabeçalho autorizado; Registro de cabeçalho autorizado e de subdivisão - cabeçalho autorizado que também pode ser usado como subdivisão.
VC3	As hierarquias de termos foram definidas por áreas do conhecimento e agrupadas conforme a Macroestrutura; organizado por listas de assunto em ordem alfabética, hierárquica, ou gênero e forma.

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Verificou-se que apenas o VC1 se diz estruturado na forma de tesauros, cujos relacionamentos associativos, hierárquicos e de equivalência entre os termos são expressos por meio de indicadores padronizados, formando uma rede semântica de conceitos. Nele é possível observar notas de escopo que indicam o uso adequado do termo e auxiliam a indexação e a recuperação da informação.

Os VC2 e VC3 são menos dinâmicos, visto não especificarem os níveis associativos e de equivalência entre os termos. Apenas o VC3 disponibiliza a lista de assuntos na forma de estruturas hierárquicas de conceitos, agrupando-os em áreas temáticas e suas respectivas subáreas. Já o VC2 foi organizado na forma de registros de autoridades em função da política de indexação da instituição e, por isso, não disponibiliza ao público uma versão estruturada. Este último afirma estar trabalhando na construção de sua macroestrutura e na hierarquização dos termos; pretende adotar um *software* para agilizar a construção e disponibilizar uma ferramenta que permita ao usuário visualizar e interagir com a linguagem mediante interface de busca.

O Quadro 23 lança os aspectos metodológicos básicos que incluem as etapas de estruturação da macroestrutura, análise das solicitações, adequação dos registros, estabelecimento das relações semânticas e desambiguação.

**Quadro 23** - Percurso metodológico na construção dos vocabulários controlados

Unidades de Informação	Percurso metodológico
VC1	a) As obras são encaminhadas à divisão técnica para proceder com a classificação e a seleção da melhor forma de entrada para a autoridade de assuntos; b) Os termos escolhidos são verificados junto ao LCSH e as diversas fontes de referência; c) Os bibliotecários recebem dos classificadores as solicitações de inclusão de termos no catálogo de assuntos; d) Os bibliotecários decidem sobre a inclusão de termos que estão em conformidade com os existentes na LCSH ou adaptam um assunto específico que lá não seja tratado.
VC2	Utiliza o método terminográfico para: a) Adequar registros inicialmente importados do BIBLIODATA. Consiste em verificar os campos a serem contemplados ou incluídos; b) Criar registros para termos novos a partir de solicitações encaminhadas por meio de formulário eletrônico ao grupo gestor do vocabulário. Os termos novos são sugeridos quando não são encontrados nas linguagens BN, LCSH e MeSH.; c) Conceber e adequar registros novos para termos importados das linguagens BN, LCSH e MeSH.
VC3	a) Elaboração da estrutura temática a partir de descritores existentes nos catálogos e de diversos tipos de fontes de referências: tesouros existentes; dicionários especializados; e grupos de pesquisas da universidade; b) Inclusão dos blocos de assuntos em ordem hierárquica no sistema de gestão do vocabulário; c) Estabelecimento das relações lógico-semânticas entre os termos; d) Definição dos termos ambíguos (em ficha terminológica) e compatibilização das estruturas temáticas com as áreas complementares.

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Fundamentado nas breves descrições dos percursos metodológicos, é prudente afirmar que a construção desses vocabulários controlados ocorre por meio de grupos de gestão, no qual fazem parte profissionais que atuam nas diferentes unidades especializadas das referidas instituições. Os sujeitos que operam na gestão dos vocabulários podem estar organizados na

forma de subgrupos por áreas temáticas, são responsáveis pela análise das sugestões de termos candidatos e pela decisão de incluir ou alterar termos nos seus respectivos vocabulários.

Vale destacar que todos os vocabulários controlados ressaltaram que a inserção ou a mudança dos termos estão condicionadas à pesquisa em vocabulários mantidos por outras instituições. Nesse sentido, destaca-se o papel das terminologias já consolidadas da área e a preocupação dos gestores em compatibilizar os termos em função das estruturas temáticas de cada vocabulário. Essa iniciativa permite considerar termos amplamente aceitos pela comunidade científica e preencher lacunas não contempladas por essas terminologias existentes, especialmente no tocante aos conceitos que exigem maior especificidade ou adaptação à realidade de uso local.

As informações coletadas demonstraram que as origens e as metodologias dos vocabulários controlados são marcadas pela necessidade de uniformizar as linguagens de indexação. Ademais, estão condicionadas à eficiência de terminologias consolidadas, enquanto que os membros das equipes de controle atuam de maneira descentralizada e em função da especialização dos acervos dessas instituições.

Por conseguinte, objetivando analisar informações mais detalhadas sobre os procedimentos, recursos e tecnologias aplicadas à construção dos vocabulários controlados, na sequência serão discutidos os dados obtidos mediante questionário. As seções seguintes tratam dos: aspectos gerais acerca da manutenção e elaboração; como ocorrem as etapas de compilação ou levantamento, especificação ou supressão, registro e validação de termos do domínio da Música; como sucede a fase de testes e avaliação do vocabulário, as ferramentas colaborativas empregues e a representação desses vocabulários na *web*; e a percepção da colaboração dos gestores nas etapas de construção dos vocabulários controlados.

#### **4.1.3 Elaboração e manutenção do vocabulário controlado**

Enquanto ferramentas de representação construídas para atender demandas de informação dos usuários, os vocabulários controlados precisam estar constantemente por manutenção. Diante disso, ao serem indagados sobre a frequência de atualização de seus vocabulários verifica-se que todos os participantes afirmaram que esse processo é contínuo, executado diariamente pela equipe de construção, apenas o G2 forneceu mais detalhes:

*Diariamente. No início o vocabulário era alimentado pelo grupo de catalogação da rede de bibliotecas, hoje os bibliotecários conseguem inserir assuntos no vocabulário também. (G2)*

Esse gestor afirma que a ação é sucedida pelo grupo de catalogação da rede de bibliotecas, ou seja, havia um grupo de profissionais designados especificamente para executar a função. Contudo, prossegue declarando que atualmente todos os bibliotecários do sistema de bibliotecas podem atualizar a estrutura do vocabulário. Isso implica na possibilidade de alterações a qualquer momento, desde que observada a real necessidade e relevância do termo sugerido.

A participação de diversos profissionais na elaboração é considerada por muitos estudiosos uma alternativa positiva, pois a seleção de termos e também sua categorização em uma estrutura de classes conceituais pode envolver subjetividade. Nesse ponto, a participação efetiva de diversos profissionais e especialistas pode ajudar a estabelecer acordos e promover o consenso quanto às decisões de inclusão ou exclusão de termos (NOHAMA et al., 2012; ÇAĞDAŞ; STUBKJÆR, 2015; STELLATO et al., 2015). Todavia, a efetividade nessa ação pode estar condicionada à experiência dos membros da equipe de construção com o domínio de especialidade.

Com base nessas evidências buscou-se averiguar quais profissionais estão envolvidos diretamente com a construção de vocabulários controlados do domínio da Música. O Quadro 24 mostra a participação majoritária de bibliotecários auxiliados por músicos e pesquisadores que atuam como docentes nas referidas instituições.

**Quadro 24** - Profissionais envolvidos com a construção dos vocabulários controlados

<b>Gestores</b>	<b>Quais são os profissionais envolvidos na elaboração e manutenção do vocabulário do domínio da Música?</b>
G1	<i>Bibliotecários e Músicos. A bibliotecária chefe da Seção de Música também possui formação em Música.</i>
G2	<i>Bibliotecários.</i>
G3	<i>Bibliotecários. O Vocabulário Controlado [nome da instituição] integra o conjunto de processos sistêmicos da agência [nome da instituição] de gestão da informação acadêmica [sigla do grupo de gestão] e, dessa forma, possui um grupo de gerenciamento específico para sua manutenção, constituído por profissionais representantes das bibliotecas da [nome da instituição] por área do conhecimento (3 bibliotecários da área de Ciências Biológicas; 3 da área de Ciências Exatas e Tecnologias, e 3 da área de Ciências Humanas - os termos da área de música são avaliados nesta área), além de representantes da Coordenadoria da [sigla do grupo de gestão] e de representantes acadêmicos da área de Biblioteconomia e Ciência da Informação.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

O G1 afirma haver um profissional bibliotecário que também tem formação em Música atuando na elaboração e manutenção do vocabulário. O G3, por sua vez, enfatiza que além dos bibliotecários e especialistas do domínio da Música há representantes acadêmicos da área de Biblioteconomia e Ciência da Informação que colaboram. É importante destacar que as bibliotecas públicas e universitárias estão organizadas em redes e, por isso, a construção dos vocabulários controlados nessas instituições pode estar sob a responsabilidade de um grupo gestor do qual faz parte representantes de diversas áreas de conhecimento.

A participação efetiva de profissionais da área de Terminologia e de especialistas do domínio se faz necessária porque podem surgir eventuais dúvidas sobre a pertinência de um termo. Assim, esses profissionais são consultados geralmente para resolver questões de ambiguidade ou para aconselhar quanto à preferência no uso de termos mais usuais ou representativos em detrimento a outros que se mostrarem obsoletos (NOHAMA et al., 2012; CARACCILO et al., 2013; ÇAĞDAŞ; STUBKJÆR, 2015; BINDING; TUDHOPE, 2016).

Nesse ponto, voltamos ao aspecto da atualização dos vocabulários. Uma vez destinados ao usuário, estes devem ser constituídos de termos correntes e a linguagem de representação se aproximar ao máximo daquela utilizada pela comunidade. O Quadro 25 aponta as respostas dos gestores ao serem indagados quanto à participação da comunidade na elaboração de seus vocabulários, em especial se esta puder sugerir inclusões ou modificações de termos.

**Quadro 25** – Modificações ou sugestões de termos do vocabulário

<b>Gestores</b>	<b>A comunidade pode sugerir inclusões ou modificações de termos no vocabulário controlado?</b>
G1	<i>Sim. São sugeridos por membros da equipe de controle, usuários e pessoas ou grupos externos à instituição.</i>
G2	<i>Sim. São sugeridos por membros da equipe de controle, usuários, docentes e pesquisadores.</i>
G3	<i>Sim. A sugestão de novos termos é realizada pela equipe de indexadores da [sigla do grupo de gestão]. Mas, usuários em geral, ao perceberem a ausência de termos específicos na indexação de determinados materiais ao pesquisarem no [nome do catálogo online utilizado], poderão encaminhar suas sugestões para os bibliotecários de suas Unidades que, após avaliarem a relevância do termo (conceito, nota de escopo etc) irão fazer a sugestão do termo pela Base de Sugestões do Vocabulário [nome da instituição].</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Todos afirmaram ser possível a sugestão de modificações pela comunidade de usuários, especialmente quando notar a ausência de um termo específico durante as consultas nos catálogos. O G3 especifica que as opiniões são encaminhadas aos bibliotecários das unidades especializadas, os quais avaliam sua relevância antes de incluí-las na base de sugestões. Por conseguinte, uma avaliação final é realizada pela equipe de gestão.

Para que os usuários participem desse ato é fundamental que as instituições disponham de canais efetivos de comunicação. Desse modo, buscou-se identificar quais são os canais utilizados e como são realizadas as recomendações pela comunidade. Mediante o Quadro 26, é verossímil notar que a ferramenta de comunicação mais empregada é o correio eletrônico ou o contato através do “fale conosco”, disponível nos endereços eletrônicos das unidades de informação na *web*. Ressalta-se que o G3 foi o único a dizer que a instituição desenvolveu uma ferramenta específica para acolher essas sugestões.

**Quadro 26** - Canais de comunicação e procedimentos para acolher as sugestões da comunidade

<b>Gestores</b>	<b>No que se refere as recomendações de inclusão ou modificação de termos no vocabulário realizados pela comunidade (usuários do sistema): Quais são os canais de comunicação utilizados para que a comunidade recomende as inclusões ou modificações?</b>
G1	<i>E-mail ou contato através de “fale conosco”.</i>
G2	<i>E-mail do setor de tratamento da informação.</i>
G3	<i>O canal utilizado para sugestão de novos termos, exclusão ou alteração de termos é o Sibix-650B versão 1.7.6 de maio de 2016 - desenvolvido pelo Analista de sistemas [nome do analista] da equipe [sigla do grupo de gestão da informação acadêmica].</i>
<b>Como são realizadas as recomendações de inclusão ou modificação dos termos pela comunidade?</b>	
G1	<i>Depois de recebida a solicitação, a mensagem é enviada ao Setor de Autoridades para verificação. Como regra geral, não fazemos inclusões simplesmente a pedido, já que a tabela de autoridades deve estar correlacionada às obras do acervo.</i>
G2	<i>As sugestões são analisadas e nestes casos são encaminhadas para o grupo de catalogação. Os bibliotecários só tem permissão para inserir assuntos que já existam em vocabulários consagrados, como BN, LC, etc.</i>
G3	<i>1. O bibliotecário indexador faz a sugestão do termo; 2. Grupo gestor recebe um e-mail informando que existem termos para serem analisados; 3. O pedido é avaliado em três níveis: Nível 1 (Manutenção - Os responsáveis pela área do termo fazem a avaliação do pedido) Nível 2 (Conteúdo – a gestora de conteúdo revisa o pedido, checa nota de escopo, hierarquia etc) Nível 3 (Processo – o termo é incluído no Vocabulário); 4. Ao finalizar o processo, o bibliotecário indexador recebe um email informando se o termo foi incluído ou não no Vocabulário.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Apesar da possibilidade dos usuários sugerirem termos, todas as instituições asseguram que os pedidos são avaliados pelos responsáveis de cada área de conhecimento do vocabulário. No caso do domínio da Música, majoritariamente por bibliotecários, apoiados por músicos e pesquisadores. As inclusões não são realizadas a pedido e, em alguns casos, devem estar alinhadas aos vocabulários mantidos por outras instituições (conforme descrito pelo G2) ou ter correlação com a indexação das obras do acervo (conforme descrito pelo G1). É importante frisar que o termo deve ser incorporado conforme sua pertinência e a estrutura hierárquica do vocabulário controlado posta de maneira adequada. Assim, são examinadas notas de escopo e fontes auxiliares para identificar sua adequação à linguagem.

Como algumas unidades de informação condicionam a inclusão de termos às atividades de indexação das coleções, as redes de indexação ou catalogação cooperativa podem contribuir com a identificação de termos que não sejam contemplados pelos vocabulários locais. O Quadro 27 indica que todas as instituições afirmaram não fazer parte de redes de cooperação. Algumas concordam que podem contribuir com os vocabulários controlados no aspecto da uniformidade e da qualidade da linguagem construída. Com exceção do G3, que não dispôs de resposta.

**Quadro 27** - Participação das unidades de informação em redes cooperativas de catalogação ou indexação

<b>Gestores</b>	<b>A unidade de informação faz parte de alguma rede cooperativa de catalogação ou indexação de documentos especializados no domínio da Música ou áreas correlatas?</b> <b>a) Em caso afirmativo, quais são as instituições que cooperam?</b>
G1	<i>Não.</i>
G2	<i>Não.</i>
G3	<i>Não.</i>
<b>b) Em ambos os casos, na sua opinião, qual a importância do controle do vocabulário para as redes cooperativas de indexação ou catalogação?</b>	
G1	<i>A importância é crítica, uma vez que é através da uniformidade de vocabulário que aspectos como precisão e revocação do catálogo são atendidos.</i>
G2	<i>A cooperação é fundamental para a qualidade do vocabulário controlado.</i>
G3	<i>[nenhuma resposta obtida]</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

A colaboração com outras instituições é sublinhada na literatura como um meio de agilizar a elaboração do vocabulário, trazer benefícios para sua atualização e exercer importante papel no controle terminológico. Essas colaborações podem ser formais – quando entidades e

grupos de pessoas estabelecem acordos – ou informais, quando não há acordos estabelecidos, mas as entidades e o corpo social colaboram indiretamente com a construção.

Ao serem indagadas quanto à existência de algum projeto de cooperação formal para o controle do vocabulário geral ou específico do domínio da Música, apenas o G2 afirma existir, cujos detalhes são apresentados abaixo:

*Sim. Participam do nosso vocabulário controlado todos os bibliotecários da rede de bibliotecas [...]. Usamos formulários online para fazer pedidos de novos assuntos para o grupo. No caso de assuntos encontrados na BN, por exemplo, usamos o formato MARC para leitura e inserção em nossa base (G2).*

Segundo o gestor, a cooperação formal é garantida pela participação de todos os bibliotecários do sistema que atuam em conjunto com o grupo designado especificamente para gestão do vocabulário, os quais usam formulários para comunicar os pedidos e sugestões e padrões de dados para a leitura e inserção de registros em suas bases de autoridades. Como o MARC é um formato para a leitura e intercâmbio de registros bibliográficos bastante empregue pelas bibliotecas, o que de fato corre no caso apresentado pelo G2 é uma importação de registros de autoridades de vocabulários controlados mantidos por outras instituições para uma base local. Isso pode ocorrer independente do estabelecimento de acordos entre as duas instituições, pois a maioria dos catálogos *online* oferecem gratuitamente seus registros em formato MARC.

Quanto à existência de cooperação informal, apenas o G2 sinaliza a participação dos alunos e professores, que contribuem com as atividades de manutenção ao sugerir mudanças no vocabulário mediante correio eletrônico.

*Sim. Alunos e professores. Enviam as sugestões por e-mail que serão analisadas pelos bibliotecários (G2)*

Apesar da resposta negativa obtida pelo G1 e G3, estes afirmam em questões anteriores (Quadro 25 e Quadro 26) haverem grupos de profissionais que atuam internamente na gestão dos vocabulários, bem como a existência de canais de comunicação e processos que garantem a participação da comunidade nas fases de manutenção. Desse modo, não se sabe ao certo os motivos pelos quais esses gestores não consideraram tais atividades como características da cooperação formal ou informal. Não é clara suas percepções quanto às atividades de cooperação.



Nota-se, contudo, que em todas as instituições há algum tipo de cooperação, seja com profissionais bibliotecários, pesquisadores especialistas do domínio da Música, internos ou externos à instituição, seja com a comunidade de usuários. Nenhuma parceria, formal ou informal, com entidades externas às instituições foram destacadas por eles, apesar de declararem constantemente que utilizam vocabulários existentes como base para a construção de seus próprios vocabulários.

#### 4.1.4 Compilação ou levantamento de termos

A compilação de termos é uma etapa que ocorre geralmente no início da elaboração dos vocabulários controlados. Está naturalmente condicionada a avaliação de especialistas do domínio, em que a partir de um *corpus* de documentos os termos representativos são coletados e selecionados. Nessa etapa, a equipe de construção deve assegurar que os termos selecionados representem efetivamente o domínio, executando a garantia literária do vocabulário desenvolvido.

O Quadro 28 mostra que ao serem indagados a respeito de como são selecionados os termos candidatos para inclusão nos vocabulários, especificamente do domínio da Música, os gestores afirmaram que são consultadas terminologias existentes. Além dessas terminologias, esses termos são recolhidos mediante literatura da área ou sugeridos pelos bibliotecários indexadores.

**Quadro 28** - Seleção dos termos candidatos à inclusão no vocabulário controlado

Gestores	Como são selecionados os termos candidatos a inclusão no vocabulário do domínio da Música?
G1	<i>Majoritariamente através da utilização prévia na indexação da Library of Congress, catálogo em que a Biblioteca [nome da instituição] se baseia para construção de seus registros bibliográficos.</i>
G2	<i>Termos já disponíveis em vocabulários controlados consagrados ou em literatura da área.</i>
G3	<i>Os termos candidatos são sugeridos pelos bibliotecários indexadores.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Embora o G3 tenha destacado a participação de bibliotecários e indexadores na compilação, declarando que os termos candidatos são sugeridos por esses profissionais, percebeu-se na metodologia que a instituição também realiza consultas em fontes de referências

como, por exemplo, dicionários e tesouros especializados. Apenas o G2 disse consultar a literatura da área.

É prudente ressaltar a importância de executar a garantia literária para que o vocabulário controlado exerça fundamentalmente seu papel: representar um domínio do conhecimento assegurando que os termos selecionados sejam amplamente conhecidos e compartilhados pela comunidade científica (HARPRING, 2018; NORUZI, 2018).

O Quadro 29 mostra que todos os vocabulários controlados têm em conta vocabulários existentes. Evidenciaram-se no nível internacional os vocabulários da LC e nacionalmente a Terminologia de assuntos da BN e o Vocabulário Controlado da USP. Logo, basear-se em vocabulários existentes é considerado fator positivo porque otimiza o tempo dessas instituições, reduz custos, evita a duplicidade de vocabulários cobrindo um mesmo domínio e ajuda a manter o controle terminológico à medida que as unidades de informação selecionam termos já consolidados na área de domínio (JUPP et al., 2008; L'ABATE et al., 2015; ÇAĞDAŞ; STUBKJÆR, 2015).

**Quadro 29** - Uso de vocabulários existentes para a seleção de termos candidatos

<b>Gestores</b>	<b>Os termos para inclusão no vocabulário controlado são selecionados a partir de terminologias ou vocabulários existentes?</b> <b>a) Em caso afirmativo, quais são as terminologias ou vocabulários existentes utilizados?</b>
G1	<i>Sim. Library of Congress Subject Headings.</i>
G2	<i>Sim. Vocabulário Controlado da USP, Library of Congress Subject Headings e Terminologia de Assuntos da Biblioteca Nacional.</i>
G3	<i>Sim. Library of Congress Subject Headings, Musical Instrument Museums Online, Tesouro EuroVoc, UNESCO Thesaurus, Terminologia de Assuntos da Biblioteca Nacional.</i> <i>Outros: No caso de termos Gênero/Forma – LCGFT (Library of Congress Genre/Form Terms), e dos Meios de Expressão – LCMPT (Library of Congress Medium of Performance Thesaurus).</i> <i>Observação: a orientação na [nome da instituição] é para que os indexadores façam a inclusão de termo novo desde que constem nos principais tesouros das áreas (DeCs, CAB etc) e que tenham registros no [nome do sistema de gestão do acervo] que tratem do assunto.</i>
<b>b) Em caso afirmativo, existe uma parceria formal com as instituições que mantêm esses vocabulários?</b>	
G1	<i>Não existe uma parceria formal, os termos são apenas recolhidos de seus endereços na web</i>
G2	<i>Não existe uma parceria formal, os termos são apenas recolhidos de seus endereços na web.</i>
G3	<i>Não, os termos são apenas recolhidos de seus endereços na web.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Além do LCSH, o G3 destaca o uso de dois tesouros específicos, também mantidos pela LC: o *Library of Congress Genre/Form Terms* (LCGFT) e o *Library of Congress Medium of Performance Thesaurus* (LCMPT). Este último está voltado para a representação de entidades necessárias à execução de obras musicais, como vozes e instrumentos. O LCMPT é aplicado para especificar os gêneros ou formas da Música, a exemplo Concertos (piano e orquestra) e Sonatas (clarinete). Concertos e Sonatas são tipos de composições musicais: clarinete, piano e orquestra são os instrumentos utilizados para executar as referidas composições. Particularmente na Música, uma mesma obra pode se apresentar em diferentes meios de expressão, como partituras, gravações de vídeo ou registros sonoros.

Embora tenha-se destacado o uso de termos de vocabulários existentes, no Quadro 30 os gestores esclarecem que os termos recolhidos podem ser adaptados, principalmente em virtude da tradução ou da linguagem documentária empregue na indexação do acervo. O G3 concorda que é possível executar modificações, a considerar as tabelas auxiliares do sistema de classificação em uso na unidade de informação.

**Quadro 30** - Adaptação dos termos compilados a partir de vocabulários controlados existentes

<b>Gestores</b>	<b>Os termos são incluídos no vocabulário exatamente como se apresentam nas terminologias ou vocabulários existentes, ou podem ser adaptados à realidade da instituição?</b>
G1	<i>Passíveis de adaptação. São adaptados principalmente em virtude de tradução.</i>
G2	<i>Usados exatamente como são apresentados nos vocabulários existentes ou Passíveis de adaptação. São adaptados de acordo com os campos MARC usados pela rede de bibliotecas [...], mas o assunto em si não pode ser alterado a não ser os termos em inglês que são traduzidos.</i>
G3	<i>Passíveis de adaptação. Os termos são convertidos para a linguagem documentária utilizada no Vocabulário [nome da instituição], ou seja, linguagem de indexação pós-coordenada onde o conteúdo dos documentos é representado pelo uso direto (Campos 650, 651, 655 e 656) ou combinação com as Tabelas Auxiliares (Geográfica, Gênero e forma, Qualificadores, Profissões).</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Em geral, a adaptação por encargo da tradução é necessária em virtude de aspectos multiculturais da organização do conhecimento (BACA; GILL, 2014). Desse modo, devem ser evitadas simples traduções de palavras, visto que elas podem não agregar valor à indexação e prejudicar a recuperação da informação. Assim, a atividade de tradução exige uma equipe especializada no idioma, pois alguns termos nos idiomas de origem podem não dispor de termos

correspondentes no idioma de destino, havendo a necessidade de adaptação em razão dos aspectos culturais do público local (MOREIRA; MOURA, 2006; ABECKER et al., 2015).

Quando indagados sobre quem são os profissionais, unidades especializadas ou departamentos responsáveis pela compilação dos termos do domínio da Música, todos afirmaram que os bibliotecários executam essa função – conforme descrito no Quadro 31 – com destaque para o G2 e G3, que enfatizaram a participação de grupos de catalogação ou grupos de gestão do vocabulário.

**Quadro 31** - Responsáveis pela compilação dos termos do domínio da Música

<b>Gestores</b>	<b>Quais são os profissionais, unidades especializadas ou departamentos responsáveis pela compilação dos termos do domínio da Música?</b>
G1	<i>Bibliotecários do Setor de Autoridades e da Seção de Música e Arquivo Sonoro.</i>
G2	<i>Para inserção de termos são os mesmos de todos os outros termos: bibliotecários. Para a compilação que foi feita somente no início os profissionais além de bibliotecários teriam que ser membros do grupo de catalogação.</i>
G3	<i>Na [nome da instituição] os termos da área de Música são sugeridos principalmente pelos bibliotecários indexadores da Biblioteca da [nome do UI especializada em Música da universidade] (A [nome da instituição (campus avançado)] também possui curso de Música, mas raramente sugerem assuntos). A [nome da biblioteca] possui o maior acervo de partituras cadastrado no [nome do sistema de gestão], pois forma e organiza documentos musicais desde a década de 1970.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

As respostas apresentadas por G1 e G2 mostram que os profissionais estão situados em unidades de informação especializadas, que mantêm os acervos musicais e atendem às necessidades informacionais dos alunos do curso de Música nessas instituições. Ainda, são envolvidos com a atividade de compilação de termos candidatos, necessitando, por essa razão, ter experiência com a organização de coleções da área ou o mínimo de interação com a comunidade de usuários.

Questionados com relação às competências básicas dos membros da equipe de construção na fase de compilação de termos, especificamente no que concerne ao domínio da Música, o Quadro 32 apresenta os gestores que anunciaram ser essencial que os profissionais precisem ter formação em biblioteconomia, participem dos grupos de catalogação ou gestão do vocabulário e conheçam não apenas as políticas de indexação de acervos musicais, mas também

as metodologias formuladas pelos grupos de gestão. Não percebeu-se a participação de especialistas de domínio na fase de compilação em nenhuma dessas instituições.

**Quadro 32** - Competências básicas atribuídas aos profissionais responsáveis pela compilação de termos do domínio da Música

<b>Gestores</b>	<b>Quais as competências básicas atribuídas aos profissionais responsáveis pela compilação dos termos do domínio da Música?</b>
G1	<i>Formação em biblioteconomia e experiência no uso do formato MARC21 Authorities.</i>
G2	<i>Ser bibliotecário e participar do grupo de catalogação.</i>
G3	<i>Conhecer e aplicar as políticas de indexação de acervos musicais e do Vocabulário [nome da instituição].</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Parece ser restrito aos grupos de gestão a seleção dos termos candidatos, provavelmente em função da necessidade de manter certo grau de controle nessa etapa. É natural que lexicógrafos ou terminólogos executem essa tarefa, pois na fase inicial de compilação pode não ser preciso avaliar o nível de especificidade dos termos. A avaliação e especificação de termos geralmente ocorre numa etapa posterior, na qual pesquisadores são consultados e levados a analisar não só a pertinência, bem como as relações lógico-semânticas.

Embora a compilação de termos seja uma atividade característica da fase inicial, é possível executá-la nas demais etapas de construção. Isso pode ser vital em razão das sugestões de novos termos por membros da equipe ou por usuários do sistema. Essas sugestões podem ser o gatilho para a realização de um levantamento na literatura a fim de investigar os desdobramentos de uma nova classe de assuntos até então desconhecida ou não contemplada pelo vocabulário.

Nesse sentido, questionou-se aos gestores se a compilação de termos pode ocorrer em outras etapas. O Quadro 33 mostra que o G2 ratifica que a compilação de termos sucedeu apenas na fase inicial. Porém, os G1 e G2 consideraram a possibilidade de efetuar levantamentos de novos termos durante outras fases. A incorporação de novas classes de assuntos no decorrer da construção tende a preencher lacunas percebidas só depois da fase aludida.

**Quadro 33** - Etapas em que podem ocorrer a compilação de termos

<b>Gestores</b>	<b>O processo de compilação dos termos ocorre apenas na fase inicial de construção do vocabulário, ou em outras etapas do processo de construção? Especifique quais etapas podem ocorrer a compilação de termos.</b>
G1	<i>A construção do vocabulário é um processo contínuo.</i>
G2	<i>A compilação foi feita para todos os assuntos pelo grupo de catalogação somente na fase inicial do vocabulário, hoje fazemos a manutenção do mesmo.</i>
G3	<i>A compilação de termos se deu durante a fase inicial de construção do Vocabulário Controlado [nome da instituição]. Sim, às vezes a unidade tem interesse em revisar sua área, ou seja, analisar e revisar áreas do conhecimento pertencentes ao Vocabulário [nome da instituição] para agilizar sua atualização. Antes de iniciar a revisão, é necessário comunicar ao Grupo Gestor do Vocabulário para organização de uma agenda que viabilize a conferência da revisão e a implementação do novo vocabulário - tarefas que são responsabilidade do grupo gestor. Também é necessário consultar outras unidades da mesma área do conhecimento e convidá-las a participarem da revisão.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Ao considerar que é um processo contínuo, mostra que os vocabulários estão sempre sendo revisados conforme as áreas de conhecimento. Essa ação garante a longevidade da terminologia, pois uma vez mantidas as atualizações periódicas os vocabulários controlados asseguram que os termos empregues são capazes de promover uma indexação eficiente e, igualmente, auxiliar os usuários na recuperação da informação (MOREIRA; MOURA, 2006).

As normas de elaboração, bem como estudos da área, consideram o uso de tecnologias ou ferramentas automáticas e semiautomáticas para compilar termos candidatos. Desse modo, os gestores foram questionados se na construção é aplicado algum tipo de técnica de PLN, como mineração de texto e dados, pré-processamento de documentos, indexação automática ou ferramentas bibliométricas. Contudo, todos os gestores afirmaram não lidar com tecnologias ou ferramentas para a compilação de termos, seja de maneira automática, seja de maneira semiautomática.

Apesar de confirmada a ausência de técnicas de PLN, sua importância é ratificada na literatura (MOREIRA; MOURA, 2006; LIEBESKIND; DAGAN; SCHLER, 2016, 2018). Isto posto, aplicar técnicas automáticas a esse procedimento pode agilizar a construção à medida que as ferramentas são capazes de identificar ou agrupar termos representativos em classes de assunto, facilitando a posteriori a avaliação por especialistas humanos, inclusive quanto à seleção de termos em outros idiomas.

Algumas linguagens de indexação são construídas para atender, em certa medida, um nível de representação multilíngue, possibilitando a superação de barreiras impostas pelo idioma na recuperação da informação. O Quadro 34 indica que os vocabulários controlados, em sua maioria, contemplam termos em outros idiomas.

**Quadro 34** - Compilação e tradução de termos em outros idiomas

<b>Gestores</b>	<b>O vocabulário controlado contempla termos em outros idiomas? Em caso afirmativo, como ocorre o processo de compilação e/ou tradução?</b>
G1	<i>Sim. A compilação ocorre no momento da indexação, quando o bibliotecário identifica o uso de um termo ainda não existente em nossa tabela. Os termos originais extraídos da LCSH são incluídos como entrada secundária no registro traduzido.</i>
G2	<i>Sim. A tradução é feita pelo bibliotecário catalogador baseado em dicionários e outras literaturas da área. Usamos também o campo 750 do MARC para termo em inglês.</i>
G3	<i>Não.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Os G1 e G2 certificaram que os termos são extraídos de vocabulários existentes e os registros inseridos no sistema como entradas secundárias. Apenas o G3 afirmou que na instituição nenhum termo em outro idioma é admitido. Ao se referir ao campo 750 do MARC, o G2 demonstrou que a admissão de termos em outros idiomas está condicionada à indexação das coleções nessas instituições.

Constatada a ausência de ferramentas automáticas e com base na afirmação do G2 quanto ao uso de dicionários, é conveniente asseverar que as traduções são feitas manualmente. Nesse aspecto, não foi informado por qualquer um dos gestores como são avaliados os termos em outros idiomas, o que pode comprometer a qualidade das representações bilíngues.

#### **4.1.5 Especificação ou supressão de termos**

Os termos compilados geralmente são substantivos ou grupos de substantivos selecionados em função da ocorrência frequente na literatura da área. Uma fase posterior de especificação pode ser necessária para que termos genéricos possam ser melhor explicitados, a depender se o vocabulário está destinado a uma indexação exaustiva ou não. Nessa fase é possível também eliminar termos semanticamente similares, resolvendo o problema da ambiguidade de conceitos ou redundâncias.

Nesse sentido, os gestores foram questionados se os termos do vocabulário podem ser especificados e em quais casos isso seria possível de ocorrer. O Quadro 35 aponta que apenas o G1 declara não executar nenhum tipo de especificação, enquanto que o G2 e G3 testificaram que podem ocorrer em alguns casos, principalmente quando o termo é coletado em vocabulários existentes ou um mesmo termo assume significados diferentes.

**Quadro 35** - Especificação de termos

<b>Gestores</b>	<b>Os termos compilados são passíveis de especificação? Em caso afirmativo, quais são os casos que exigem o processo de especificação?</b>
G1	<i>Não.</i>
G2	<i>Sim. Quando um assunto já incluído é encontrado em outra base com mais informações, como sinônimos ou termo traduzido por exemplo, estas informações podem ser incluídas no registro.</i>
G3	<i>Sim. É necessária a especificação dos termos quando possuem significados diferentes.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Ao afirmar que um assunto “[...] é encontrado em outras bases com mais informações [...]”, bem como “[...] estas informações podem ser incluídas no registro”, o G2 parece sinalizar que o grupo gestor necessita analisar as notas de escopo ou a estrutura hierárquica dos termos nos vocabulários de origem. As relações associativas e de equivalência do termo podem auxiliar os profissionais a compreender como organizá-lo dentro de uma classe de conceitos. Não obstante, as decisões sobre sua especificação dependerão também do nível de especialidade ou da familiaridade que os gestores dos vocabulários têm com o domínio.

Diante disso, ao observar a inconformidade de alguns termos em relação ao domínio, os gestores dos vocabulários devem excluí-los. É comum ocorrer algum tipo de exclusão quando palavras não representativas são frequentes no *corpus* de coleta, em especial expressões popularmente conhecidas que não são compartilhadas pela comunidade científica. De outra forma, os gestores podem optar por utilizar um termo preferido em detrimento a outro, ou, nos casos dos vocabulários multilíngues, descartar expressões do termo em outro idioma que não podem ser traduzidas ou aplicadas como aparecem nos vocabulários de origem.

Apesar de listadas as inúmeras possibilidades, ao serem questionados a respeito da necessidade de supressão ou exclusão de alguns termos compilados, apenas o G3 concorda ser possível, em particular para os casos de ambiguidade e obsolescência de termos:



*Sim [há necessidade de supressão ou exclusão de termos]. Para a eliminação das ambiguidades durante o processo de construção do Vocabulário [nome da instituição]. E, de termos que ficaram obsoletos e não estejam mais sendo utilizados na indexação (G3).*

Nessa etapa é adequado submeter os termos a uma minuciosa análise, que pode ser efetuada mediante conferência das definições em dicionários especializados ou demais terminologias que cobrem o domínio. Esse ato confere rigor à atividade de compatibilização de termos, permitindo minimizar ambiguidades à medida que as relações lógico-semânticas são estabelecidas. Apesar disso, ao persistirem as dúvidas os especialistas do domínio devem ser consultados (CHEN; YEH; SIE, 2005; GRANADA; VIEIRA; LIMA, 2012; ÇAĞDAŞ; STUBKJÆR, 2015).

Quanto à participação de especialistas do domínio da Música, apenas o G3 afirmou consultá-los quando necessário. Eles podem ser convidados a avaliar a pertinência dos termos “[...] em caso de dúvida [...]” sobre uma definição compilada e seus relacionamentos. Destaca-se, também, que a comunicação entre o grupo gestor e os especialistas citados é realizada mediante mensagens eletrônicas (*e-mail*).

*Sim. Em caso de dúvida, especialistas com conhecimento na área de música são consultados para validação da definição compilada do termo e de seus relacionamentos. Os especialistas são consultados por email (G3).*

Embora o G1 tenha anunciado que especialistas do domínio não são consultados, é importante lembrar que na sua equipe há um bibliotecário com formação em Música (Quadro 24). Assim, a negativa desse gestor pode estar relacionada a ausência de especialistas que não têm vínculo direto ou formal com a construção do vocabulário controlado.

A partir das respostas obtidas nesta seção verificou-se que a especificação ou a supressão dos termos é comum apenas nas instituições que constroem seus vocabulários com base nas sugestões do grupo gestor ou de especialistas do domínio. As instituições que utilizam vocabulários mantidos por outras organizações preferem manter a conformidade com as terminologias já consolidadas na área.

#### **4.1.6 Registro e validação de termos**

Após a equipe de construção executar várias ações relativas à compilação e à supressão de termos, um grupo mais restrito, geralmente formado por editores ou gestores dos projetos,

são responsáveis pelo registro e validação. Nessa fase, todas as atividades ocorrem colaborativamente. Os editores registram e estruturam os termos dentro das classes de conceitos e os gestores procedem com a validação daqueles que julgam serem pertinentes.

De todo o modo, a construção dos vocabulários controlados necessita do suporte de ferramentas de edição distribuídas e colaborativas, devendo atuar em um fluxo de trabalho editorial formalizado e permitindo que as atividades sejam monitoradas e avaliadas continuamente (CARACCILO et al., 2011; LAPORTE; MOUGENOT; GARNIER, 2012).

Ao serem indagados como a etapa de registro é realizada, os gestores frisaram utilizar dois tipos de ferramentas: as bases de autoridades e os *softwares* específicos para a construção de vocabulários controlados. Os registros dos termos, quando executados nas bases de autoridades, incorporam o formato bibliográfico MARC 21 e permitem que os registros sejam legíveis e intercambiáveis por computador, visto que cada assunto ou subdivisão de assunto é descrito em uma etiqueta de marcação. Diante disso, a estrutura de termos nessas bases constitui os índices de assuntos usados para indexar as coleções.

O *software* de gestão de vocabulários controlados, mencionado pelo G3, foi desenvolvido na linguagem PHP com banco de dados MySQL. No entanto, seu endereço eletrônico na *web* consta a informação de que o vocabulário não está disponível para importação de sistemas externos à instituição. O sistema possui dois módulos principais: o Módulo de usuários, que controla os níveis de permissão dos usuários no sistema; e o Módulo de sugestões, que controla o fluxo de validação de sugestões até a inclusão no vocabulário.

É por meio da validação que são firmados os acordos sobre a pertinência dos termos compilados. Nesse quesito, buscou-se investigar se os demais membros da equipe de controle dos vocabulários participam da etapa de validação. Nesse caso, apenas o G3 afirmou haver uma etapa de validação, sendo esta efetivada por uma equipe de gerenciamento do vocabulário.

*Sim. A validação é feita pela Equipe de Gerenciamento do Vocabulário Controlado. Mas, devido à incorporação do [sistema de bibliotecas] pela agência de gestão da informação acadêmica no primeiro semestre de 2019, a equipe para o biênio 2020-2022 não foi designada. Os membros do último biênio podem ser conferidos no link: [fornece endereço eletrônico] (G3).*

Ao relatar “[...] a equipe para o biênio 2020-2022 não foi designada”, indica que a equipe de profissionais sugerida para a atividade de validação pode ser alterada. Contudo, declara que as propostas de inclusão ou alteração são encaminhadas a um grupo restrito, distribuídos por áreas do conhecimento.

Reiterada a importância de agilizar a elaboração dos vocabulários controlados na literatura para além da forma manual, a atividade de validação pode ser executada de maneira automática ou semiautomática. Nesse caso, os gestores foram indagados com relação às tecnologias empregues nessa etapa. As respostas apresentadas no Quadro 36 mostram que apenas os G1 e G2 confirmaram utilizar instrumentos ou tecnologias na etapa de validação.

**Quadro 36** - Processamento automático, semiautomático ou manual aplicado à validação dos termos

Gestores	Existe alguma espécie de processamento automático, semiautomático ou manual aplicado à validação dos termos? Em caso afirmativo, especifique quais são os instrumentos ou tecnologias utilizadas?
G1	<i>Sim. A inclusão de termos por parte de servidores externos ao Setor de Autoridades é informada e a chefia do setor valida o registro.</i>
G2	<i>Não.</i>
G3	<i>Sim. Utilizamos um aplicativo que proporciona o registro de sugestões pelas Bibliotecas; análise e validação pelo Grupo de Gerenciamento; atualização do Vocabulário pela equipe [gestão da informação acadêmica]. Por meio de formulário eletrônico, as bibliotecas podem sugerir novos termos nas diferentes áreas do conhecimento, assim como solicitar alterações ou remoções de termos já existentes. Além desses procedimentos, há a possibilidade, ainda, de acompanhamento dos pedidos feitos até a etapa de finalização do processo. Para facilitar a administração dos pedidos, o Grupo de Gerenciamento do vocabulário [nome da instituição] estabeleceu que cada biblioteca tenha um bibliotecário responsável pelo encaminhamento das solicitações. O funcionário indicado pela Chefia Técnica da Biblioteca recebe uma senha para entrada nesse sistema, de modo a centralizar os pedidos de cada biblioteca.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Ao dizer “[...] A inclusão de termos por parte de servidores externos ao Setor de Autoridades é informada [...]” à chefia do setor, o G1 revela que o procedimento é realizado quando profissionais que não atuam na mesma divisão executam modificações na base. Após serem informados acerca das mudanças via correio eletrônico, essa chefia valida o registro feito por esses profissionais de forma manual.

A validação é descrita pelo G3 com detalhes, que menciona ocorrer mediante formulário eletrônico. Os dados das solicitações são encaminhados ao profissional responsável pelo gerenciamento do vocabulário do domínio específico designado pela chefia técnica da biblioteca especializada, que analisa e valida a inclusão de novos termos. Quando afirma que “[...] há a possibilidade, ainda, de acompanhamento dos pedidos feitos até a etapa de finalização do processo [...]”, sinaliza que a validação está incorporada a um fluxo de trabalho bem

definido, cujas ferramentas eletrônicas são operadas apenas para encaminhar e acompanhar as solicitações junto ao profissional ou setor responsável pela validação. Apesar do uso dessas ferramentas, percebe-se que a validação nessas instituições é feita de forma semiautomática.

Embora não seja comum a atuação da comunidade na etapa de validação, buscou-se conhecer se há alguma iniciativa para integrar os usuários do sistema a essa etapa. Constatou-se que esses sujeitos não contribuem com a fase em destaque.

As decisões e acordos quanto à inclusão de termos é uma etapa que pode envolver um grande número de profissionais, principalmente quando os procedimentos são executados continuamente. Todavia, ao serem conduzidos por diferentes especialistas pode ocorrer de nem todos estarem em harmonia com as deliberações tomadas. Isso pode ser aprimorado se as atividades de edição estiverem orientadas por diretrizes bem definidas tentando promover o consenso.

De outra forma, as ferramentas utilizadas na edição devem dispor de espaços colaborativos para promover discussões entre os membros da equipe. Isso pode ajudar a manter o controle do vocabulário em função da uniformidade de decisões e redução de erros. Todavia, o Quadro 37 mostra que entre as instituições pesquisadas não é comum haver o diálogo sobre a inclusão ou a validação de termos nos vocabulários.

**Quadro 37** - Consenso sobre a inclusão de termos no vocabulário

<b>Gestores</b>	<b>Como a equipe chega ao consenso sobre a inclusão de um termo no vocabulário?</b>
G1	<i>O vocabulário utilizado é copiado dos registros da LCSH.</i>
G2	<i>Não existe essa consulta formal, pois os bibliotecários só estão autorizados a inserir termos já disponíveis em outras bases.</i>
G3	<i>Conforme mencionado anteriormente, para a inclusão de um termo no vocabulário, após a solicitação via Base Sugestões, o pedido é avaliado em três níveis: Nível 1 (Manutenção - Os responsáveis pela área do termo fazem a avaliação do pedido – verificam a Nota de Escopo, justificativa para a inclusão do termo, e fontes consultadas); Nível 2 (Conteúdo – a gestora de conteúdo revisa o pedido, checa nota de escopo, hierarquia etc) Nível 3 (Processo – o termo é incluído no Vocabulário).</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

O G1 diz que o vocabulário “[...] é copiado dos registros da LCSH”, e o G2 confirma que “[...] os bibliotecários só estão autorizados a inserir termos já disponíveis em outras bases”, ou seja, nessas instituições não há necessidade de produzir acordos ou consensos, já que os

termos são retirados de vocabulários existentes, mesmo para os casos em que alguns precisam ser adaptados. Isso demonstra que em alguns fatos a validação pode ser arbitrária e centralizada.

Apenas o G3 declara haver um procedimento de avaliação dos pedidos de solicitação e checagem final em que se busca o consenso sobre a inclusão. As sugestões são encaminhadas ao grupo gestor e só após passar pelos procedimentos citados são incluídos no vocabulário. Embora não destacado a relevância de se discutir a respeito da inclusão ou não de termos, o gestor em realce certifica que “[...] o pedido é avaliado em três níveis [...]” e sinaliza que a inclusão só ocorre depois que o termo passa por uma rigorosa avaliação, como em um fluxo de trabalho, o qual diferentes profissionais fazem parte.

Em relação à resposta fornecida pelo gestor acima, entende-se que são nas atividades de solicitação e checagem que os acordos são estabelecidos. As sugestões passam por uma minuciosa conferência que leva em conta as justificativas para a inclusão, as fontes consultadas, notas de escopo e a estrutura hierárquica dos termos. Assim, a decisão quanto a incluir ou não o termo no vocabulário é fundamentado em todas as evidências levantadas pelo membro que o sugeriu e pela mostra do validador de que sua inclusão de fato seria relevante para o vocabulário.

#### **4.1.7 Testes e avaliação**

A etapa de testes e avaliação é sugerida nas normas e difundida na literatura (SPITERI, 2005; LAWRENCE et al., 2006). Diversas abordagens de avaliação são possíveis, cujos objetivos variam entre verificar se os relacionamentos semânticos estão bem veiculados nos níveis associativos ou hierárquicos e perceber se os termos selecionados refletem o uso comum da linguagem pela comunidade de usuários. Ademais, a manutenção permanente é requerida, pois pode haver mudanças no escopo do vocabulário controlada com a evolução natural da terminologia (INTERNATIONAL STANDARD ORGANIZATION, 2011, 2013).

Ao submeter o vocabulário a testes com participação do usuário, estes podem mostrar que um conceito sugerido já estava presente no vocabulário, porém designado pelo uso de um termo preferencial diferente. De outra forma, os membros da equipe de controle podem descobrir que alguns ajustes complementares são necessários e haver mudanças adicionais nas notas de escopo; e nos relacionamentos hierárquicos ou associativos podem melhorar os níveis de especificidade e compreensão quanto ao uso.

Embora seja uma recomendação, não foi relatada a realização de testes com usuários dos vocabulários controlados tratados nesta pesquisa. Apesar disso, os gestores julgam

relevante a participação desses atores sociais na etapa de testes e avaliação. É possível verificar no Quadro 38 que apenas o G1 se mostrou desfavorável.

**Quadro 38** – Importância da participação dos usuários na etapa de testes e avaliação dos vocabulários controlados do domínio da Música

Gestores	Conforme sua experiência, qual a importância dos usuários na etapa de testes e avaliação do vocabulário controlado?
G1	<i>De modo geral, a comunidade de usuários não compreende o escopo de atuação da Biblioteca [...] e faz críticas não pertinentes, como de falta de especificidade na indexação – algo que não compete necessariamente a uma biblioteca de última instância, não especializada.</i>
G2	<i>Seria bem interessante ter o apoio de alunos e professores.</i>
G3	<i>Acredito ser muito relevante a avaliação dos termos do vocabulário, assim, os usuários devem estar atentos na frequência de sucesso e fracasso na recuperação da informação por tema/assunto do Banco de Dados Bibliográfico [nome do sistema de gestão do acervo].</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

O G1 argumenta que alguns usuários fazem “[...] críticas não pertinentes, como de falta de especificação na indexação [...]”, ou seja, a instituição se reserva ao direito de realizar uma indexação pouco exaustiva devido o papel ou a missão que exerce junto à comunidade. Prossegue afirmando que é “[...] uma biblioteca de última instância [...]”, deixando a entender que supostamente não teria a missão de atender ao público especializado, apesar de ter um setor específico dedicado ao tratamento da informação musical.

Os demais gestores concordam que a participação dos usuários pode ser relevante. O G3 certifica que “[...] os usuários devem estar atentos na frequência de sucesso e fracasso na recuperação da informação [...]”, isto é, que a efetividade da recuperação da informação está condicionada à qualidade na manutenção da linguagem.

As abordagens mais aplicadas na fase de testes mencionadas na literatura têm em conta as análises das consultas dos usuários, podendo estas serem empregues para avaliar os termos usados pelos usuários tencionando verificar sua validade em função da recuperação da informação (SPITERI, 2005; MOURA, 2009; LAIPELT, 2015). Se um usuário frequentemente não obtém êxito ao usar um termo de consulta, esse termo pode ser considerado inválido, ser encaminhado para análise e possivelmente a equipe deve proceder com a exclusão.

#### 4.1.8 Ferramentas colaborativas e representação dos vocabulários controlados na *web*

As ferramentas colaborativas estão presentes em muitas etapas de construção e favorecem aos profissionais executar procedimentos de forma compartilhada, síncrona ou assincronamente, estando geograficamente dispersos ou no mesmo ambiente. À vista disso, os gestores foram indagados em relação à possibilidade de efetuar o controle do vocabulário de forma distribuída. As respostas são descritas no Quadro 39.

**Quadro 39** - Controle distribuído e compartilhado do vocabulário controlado

Gestores	<b>O controle do vocabulário é realizado de maneira distribuída e compartilhada? Ou seja, os profissionais envolvidos operam geograficamente dispersos, criando ou editando a estrutura de termos por meio de ferramentas colaborativas? Em quais etapas de elaboração do vocabulário controlado a edição colaborativa é necessária? Por quê?</b>
G1	<i>Sim. O trabalho é geograficamente distribuído porque a [nome da instituição] está distribuída em mais de um local, o que não necessariamente implica em trabalho “colaborativo”.</i>
G2	<i>Sim. No momento todo o processo é colaborativo.</i>
G3	<i>Não.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

O G2 relata que essa ação é possível em todas as etapas de construção do vocabulário. Apesar do G1 afirmar que “[...] o trabalho é geograficamente distribuído [...]” e a instituição “[...] está distribuída em mais de um local [...]”, assegura que essa não seria uma característica colaborativa, o que parece ser controverso, já que diversos profissionais alocados em diferentes unidades contribuem na elaboração do vocabulário.

É curioso evidenciar que o G3 declara não haver um controle compartilhado e distribuído, mesmo expondo nas questões anteriores (Quadros 24 e 25) que o vocabulário controlado é resultado de uma ação colaborativa da equipe de bibliotecários que atuam em diversas unidades de informação especializadas da instituição. Parece não haver um entendimento sobre o significado de controle compartilhado, possivelmente porque as decisões finais relativas à inclusão de termos são feitas por apenas alguns membros dos grupos gestores.

Quanto ao uso de ferramentas ou recursos de colaboração, os gestores destacaram, conforme apresentado no Quadro 40, a utilização do *e-mail*, formulários eletrônicos, processos em *workflow* (fluxo de trabalho) e reuniões presenciais.

**Quadro 40** – Recursos de colaboração e etapas em que são utilizados

<b>Gestores</b>	<b>Recursos e ferramentas de colaboração</b>	<b>Etapas em que os recursos são utilizados</b>
G1	<i>e-mail e reuniões presenciais.</i>	<i>Registro e validação de termos.</i>
G2	<i>Formulário eletrônico, e-mail, reuniões presenciais.</i>	<i>[sem resposta]</i>
G3	<i>e-mail, reuniões presenciais e processos em workflow (fluxo de trabalho).</i>	<i>Registro e validação de termos.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

No geral, o correio eletrônico e as reuniões presenciais são meios empregues para promover discussões entre os colaboradores. Ao mencionar os processos em *workflow*, o G3 se refere à ferramenta desenvolvida pela própria instituição, a qual tem sido um recurso usado para monitorar e controlar as ações da equipe de construção mediante um fluxo de tarefas na avaliação de pedidos de inclusão de termos em três níveis, descritos pelo gestor no Quadro 26.

Em relação às etapas que os recursos ou as ferramentas são aplicados, os gestores destacam as fases de registro e validação de termos. É possível verificar que o caráter colaborativo dessas etapas é mais evidente, visto ser por meio delas que os acordos são realizados e em que há o maior número de profissionais atuando em conjunto.

As ferramentas operadas durante a edição podem assumir diferentes características e, depender do nível de colaboração, ajudam a viabilizar a edição compartilhada. Por meio delas os membros da equipe acessam todos os recursos e editam os conjuntos de dados simultaneamente. Nesse sentido, é possível observar no Quadro 41 que os sistemas manipulados no decurso da edição do vocabulário controlado podem ser obtidos de forma remota, oferecidos na forma de *software as a servisse* e trabalhados com computação em nuvem ou a opção de edição simultânea. Apenas o G1 declarou que o sistema de gestão utilizado faz uso de tecnologias associadas ao conjunto de práticas do *Linked Data*.

**Quadro 41** – Características dos sistemas de gestão e uso de tecnologia do *Linked Data*

<b>Gestores</b>	<b>Características dos sistemas utilizados para a edição dos vocabulários</b>	<b>Tecnologia, padrão ou modelo de dados relacionados ao conjunto de práticas do <i>Linked Data</i></b>
G1	<i>Trabalha com computação em nuvem; Editável simultaneamente.</i>	<i>Sim. eXtensible Markup Language (XML).</i>
G2	<i>Editável simultaneamente; Oferecido na forma de Software as a Service (SaaS).</i>	<i>Desconheço.</i>
G3	<i>Acessível de forma remota.</i>	<i>Não.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).



Ao afirmar que os sistemas trabalham com computação em nuvem, o G1 e G2 concordam que o conjunto de dados do vocabulário controlado não está numa máquina específica e todos na instituição podem acessar. Porém, quando apresentada a característica de editável simultaneamente as plataformas podem restringir os papéis de atuação a profissionais específicos da equipe. Isso implica dizer que apenas aqueles que tiverem autorização poderão editá-lo. Somente o G2 afirmou usar um *software* comercial, acessado via Internet.

O oferecimento de sistemas na forma de *Software as a Service* (SaaS) é comum no âmbito das bibliotecas. Nele o cliente paga pelo serviço e o fornecedor se responsabiliza por toda a estrutura necessária para o uso (servidores, conectividade, cuidados com segurança da informação). Como relatado nas questões anteriores (Quadro 26), o G3 atesta empregar um sistema desenvolvido por analistas da própria instituição. A única característica mencionada é de que o sistema é acessível de forma remota, ou seja, é possível visualizar as informações, mediante login e senha, de qualquer lugar que se esteja.

O formato *eXtensible Markup Language* (XML), destacado pelo G1, é amplamente difundido na comunidade *Web*. Os G2 e G3 declararam não utilizar ou não conhecê-lo. Todavia, ao analisar os sistemas de gestão operados por suas respectivas unidades de informação percebe-se uma adesão ao padrão XML. Esse formato mostra uma linguagem de marcação bastante usada para descrever diversos tipos de dados na *web*. Os sistemas que lidam com o padrão MARC21 para catalogação e registro de autoridades fornecem um formato de saída de dados em XML, seja para processamento e intercâmbio, seja para a representação dos registros bibliográficos na *web*.

Diante o exposto, foi possível identificar pouca adesão das tecnologias relacionadas ao conjunto de práticas do *Linked Data*. As instituições parecem estar distantes de incorporar seus vocabulários controlados à nuvem de dados conectados.

#### **4.1.9 Percepção da colaboração nas etapas de construção**

Diante das práticas e ferramentas identificadas nas seções anteriores, os gestores foram convidados a relatar conforme suas experiências e percepções em quais etapas de construção o aspecto colaborativo é mais evidente. As respostas descritas no Quadro 42 revelam que o G1 não concorda haver experiência colaborativa na construção do vocabulário usado na instituição, mesmo ao relatar inúmeras vezes que os termos são extraídos de vocabulários mantidos por

outras instituições, sugeridos por outros profissionais da instituição e validados pelo setor de autoridades.

**Quadro 42** - Percepção da colaboração nas etapas de construção dos vocabulários controlados

Gestores	<b>De acordo com sua experiência, em qual etapa de construção ou manutenção do vocabulário controlado a colaboração é mais evidente? Por quê?</b>
G1	<i>Não se aplica. Não há experiência de colaboração na construção do vocabulário usado na instituição.</i>
G2	<i>No caso do vocabulário da [nome da instituição] a colaboração se deu em todas as etapas, porém a mais evidente é a de agora, na manutenção deste vocabulário.</i>
G3	<i>Acredito que na etapa de atualização/manutenção do vocabulário. Porque a atualização constante do Vocabulário Controlado [nome da instituição] somente é possível com a colaboração da equipe de bibliotecários indexadores da [equipe de gestão da informação acadêmica]. a partir do encaminhamento das sugestões de novos termos e do processo de revisão das áreas coordenado pelo grupo de gerenciamento do vocabulário que proporcionam a representação e a recuperação de informações que acompanham o dinamismo do desenvolvimento científico e tecnológico.</i>

**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Apenas o G2 e G3 relataram as etapas em que a colaboração parece ser mais evidente. O G2 aponta que em todas elas há algum tipo de colaboração, mas destaca a fase de manutenção como a mais relevante. Igualmente, o G3 menciona que são nas fases de manutenção e atualização do vocabulário que a equipe trabalha mais em conjunto. Prossegue argumentando que é “[...] a partir do encaminhamento das sugestões de novos termos e do processo de revisão das áreas coordenado pelo grupo de gerenciamento do vocabulário [...]” que se dá efetivamente a representação do conhecimento. Assim, as equipes de bibliotecários indexadores que encaminham as sugestões atuam com a equipe de gestão do vocabulário, demonstrando ser os principais atores sociais envolvidos com a construção do vocabulário controlado.

Na seção seguinte, serão apresentados os aspectos gerais das práticas colaborativas na construção dos vocabulários controlados do domínio da Música evidenciados com a pesquisa de campo.

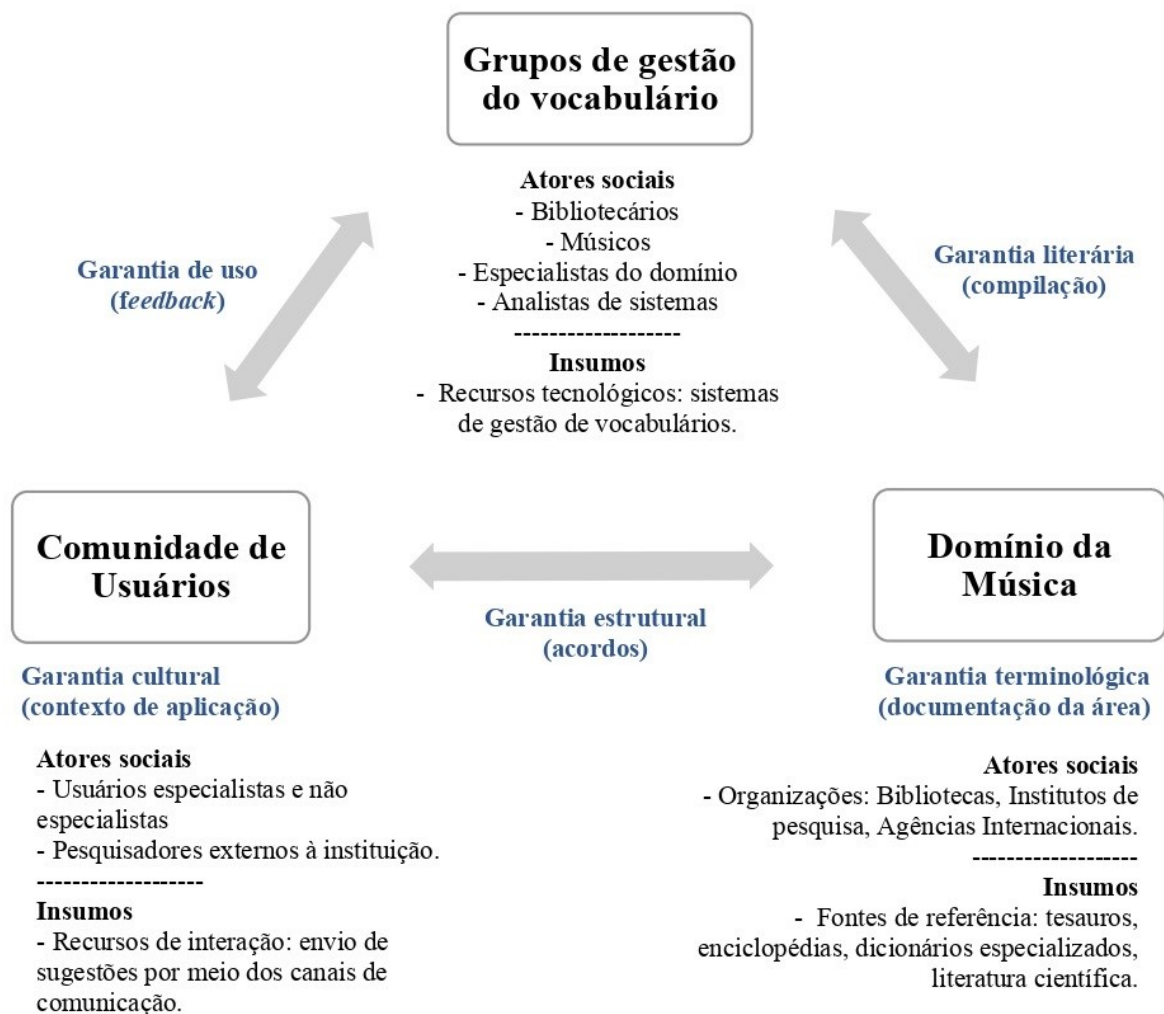
#### **4.1.10 Considerações gerais sobre as práticas colaborativas na construção de vocabulários controlados do domínio da Música**

Após a análise dos manuais técnicos, metodologias e políticas de indexação disponibilizados pelas instituições e dos relatos dos gestores conseguiu-se identificar as dimensões de colaboração, os principais atores sociais e os insumos que apoiam a construção dos vocabulários controlados. Esses atores podem estar distribuídos em três eixos principais: Grupos de gestão, Organizações ligadas ao Domínio da Música e a Comunidade de Usuários. A Figura 36 ilustra os principais aspectos dessa rede colaborativa.

Percebeu-se, ainda, que as principais atividades de edição são realizadas pelos grupos de gestão dos vocabulários controlados. Os atores sociais que integram a dimensão dos grupos de gestão dão suporte técnico, metodológico ou de especialidade do domínio, constituindo assim a equipe editorial do vocabulário, e são responsáveis por conduzir a construção. Podem estar organizados em grupos temáticos, nos quais profissionais e colaboradores contribuem conforme suas áreas de especialidade. No caso do domínio da Música, constatou-se a participação de usuários especialistas, músicos e docentes dos departamentos ligados às instituições participantes da pesquisa.

Entre os recursos colaborativos empregues pela equipe editorial destaca-se as reuniões presenciais, ferramentas de trabalho em fluxo (*workflow*) e *e-mail* e formulários eletrônicos. Esses recursos são aplicados para compartilhar decisões, estabelecer acordos, comunicar ações, sugerir melhorias e conceber o controle de edição, principalmente nas etapas de registro e validação de termos. Essas etapas evidenciaram o caráter colaborativo, mesmo em todas elas havendo a colaboração, como as de planejamento, desenvolvimento de políticas de gestão, atualização e revisão.

**Figura 36** – Visão geral das práticas colaborativas na construção dos vocabulários controlados



**Fonte:** Elaborado pelo autor com base nos dados da pesquisa (2020).

Os termos são compilados a partir de fontes de referência (dicionários especializados, enciclopédias, coleção de documentos da própria instituição), especialmente tesouros mantidos por outras instituições. Todas elas são reconhecidas pela comunidade bibliotecária e científica, as quais ajudam a executar a garantia literária e terminológica do vocabulário.

Os profissionais relataram consultar os principais vocabulários controlados do domínio da Música na etapa de compilação de termos, como: o *Library of Congress Subject Headings*, o *Library of Congress Genre/Form Terms*; o *Musical Instrument Museums Online*; o *Library of Congress Medium of Performance Thesaurus*; o Tesouro EuroVoc; o UNESCO Thesaurus; e a Terminologia de Assuntos da Biblioteca Nacional.

Visto a necessidade de adaptar os termos coletados, especialmente em função da tradução de termos coletados em vocabulários de outros idiomas, acordos devem ser estabelecidos e ajudam a executar a garantia estrutural. Nesse aspecto, termos que não se

apoiam nas garantias literárias e de uso podem ser admitidos com o propósito de melhorar a organização do vocabulário, garantindo que todos os termos se adequem à estrutura lógico-semântica.

Os relacionamentos entre os termos são essenciais para estabelecer a conectividade semântica. Alguns vocabulários buscam evitar termos órfãos, ou seja, que não possuem relacionamentos com outros termos do próprio vocabulário.

Ademais, percebe-se o papel da garantia cultural e da garantia do usuário em torno da rede de colaboração. Para projetar o vocabulário em função do público-alvo é essencial observar como a comunidade de usuários se comporta ao tentar recuperar informações no sistema de busca. A busca pode variar conforme o contexto cultural que esses sujeitos estão inseridos ou suas necessidades de uso.

No contexto de colaboração, identificou-se que os usuários podem fornecer *feedbacks* às equipes de gestão sobre a afetividade do uso dos vocabulários na recuperação da informação, as quais podem ser compostas por usuários especialistas, não especialistas e pesquisadores externos às instituições. Ao identificar termos inadequados ou desatualizados no vocabulário, a comunidade envia sugestões para a equipe editorial por meio de canais de comunicação, sendo estas avaliadas, e se comprovada a necessidade os termos indicados podem ser inclusos.

As evidências encontradas permitiram identificar os atores sociais e seus papéis dentro de uma rede colaborativa. Desse modo, buscou-se desenvolver um modelo que especifique as relações entre esses atores e a elaboração de tesouros do domínio da Música, a considerar a permeabilidade no âmbito dos projetos de LOD. Para isso, conseguiu-se estruturar um conjunto de princípios que norteiam a construção de tesouros, podendo estes serem compreendidos como fundamentos ou a base teórica-conceitual do modelo. Na próxima seção serão apresentados os princípios norteadores e como eles são discutidos na literatura.

#### 4.2 PRINCÍPIOS PARA A CONSTRUÇÃO E A PUBLICAÇÃO DE TESAuros COMO *LINKED OPEN DATA*

Esta seção atende ao cumprimento do objetivo específico b desta pesquisa: estruturar com base na literatura os princípios que norteiam a construção colaborativa e a publicação de tesouros em *Linked Open Data*. Na sequência, são apresentados os principais conceitos, padrões encontrados, a frequência que aparecem na literatura analisada e os autores que os destacaram.

#### 4.2.1 Estruturação e análise dos princípios norteadores

Realizou-se diversas análises no *corpus* teórico – especialmente nas seções que abordam as questões teóricas e metodológicas – para identificar os princípios que norteiam a construção e a publicação de tesouros no contexto do LOD. Esses princípios funcionam como uma base de conhecimentos ou um núcleo do sistema colaborativo que estabelecem as principais atividades de construção.

A construção e publicação de tesouros no contexto do *Linked Open Data* considera dimensões técnicas e legais para: garantir que todos os atores sociais envolvidos trabalhem de forma integrada; preservar a consistência e integridade dos dados; superar barreiras culturais e linguísticas quanto ao uso da terminologia; e garantir que as comunidades tenham acesso aos dados sem restrições de uso. A análise culminou em um conjunto de sete princípios gerais, conforme ilustrado na Figura 37.

**Figura 37** – Princípios para a construção e publicação de tesouros no contexto do LOD



**Fonte:** Elaborado pelo autor (2021).

Um tesouro construído e disponibilizado na *web* precisa ter esses elementos para que a sua elaboração ocorra de maneira efetiva e atenda às demandas exigidas em um ambiente de dados abertos conectados, podendo estes serem compreendidos como os fatores que estimulam a interação e a cooperação. Na sequência, serão discutidas as principais características.

### a) Consenso

Os tesouros são vocabulários estruturados que atentem a padrões internacionais de construção. São criados para dar resposta às comunidades de especialistas de um domínio e, por isso, buscam refletir o uso comum da linguagem. Todavia, definir conceitos compartilhados pela comunidade de usuários e estritamente relacionados ao domínio do tesouro pode ser uma tarefa bastante complexa. Geralmente, a elaboração é dinâmica e participativa, o que pode envolver uma equipe de construção formada por diversos profissionais.

Além de aspectos como seleção apropriada, a inclusão e exclusão de termos deve passar por um processo compartilhado de avaliação, de maneira que os relacionamentos semânticos possam ser estabelecidos corretamente. As divergências precisam ser discutidas amplamente, e para isso diferentes canais de comunicação devem ser utilizados.

Dessas questões, surge a problemática do consenso, mencionado nos estudos de (MCCULLOCH, 2005; GERBÉ; KERHERVÉ, 2010; LAPORTE; MOUGENOT; GARNIER, 2012; NOHAMA et al., 2012; ALBERTONI et al., 2014; PIOTROWSKI; COLAVIZZA; THIERY, 2014; WRIGHT; HARRISON; WATKINS, 2015; ÇAĞDAŞ; STUBKJÆR, 2015; MACULAN, 2015; BINDING; TUDHOPE, 2016; HARPRING, 2018; MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019b). O consenso pode ser definido como um acordo feito pela comunidade científica do domínio em relação ao uso da linguagem de representação, ou seja, o uso de termos e conceitos compartilhados por esse corpo social para se referir a quaisquer fenômenos da disciplina.

No campo da construção de tesouros, o consenso reflete os acordos produzidos também pelos membros da equipe de construção quanto à definição dos conceitos. É marcado pela uniformidade de opiniões e norteia a decisão quanto ao nível de abrangência ou especificidade do tesouro, ou quaisquer aspectos relacionados à edição, como inclusão ou exclusão de termos.

Considerando a abordagem pragmática no campo da Organização do Conhecimento, Maculan (2015) ressalta que a linguagem não é própria do indivíduo, mas desenvolvida para uso da comunidade, e prossegue inscrevendo que

[...] as regras a serem seguidas no uso do vocábulo é um “jogo” entre os sujeitos, que devem estabelecer consenso de conceitos em ambientes de uso da linguagem, visando à comunicação. Nesse sentido, prioriza-se o uso concreto da linguagem em distintos contextos sociais ou de especialidade, cujos significados são construídos pela interação de seus atores sociais, visando a alcançar determinados efeitos ou objetivos práticos (MACULAN, 2015, p. 66)

Dessa forma, pressupõe-se que devem ser consideradas todas as garantias (literária, estrutural e de uso) para que o tesouro possa atingir a completude ou, pelo menos, o máximo de consenso sobre o uso de conceitos do domínio. Não obstante, os propósitos de sua construção devem permanecer claros para que decisões sejam tomadas no âmbito dos grupos de gestão dos vocabulários.

Corroborando com a necessidade de considerar muitos atores sociais em relação ao uso compartilhado da linguagem, Piotrowski, Colavizza e Thiery (2014, p. 1) enfatizam que em algumas disciplinas os vocabulários precisam abranger o máximo de questões de pesquisa relacionadas ao domínio. Por isso, “Os principais desafios são obter um entendimento compartilhado dos termos entre todas as partes interessadas e encontrar o equilíbrio entre a amplitude da aplicação e a profundidade da análise”. Portanto, todos os procedimentos e acordos precisam ser estabelecidos e gerenciados por uma equipe interdisciplinar, porém com comprovada experiência no domínio.

O uso de fontes de referência em consonância com o método do comitê de especialistas ajuda a extrair o máximo de consenso não apenas sobre a relevância dos termos, mas também quanto aos seus relacionamentos (ÇAĞDAŞ; STUBKJÆR, 2015). Podem haver fontes de referência consolidadas que já desempenham um importante papel na disseminação da terminologia do domínio. Assim, para evitar duplicidade de tesouros, os gerentes dos projetos de construção precisam realizar um levantamento prévio dessas fontes e analisar se elas atendem às demandas da comunidade ou aos propósitos do tesouro que se pretende construir, considerando principalmente questões como atualização, o público-alvo e o escopo.

Apesar da existência de fontes de referência, não é possível garantir que elas atendam a propósitos específicos, muitas vezes diretamente relacionados à missão e aos objetivos das organizações as quais as terminologias estão subordinadas.

O gerenciamento de tesouros é altamente dinâmico e o tipo de decisão a ser tomada impõe desafios aos grupos que os elaboram. Por mais que hajam diretrizes bem definidas, Nohama et al., (2012) declaram que pode ocorrer desperdício de tempo quando uma especialista desfaz uma ação executada anteriormente por outros especialistas. Tais casos só podem ser resolvidos quando são aplicadas ferramentas de auditoria e orientadas ao consenso. Nesse caso, é preciso uma integração perfeita, disponível em muitos tipos de ferramentas, de maneira a monitorar as rotinas de alteração a fim de obter maior eficácia do processo. Uma vez identificados os erros, é essencial a revisão dos resultados para a classificação e correção pela equipe de construção.



No contexto do LOD, em que os tesouros são acessados e reutilizados livremente por diversas comunidades, espera-se que essas linguagens representem o conhecimento e as experiências consensuais para ganhar espaço e ter destaque como fontes de referência confiáveis dessas comunidades.

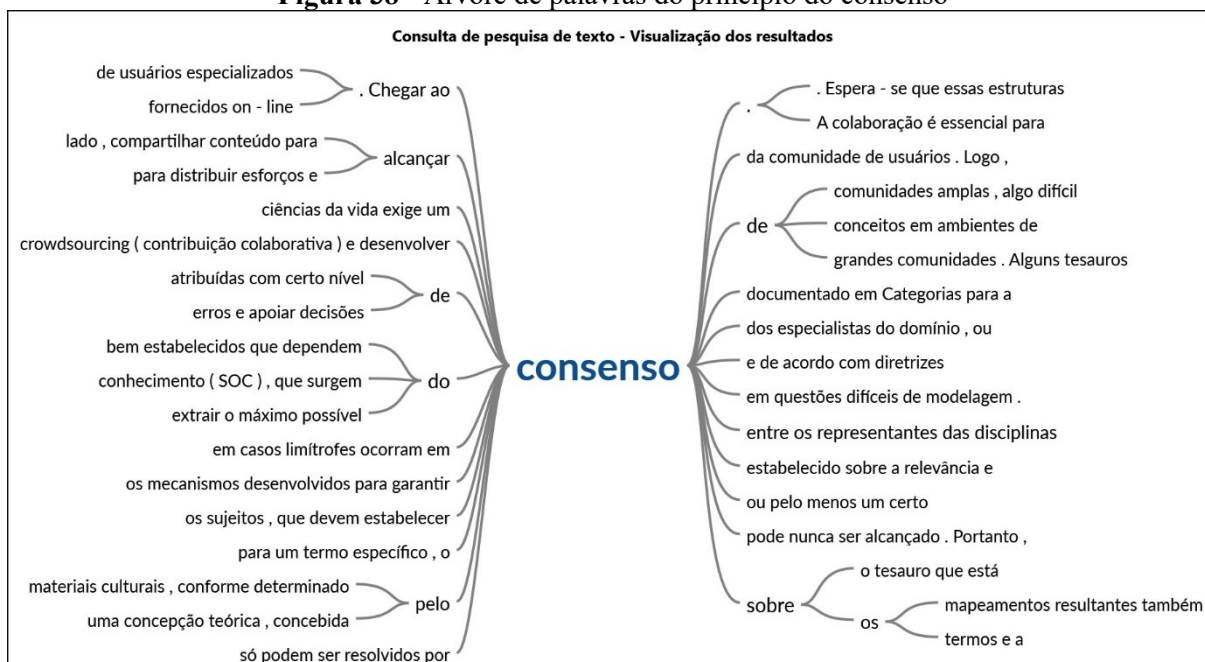
Os tesouros são Sistemas de Organização do Conhecimento (SOC), que surgem do consenso de grandes comunidades. Eles estão em uso há muitos anos e são atualizados regularmente. Enquanto no passado os tesouros eram projetados para profissionais da informação para indexação e pesquisa, hoje existe uma demanda por vocabulários conceituais que permitam inferir por máquinas (MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019b, p. 153151).

Portanto, sabe-se que no âmbito da *web* de dados abertos e conectados o tesouro é concebido como um tipo de ontologia leve, com ressaltos não apenas como um instrumento de indexação, mas também uma ferramenta terminológica que auxilia na inferência automática e processamento inteligente de dados. À vista disso, o princípio do consenso reflete na qualidade do tesouro, que uma vez construído seguindo esse critério pode ser reutilizado amplamente por aplicações na *web*.

Gerbé e Kerhervé (2010) enfatizam que o trabalho colaborativo atualmente é facilitado porque as ferramentas da WS ajudam as pessoas a compartilhar conhecimento ou dados na *web*. No entanto, antes que isso ocorra, é preciso ter um nível de concordância sobre a terminologia usada para descrever, recuperar e compartilhar esses recursos, ou seja, se não houver o mínimo de consenso em relação ao uso da terminologia a recuperação ou o processamento desses dados, tanto por pessoas quanto por máquinas, poderá ser prejudicada.

A Figura 38 apresenta a árvore de palavras que deriva as correlações semânticas entre a palavra consenso no *corpus* de análise. De maneira resumida, pode ser identificado em quais contextos o princípio do consenso é abordado.

**Figura 38 -** Árvore de palavras do princípio do consenso



**Fonte:** Fornecido pelo *software* Nvivo (2020).

É possível identificar alguns atores sociais (usuários especializados, especialistas do domínio ou comunidades científicas) interessados no consenso, o qual mediante ferramentas (como do tipo *crowdsourcing*, por exemplo) contribuem para alcançá-lo sobre os tesouros, os mapeamentos (entre vocabulários) ou o uso de termos. Um fator relevante constantemente observado nos corpora de análise é o uso de ferramentas tecnológicas para apoiar a decisão. Os aplicativos criados para a construção de tesouros ajudam a assegurar que os acordos sejam produzidos de maneira consensual.

De modo geral, o trabalho dos especialistas pode ocorrer primeiro mediante uma fase de edição, durante a qual são executadas várias ações relacionadas à construção do vocabulário (adição ou exclusão de termos e conceitos, alteração de definições, adição de um comentário, etc.), e uma fase de validação, em que especialistas podem validar ou invalidar os resultados das atividades concluídas durante aquela fase através de um procedimento de ponderação.

Nesse aspecto, Laporte, Mougenot e Garnier (2012) concordam que todas as atividades colaborativas de edição são definidas como um único processo de coconstrução. As decisões relativas às atividades de edição são conduzidas por diferentes especialistas e por isso pode não haver um consenso. Desse modo, os resultados do trabalho editorial são integrados a uma nova versão do tesouro somente após a fase de avaliação.

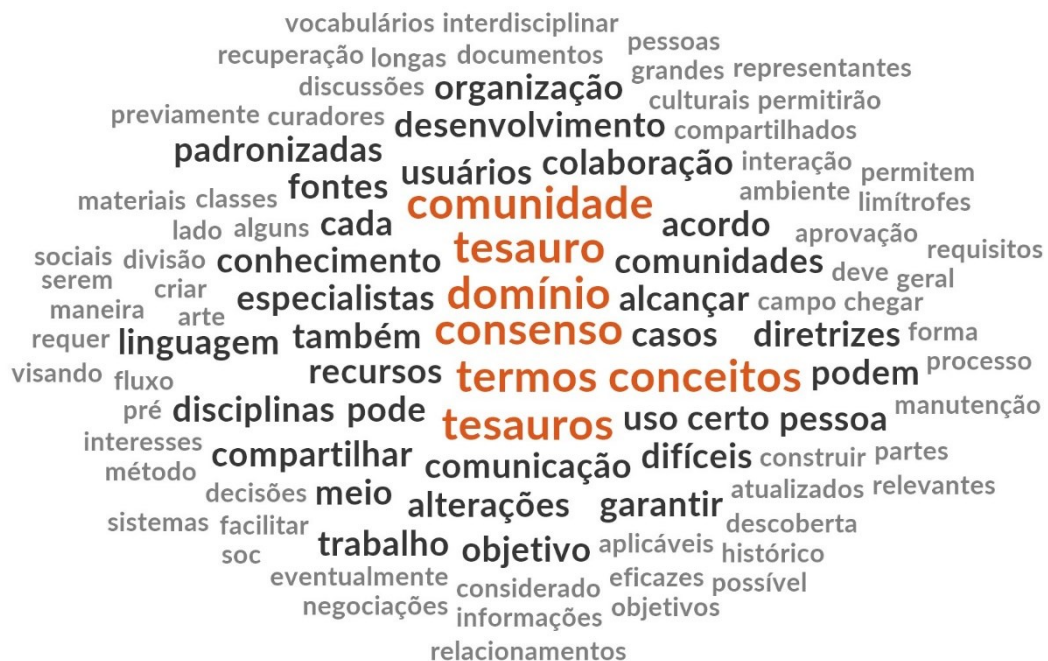
A validação é uma atividade que também compreende o princípio da avaliação da consistência, mas é norteador pelo princípio do consenso. Pode acontecer que diferentes

especialistas tenham visões distintas quanto à inclusão de determinados termos como parte de um grupo ou eixo temático na estrutura do tesouro. Desse modo, é necessário averiguar qual a função desse termo ou o uso comum na comunidade científica para incluí-lo de maneira mais adequada.

Nesse sentido, Mcculloch (2005, p. 405, tradução nossa) declara que “[...] pode ser prático atribuir áreas específicas do tesouro a subgrupos de consultores para limitar a deliberação, com uma única pessoa responsável pelo controle editorial geral.” Nota-se que os princípios do consenso e controle editorial são interdependentes, na perspectiva de que é preciso gerenciar os procedimentos para que eles reflitam as decisões ou os limites acordados por toda a equipe.

A nuvem de palavras expressa na Figura 39 enseja uma visão mais ampla de como o princípio do consenso está relacionado a outros fatores, como a comunicação, o compartilhamento e a colaboração entre os membros da comunidade científica, de especialistas ou usuários.

**Figura 39** - Nuvem de palavras do princípio do consenso



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Isto posto, decidir a respeito da inclusão de conceitos é uma tarefa subjetiva e necessariamente vinculada à experiência dos membros da equipe de construção. A maioria dos estudos analisados evidencia a importância de definir diretrizes para garantir a conformidade

de procedimentos e evitar erros que comprometam tanto a estrutura semântica como a efetividade da linguagem construída. É preciso pensar em espaços de discussão, sejam eles virtuais, sejam eles presenciais. Nesse aspecto, as ferramentas de comunicação e trabalho em grupo são colocadas como alternativas para integrar todos os profissionais envolvidos.

### **b) Controle editorial**

O controle editorial é o princípio que rege a manutenção compartilhada ou colaborativa dos dados do tesouro em um sistema de informação em rede. Pode estar relacionada à atividade de edição simultânea, sob o prisma de que dois ou mais profissionais possam executar ações no mesmo conjunto de dados sem que a qualidade do produto final seja prejudicada. Esse princípio é fundamental quando o tesouro é construído e usado por uma ampla comunidade de especialistas.

Estes profissionais podem estar geograficamente dispersos, apesar de que alguns tesouros são gerenciados de maneira centralizada. Assim, as contribuições são frutos do trabalho de muitos especialistas distribuídos em diversas regiões ou países. Portanto, é possível que algum tipo de limitação tecnológica interfira na manutenção.

Alguns vocabulários, como o AGROVOC, apoiam há décadas uma comunidade mundial de usuários e instituições que desenvolveram vários aplicativos baseados em modelos computacionais tradicionais para armazenar e manipular os dados dos tesouros. Nesses casos, pode ser essencial que alguns aplicativos façam interface com esses modelos legados (CARACCILO *et al.*, 2011). Todavia, faz-se fundamental ambientes bem elaborados e procedimentos e tecnologias uniformes para que o corpo social de usuários aproveite todos os benefícios da manutenção colaborativa. Desse modo, o suporte para versões anteriores é um requisito que não pode ser ignorado.

Por conseguinte, discussões sobre o controle editorial podem ser observadas nos seguintes estudos: (PASTOR SÁNCHEZ, 2009; ZHU; QIAO; SI, 2009; MORSHED *et al.*, 2010; BELLINI; CAPPUCCIO; NESI, 2011; STELLATO *et al.*, 2011, 2015; CARACCILO *et al.*, 2011, 2013; JORGE, 2011; NOHAMA *et al.*, 2012; LAPORTE; MOUGENOT; GARNIER, 2012; BENEVENTANO *et al.*, 2015; COBB, 2016; KEMPF; NEUBERT, 2016; ROSATI *et al.*, 2017; MOCHÓN; MÉNDEZ; BUENO DE LA FUENTE, 2017). Os autores em realce tratam de aspectos relacionados à editoração compartilhada, colaborativa e distribuída. Abordam a necessidade de estabelecer um fluxo de trabalho formal, no qual os profissionais

assumam uma função ou papel específico na edição. Esta evidência pode ser elucidada com a análise da árvore de palavras ilustrada na Figura 40.

Para gerá-la usou-se o parâmetro que identifica a frequência de palavras com generalizações. O processamento dos dados é fundamentado não apenas no número de vezes que a palavra aparece nas referências codificadas, mas também considera suas formas variantes. As palavras “tesauro” e “colaborativo” e suas formas variantes foram as mais frequentes nos fragmentos de texto que receberam o código “controle editorial”, resultando na árvore ilustrada. A árvore em alusão enfatizou que o trabalho em fluxo é um dos principais recursos que estabelecem o controle editorial.

**Figura 40** - Árvore de palavras do princípio do controle editorial



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Pode-se verificar que de modo geral a construção dos tesouros está diretamente relacionada ao gerenciamento dos grupos de trabalho que executam suas atividades em um fluxo de operações. Este fluxo é suportado pelos recursos ou ferramentas que ajudam as equipes a colaborar com os tesouros, com realce para as etapas de manutenção e validação da estrutura de conceitos.

Segundo Stellato et al. (2015), a colaboração é essencial para que se possa distribuir esforços e alcançar consenso sobre o tesauro que está em desenvolvimento. Para facilitar a colaboração, as ferramentas devem fornecer um fluxo de trabalho formal cujas alterações dos editores são rastreadas e armazenadas para aprovação pelos validadores. Esse gerenciamento é

suportado pelo recurso de controle de acesso baseado em uma função que fornece finalidades diferentes para cada usuário, categorizados conforme suas especialidades e responsabilidades editoriais.

Na maioria dos projetos, especialistas em terminologias são responsáveis por criar os tesouros. No entanto, é preciso oferecer uma plataforma na qual vários profissionais possam fazer contribuições. No estudo de Zhu, Qiao e Si (2009), três tipos diferentes de colaboradores foram identificados: editor comum, especialistas em domínio e administrador.

Na prática, todos esses colaboradores têm autoridade para criar, modificar e excluir um conceito. No entanto, quando um conceito é criado por cada um desses colaboradores, o conceito precisa ser verificado por outros colaboradores por meio do **fluxo de trabalho**. O processo de verificação pode ser descrito como "**rascunho de envio - verificação - verificado**". Durante esse processo, diferentes colaboradores podem fazer sugestões para alterar os conceitos e solicitar a aprovação de outros colaboradores (ZHU; QIAO; SI, 2009, p. 286, tradução nossa, grifo nosso)

Nesse sentido, o princípio do consenso ganha evidência. As estruturas hierárquicas são construídas e avaliadas por muitos profissionais envolvidos nos projetos, os quais são guiados por diretrizes definidas e compartilhadas com todos da comunidade. Por conseguinte, assegura que o produto final reflete a terminologia de uso comum do domínio. Como a terminologia pode mudar ou evoluir ao longo do tempo a construção requer uma equipe de especialistas em domínio e cientistas da informação que, segundo Kempf e Neubert (2016), devem trabalhar conjuntamente.

De modo igual, Caracciolo et al. (2013) afirmam que a elaboração do tesouro como dados conectados se apoia em três níveis gerenciais: a manutenção de dados, armazenamento de dados e a publicação de dados. Os tesouros, quando construídos e mantidos por comunidades distribuídas, devem dispor de uma plataforma comum a fim de integrar todos os recursos compilados e oferecer o suporte para que esses níveis sejam mantidos. As ferramentas de construção devem assegurar que a consistência de todas as réplicas de dados seja mantida, mesmo quando os colaboradores executarem modificações em ordens diferentes.

Como os dados também podem estar armazenados de forma distribuída, as plataformas desenvolvidas para gerenciá-los necessitam de mecanismos que assegurem o controle de operações simultâneas, de maneira que quando dois ou mais profissionais alterarem ao mesmo tempo nos conjuntos de dados o sistema possa preservar a estrutura lógico-semântica original até que uma versão mais estável do tesouro seja validada pelo gerente do projeto. Caso o sistema não ofereça a funcionalidade, o usuário final pode acessar uma versão não confiável do tesouro.

A escolha de um *software* ou de outro dependerá da avaliação de diferentes fatores, como recursos financeiros disponíveis, número de estruturas de dados a serem operadas simultaneamente, requisitos de controle de integridade, controle automático de eventos em cascata ou grau de velocidade da resposta de operação do aplicativo, entre outros (PASTOR SÁNCHEZ, 2009, p. 145)

A escolha da plataforma adequada é requisito a ser considerado no planejamento do tesouro, pois é a partir dela que todas as etapas de construção são operacionalizadas. Em um ambiente colaborativo, onde os usuários podem editar proativamente um recurso compartilhado, é importante ter meios para monitorar os procedimentos. Isto posto, a capacidade de controlar alterações recentes no tesouro é essencial em razão da necessidade de detectar alterações coocorrentes e coordenar o perfil de outros editores. Portanto, os estudos analisados recomendam que a plataforma tecnológica empregada deve oferecer mecanismos síncronos de trabalho com vistas a garantir que as atividades de colaboração possam ocorrer.

Ainda em relação às ferramentas, o ambiente operacional de sistemas baseados na *web* deve ser adaptado ao modelo de computação em nuvem, visto oferecer condições apropriadas para a criação colaborativa e um bom nível de interoperabilidade entre os formatos de metadados. Autores como Mochón, Méndez e Bueno de La Fuente (2017) e Laporte, Mougnot e Garnier (2012) ressaltam que esses programas de gerenciamento de tesouros devem fornecer relatórios ou registrar a data e a hora de criação e das modificações, além de identificar o usuário responsável por cada alteração, facilitando com o ato a supervisão e o controle das operações.

Como a maioria das ferramentas atuais possibilita aspectos colaborativos, a supervisão e o controle dos processos de interação devem ser cuidadosamente considerados. Cada especialista precisa ser informado em tempo real acerca das alterações trazidas ao tesouro por seus pares. Igualmente, é importante que o sistema forneça informações quanto ao número de especialistas que contribuíram para o tesouro, a distribuição geográfica desses especialistas, o número de conceitos modificados, adicionados ou excluídos, ou, ainda, aqueles que foram mantidos inalterados. Em caso de desacordo referente ao conceito, termo ou definição em particular os especialistas podem se comunicar indiretamente mediante esses ambientes.

Outra vantagem inerente à supervisão e o controle são que eles oferecem um alto grau de confiança na qualidade e na consistência do tesouro em desenvolvimento. Diante disso, um tesouro editado e validado pela maioria dos especialistas de um domínio terá alta credibilidade em termos de representação do conhecimento dessa comunidade.

Dada a natureza multilíngue e intrinsecamente colaborativa de parte significativa dos tesouros publicados na *web*, há uma necessidade de elaboração de ferramentas que ofereçam suporte para a edição compartilhada, bem como a validação de alterações em um fluxo de

trabalho editorial formalizado também no nível do idioma (CARACCILO et al., 2013). Os tesouros multilíngues são estruturas complexas que buscam representar conceitos, na maioria dos casos, difíceis de encontrar correspondentes semânticos no idioma de destino.

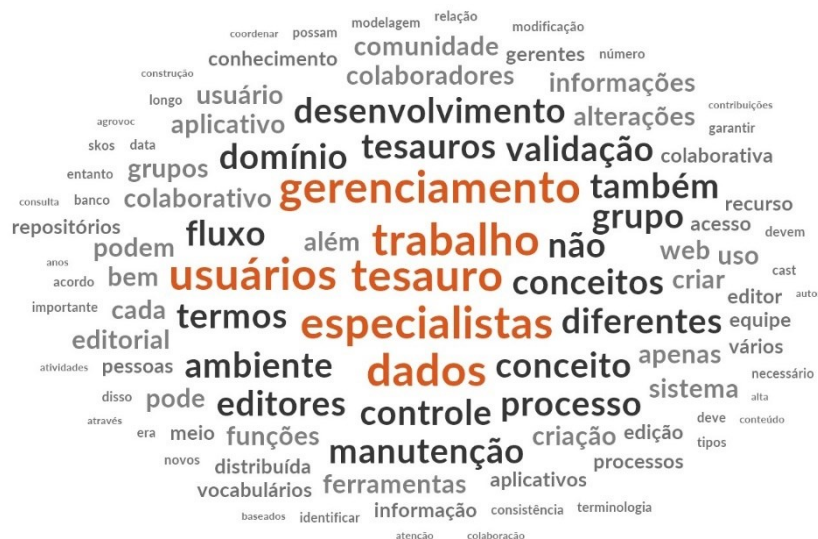
Nesse contexto, Cobb (2016, p. 143) enfatiza a participação de diferentes instituições na construção de vocabulários do Getty Research Institute, na perspectiva de que

Os vocabulários são avaliados como recursos de autoridade que crescem por meio de contribuições de especialistas e contêm citações para as instituições contribuintes e para as fontes bibliográficas dos termos, nomes e outras informações nele contidas. [...] As contribuições são de instituições e organizações da comunidade internacional de organizações do conhecimento, incluindo repositórios de arte e patrimônio cultural, além de projetos relacionados à indexação e catalogação de arte, arquitetura, cultura material e conservação (COBB, 2016, p. 143, tradução nossa)

O controle de qualidade do conteúdo, tanto em aspectos terminológicos como estruturais, deve ser contínuo. Nesse contexto, precisam ser consideradas a possibilidade de edição *offline* para atender às necessidades de usuários com baixa velocidade de conexão à Internet e melhorias contínuas na usabilidade, incluindo a adaptação da interface do usuário às comunidades de vários idiomas. Esses aspectos foram identificados como problemáticos quando se pensa em editar simultaneamente dados compartilhados e distribuídos, porém é preciso prestar uma assistência comum a todos os colaboradores, visto que o trabalho em rede é fundamental para construir e vincular dados de tesouros em uma escala global.

A nuvem de palavras ilustrada na Figura 41 reflete os principais temas abordados no *corpus* de análise e mencionados nesta seção.

**Figura 41** - Nuvem de palavras do princípio do controle editorial



**Fonte:** Fornecido pelo *software* Nvivo (2020).



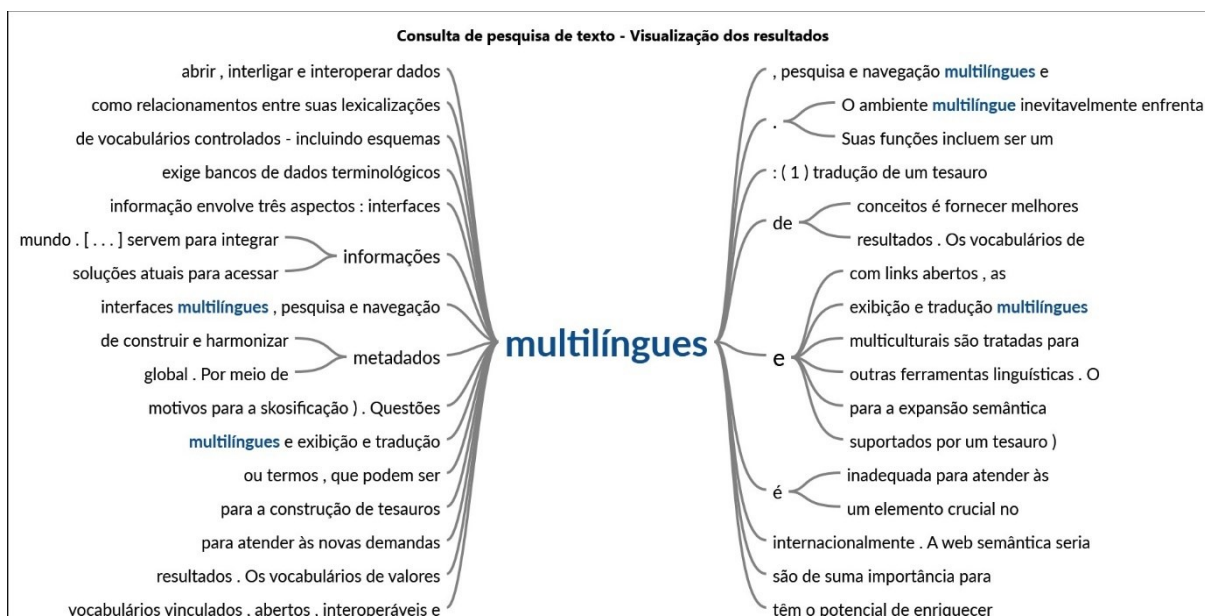
Em geral, o princípio do controle editorial pode desencadear abordagens promissoras para construir infraestruturas que integrem dados dos tesouros, deixando o desenvolvimento e a manutenção destes descentralizada com os criadores, mas alavancando o valor do tesouro local por sua conexão e interligação. Essas infraestruturas compartilhadas podem representar uma economia de recursos numa esfera global e aumentar a valia da linguagem de representação localmente.

### **c) Representação multilíngue**

A representação multilíngue é tratada neste estudo como a capacidade do tesouro reproduzir, estruturar e disponibilizar seus conceitos em dois ou mais idiomas; busca superar as barreiras culturais da organização do conhecimento e facilitar a recuperação da informação em ambientes informacionais multilíngues. No contexto do LOD, os tesouros multilíngues tornam possível os vínculos entre os conceitos de diferentes idiomas e SOC na *web* e permitem que os conjuntos de dados sejam explorados e reutilizados por comunidades de diferentes nacionalidades e culturas.

A importância da representação multilíngue dos tesouros é citada nos estudos de (PASTOR SÁNCHEZ, 2009; MORSHED et al., 2010; BELLINI; CAPPuccio; NESI, 2011; JORGE, 2011; BACA; GILL, 2014; ÇAĞDAŞ; STUBKJÆR, 2015; VANDERBILT et al., 2017; NIININEN; NYKYRI; SUOMINEN, 2017; ALBERTONI; DE MARTINO; PODESTÀ, 2018; FABEIRO; MASAGUÉ; ROVIRA JARQUE, 2019; MOCHÓN-BEZARES; MÉNDEZ-RODRÍGUEZ; LA FUENTE, 2020). A árvore de palavras ilustrada na Figura 42 mostra que as pesquisas tratam de aspectos como multiculturalidade, interoperabilidade, suporte à busca, navegação e recuperação e enriquecimentos de dados.

**Figura 42 -** Árvore de palavras do princípio da representação multilíngue



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Caso um tesouro local utilize um dado termo no idioma de origem para indexar um conjunto de documentos de sua coleção e esteja vinculado a um termo correspondente em outro idioma, as possibilidades de pesquisa e recuperação da informação são expandidas e os usuários podem acessar e explorar um conjunto de dados mais diversificado. Com isso, as instituições e repositórios que trabalham com o LOD sublinham que vocabulários abertos e vocabulários conectados precisam considerar o aspecto multilíngue, visto que dessa forma podem contribuir com a pesquisa semântica e são capazes de enriquecer os metadados descritivos, fornecendo pontos de acesso adicionais para um conceito.

Quanto à recuperação da informação, Bellini, Cappuccio e Nesi (2011) ressaltam que as questões multilíngues contribuem para uma exploração mais ampla e eficaz dos dados, porquanto ajudam os sistemas de busca a melhorar na recuperação de informações. Igualmente, Vanderbilt *et al.* (2017) endossam que uma das principais vantagens do tesouro multilíngue sobre a pesquisa de texto livre é a precisão dos resultados, cujas coleções são indexadas com o uso de termos preferidos, os significados são conhecidos e compartilhados pela comunidade do domínio e os sistemas de recuperação podem trazer resultados mais precisos. Ademais, o uso do tesouro também pode reduzir retornos irrelevantes, geralmente causados pela ambiguidade e incompatibilidades na tradução desses termos.

Além dos termos multilíngues, ganha destaque nesse cenário as interfaces de navegação e as possibilidades de exploração dos dados, que devem apresentar mecanismos que

facilitem o entendimento dos usuários relativos às diferentes opções de busca e exibição dos resultados. As comunidades do LOD estão cada vez mais incorporando dados multilíngues para enriquecer os conjuntos de dados e torná-los mais interoperáveis. Quanto a esses aspectos, considera-se que

Os princípios dos dados vinculados são fornecer dados para serem compartilhados, expandidos, usados e reutilizados no contexto da *web* semântica e em um formato independente de idioma, baseado em identificador, que permita a funcionalidade multilíngue. Esses recursos permitem a descrição de recursos e o enriquecimento de conjuntos de dados de maneira eficiente e colaborativa (FABEIRO; MASAGUÉ; ROVIRA JARQUE, 2019, p. 5, tradução nossa)

Os tesouros na *web* crescem com contribuições de organizações da comunidade internacional. Essas parcerias aprimoram o desenvolvimento da terminologia multilíngue. A tradução dos vocabulários para um idioma de destino é trabalhosa, exige equipes de especialistas no idioma a fim de mapear seus vocabulários locais para vocabulários externos. Pode ocorrer que um conceito muito utilizado localmente esteja ausente no vocabulário de destino, logo um novo conceito deve ser enviado para a inclusão.

A estrutura de dados e os princípios básicos sob os quais os vocabulários são construídos e mantidos enfatizam o multilinguismo e o multiculturalismo. Baca e Gill (2014) frisam que a elaboração de tesouros multilíngues é fruto de um trabalho minucioso e somente são possíveis graças à cooperação de diversos profissionais e instituições de interesse no domínio. Os termos são compilados mediante contribuições ao longo do tempo e os conjuntos de dados evoluem constantemente com a adição de novos termos e conceitos. Portanto, o desenvolvimento dos tesouros multilíngues é naturalmente ligado às atividades de colaboração, visto serem construídos principalmente a partir de contribuições de instituições parceiras.

Como os tesouros são considerados recursos de autoridade, as parcerias concernentes às atividades de construção multilíngue precisam ser estabelecidas com instituições confiáveis e profissionais capacitados. A participação dos usuários nesse processo pode ser considerada, porém com cautela, tendo em vista que o emprego ou o significado dos termos podem variar expressivamente em diferentes regiões ou culturas.

Cada cultura conceitua o mundo a partir de seu próprio ponto de vista; os significados raramente são simétricos entre as línguas. Portanto, o objetivo não é buscar a equivalência exata entre os idiomas, mas conduzir o usuário a resultados de pesquisa relevantes, independentemente do idioma usado na consulta. No entanto, um ambiente multilíngue apresenta vários desafios relacionados ao idioma e à cultura, e a construção de uma hierarquia completa

e consistente em mais de um idioma é um processo complexo que requer acordos (NIININEN; NYKYRI; SUOMINEN, 2017, p. 458, tradução nossa)

A construção multilíngue dos tesouros pode enfrentar desafios em volta de questões de idioma, das culturas e do aspecto inter vocabulários. Pode ocorrer que um conceito não possa ser representado no idioma de destino com um simples rótulo equivalente. Ademais, as culturas são baseadas em conjuntos de sistemas, valores e práticas e podem carregar diferentes entendimentos do mundo.

Essas diferenças são sentidas também quando se busca vincular conceitos entre vocabulários, pois nem sempre é possível conectar idiomas de duas esferas culturais diversas (NIININEN; NYKYRI; SUOMINEN, 2017). Por isso, a maioria dos tesouros em LOD que buscam mapear correspondentes semânticos em outros vocabulários de idiomas heterogêneos não o fazem em sua completude.

Assim, em razão da forte vinculação dos dados em nível global, divergências estruturais que possam envolver tesouros multilíngues devem ser minimizadas. Os tesouros são criados com diferentes pontos de vista e podem se basear em inúmeras formas de conceituação. Embora os projetos de tesouros em LOD se deparem com essa dificuldade, observa-se que as instituições têm se esforçado para mapear e vincular termos em diferentes idiomas. Ainda, que as adversidades podem ser minimizadas quando esses mapeamentos são estabelecidos em cooperação com muitos parceiros.

O multilinguismo precisa ser levado em consideração, não apenas pela variedade de tesouros nacionais que necessitam ser integrados, mas também pela futura criação de ferramentas terminológicas comuns para superar barreiras culturais. Construir ou vincular vocabulários multilíngues é um esforço colaborativo que tem como foco a integração de infraestruturas de dados existentes para permitir o uso de conjuntos de dados e serviços distribuídos.

Dessa maneira, é fundamental projetar mais aplicativos para abrir e consumir dados multilíngues internacionalmente. Para Xian e Sun (2017), na era do *Big Data* a ciência é cada vez mais orientada a dados e mais recursos de dados serão compartilhados, abertos e disponíveis em todo o mundo. *Big Data*, LOD e SOC seriam partes integrais da futura infraestrutura da *web*. Em razão disto, é importante que as organizações semelhantes alonguem e interliguem seus dados, permitindo a interoperabilidade de seus dados com outros dados ou sistemas em todo o mundo em um contexto multilíngue.

A Figura 43 ilustra a nuvem de palavras com aquelas que tiveram maior representatividade no *corpus* de análise em torno do aspecto multilíngue.

**Figura 43** - Nuvem de palavras do princípio da representação multilíngue



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Apesar dos muitos desafios, todos os esforços devem ser considerados para que a dimensão multilíngue dos tesouros possa ser maximizada, haja vista ser um requisito fundamental para se alcançar a interoperabilidade entre vocabulários. Uma vez os tesouros publicados a partir de modelos de dados adotados pela comunidade do LOD e conectados com vocabulários de diversos idiomas, tornam-se ferramentas terminológicas mais acessíveis e aumentam o potencial de reuso por outras aplicações.

#### **d) Interoperabilidade**

De modo geral, a interoperabilidade concerne à capacidade de um sistema trocar ou compartilhar informação; considera essencialmente suas características sintáticas ou estruturais. Assim, ao tratá-la pensa-se em dois tipos: interoperabilidade lexical e interoperabilidade semântica. Nesta última, a atenção se volta não apenas para a compatibilidade de termos no nível lexical, mas principalmente ao nível conceitual, ou seja, no significado desses termos em diferentes sistemas de informação.

No âmbito dos tesouros, a interoperabilidade é viabilizada quando vínculos são estabelecidos entre termos de diferentes vocabulários. No contexto do LOD, estabelecer esses relacionamentos com recursos externos além de facilitar a identificação de correspondentes lexicais ou semânticos em conjuntos de dados heterogêneos, tem o potencial de estabelecer uma base lógica e conceitual consistente que ajuda aos usuários a explorar e recuperar dados contextuais diretamente relacionados aos domínios de aplicação.

Assim, observou-se menções à interoperabilidade nos estudos de (ISAAC; MATTHEZING, 2007; ZHU; QIAO; SI, 2009; SOLOMOU; PAPTODOROU, 2010; PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012; WETZ et al., 2012; LARA, 2013; PASTOR-SÁNCHEZ, 2013; SANTOS et al., 2013; SPAMPINATO; ZANGARA, 2013; ZAPILKO et al., 2013; BERGAMIN; BERGAMIN; LUCARELLI, 2013; ALBERTONI et al., 2014; PIOTROWSKI; COLAVIZZA; THIERY, 2014; LAPORTE et al., 2014; MACULAN et al., 2015; WRIGHT; HARRISON; WATKINS, 2015; L'ABATE et al., 2015; ALEXIEV; ISAAC; LINDENTHAL, 2016; BINDING; TUDHOPE, 2016; KEMPF; NEUBERT, 2016; NEUBERT, 2017; ROSATI et al., 2017; XIAN; SUN, 2017; ALBERTONI; DE MARTINO; PODESTÀ, 2018; MAZIARZ; PIASECKI, 2018; VITI, 2018; ZOU, 2018; DIMITROVA et al., 2018; MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019b; MORAIS; RAMALHO, 2019; MOCHÓN-BEZARES; MÉNDEZ-RODRÍGUEZ; LA FUENTE, 2020).

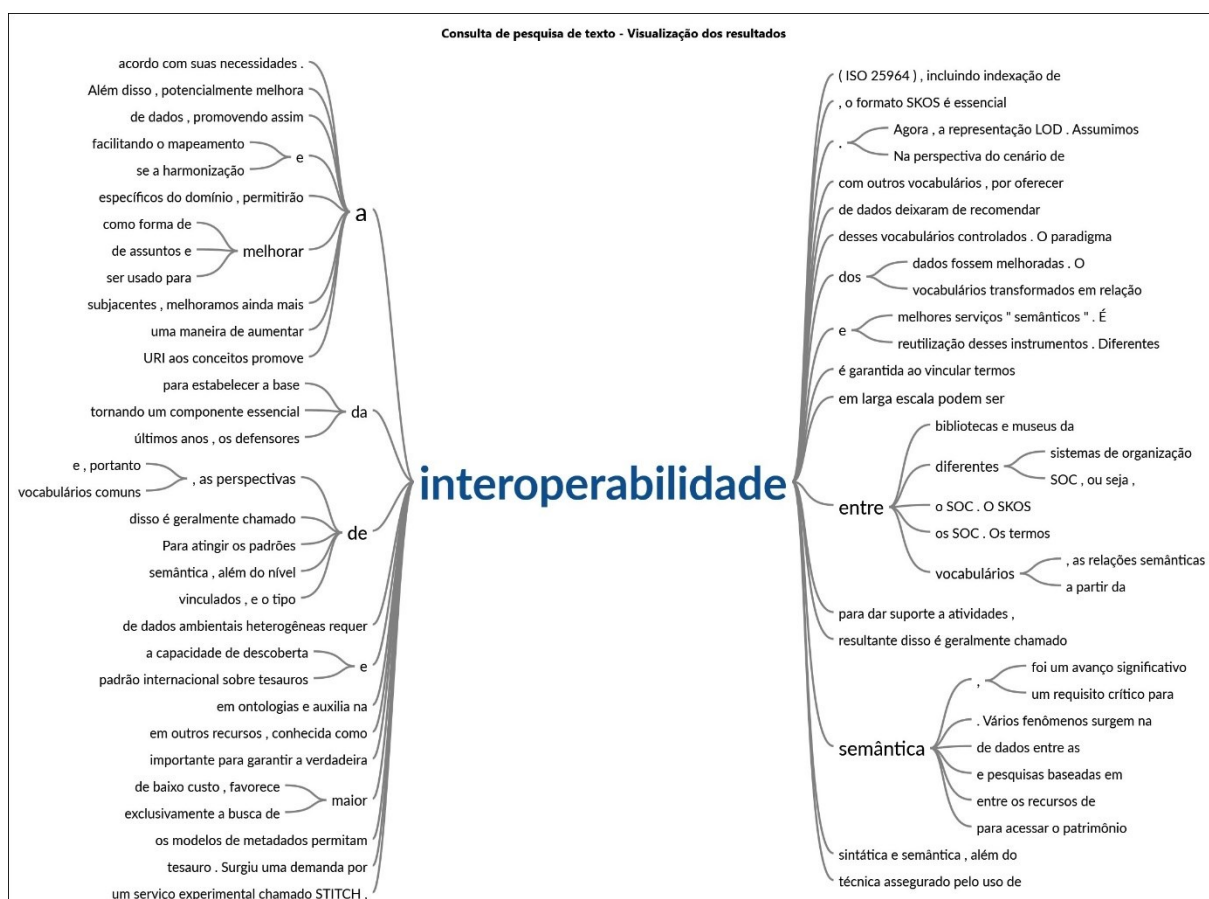
Neste escopo, temas como mapeamento de correspondências e enriquecimento semântico estão naturalmente relacionados ao princípio da interoperabilidade. Enriquecimento semântico é um termo abrangente usado para descrever a incorporação de representações ou descrições adicionais às já existentes, possibilitando que as máquinas compreendam e obtenham conhecimento sobre as relações existentes entre os vários itens de dados (SIMOU; CHORTARAS; STAMOU, 2017). No contexto das tecnologias WS e LOD, não basta mais conceber uma lista fechada de conceitos e usá-las como vocabulário para caracterizar semanticamente um conjunto de dados. É desejável usar vocabulários estabelecidos e compreensíveis por máquina para que os dados possam ser reutilizados e relacionados com outras fontes de dados semelhantes.

A capacidade aumentada de combinar dados proprietários com dados existentes em outros recursos, conhecida como interoperabilidade semântica, foi um avanço significativo para muitos tipos de SOCs, especialmente tesouros e ontologias. A conversão do SOC para a *Web* não deve se limitar a uma mera transferência de conteúdo para um formato eletrônico que só pode ser interpretado por pessoas, mas também deve ser compreensível por computadores, para que eles possam gerenciar melhor os recursos

(MOCHÓN-BEZARES; MÉNDEZ-RODRÍGUEZ; LA FUENTE, 2020, p. 55, tradução nossa)

À vista disso, a grande vantagem do uso de padrões da WS para a construção de tesouros está na capacidade de fazer com que humanos e máquinas possam conceber significado dos dados. A interoperabilidade é um requisito que deve ser considerado para os vocabulários publicados como LOD. As relações entre a palavra interoperabilidade e fragmentos codificados no *corpus* de análise, ilustrados na Figura 44, ressaltam sua contribuição, no sentido de melhorar os serviços semânticos e potencializar o reuso de dados dos SOC em diferentes tipos de aplicações.

**Figura 44 -** Árvore de palavras do princípio da interoperabilidade



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Na prática, a interoperabilidade está associada à capacidade de mapear conceitos correspondentes em diferentes tesouros e demais vocabulários em dois níveis: sintáticos e semânticos. As correspondências entre termos podem ocorrer no nível dos rótulos e/ou no nível dos conceitos. No entanto, para que a interoperabilidade semântica seja assegurada é

fundamental realizar comparativos entre vocabulários controlados no nível conceitual. Esses mapeamentos podem ser unidirecionais – ou seja, de um tesouro para o outro – ou recíprocos – quando são verificadas as correspondências mútuas entre os dois vocabulários.

A principal dificuldade em realizar os mapeamentos inter vocabulários está na maneira de fazê-lo; podem ser realizados de maneira manual por especialistas, automática ou semiautomática. Apesar da marcação manual oferecer uma precisão devido a participação de especialistas do domínio, ela não pode ser aplicada a grandes volumes de dados. Para esses casos em particular, a alternativa é utilizar o mapeamento automático, cuja geração dos valores de metadados é sucedida automaticamente por um sistema de *software*.

Esses sistemas contam tradicionalmente com técnicas de extração e PLN, desambiguação de palavras, recuperação de informações e aprendizado de máquina. A maior vantagem é que as abordagens automáticas executam procedimentos em maiores quantidades de dados de maneira rápida, porém, a depender do domínio, da forma e da qualidade dos dados, elas podem ter relativamente um baixo nível de precisão. Zapilko et al. (2013) enfatizam o uso do abordagens semiautomáticas, nas quais os sistemas de *software* geram alguns mapeamentos e, em seguida, os especialistas do domínio revisam-nos para validá-los ou rejeitá-los

O principal objetivo das ferramentas de mapeamento é permitir a descoberta de novos vínculos, não só em tesouros conhecidos, mas também em outros conjuntos de dados na *web*. Logo, para possibilitar a colaboração e o intercâmbio semântico, o uso de padrões e modelos de dados compartilhados pela comunidade do LOD são requeridos como RDF, SKOS e OWL, por exemplo. Os vocabulários controlados compartilhados permitem a combinação e a análise conjunta de dados de diferentes fontes, além de implementar ferramentas genéricas que ajudam a acessar e processar semanticamente os dados, como as API (PIOTROWSKI; COLAVIZZA; THIERY, 2014).

O mapeamento pode ocorrer entre os mesmos tipos de vocabulários ou para diferentes tipos de bases de conhecimento como, a exemplo, o Wikidata<sup>54</sup>. Neubert (2017, p. 12) ressalta que “devido à natureza do Wikidata, a rede de conceitos de muitos domínios é acessível e mantida como uma única base de conhecimento aberta e colaborativa”. Inscreve, ainda, duas razões em mapear um tesouro para o Wikidata: em primeiro lugar, o tesouro estará interligado com várias páginas da Wikipedia em diferentes idiomas, o que pode ajudar a estender as

---

<sup>54</sup> Lançado pela fundação Wikimedia em 2012 com o objetivo de criar uma base de conhecimento compartilhada e fornecer dados estruturados para as Wikipédias em diferentes idiomas, *Wikimedia Commons* e outros projetos da fundação.



explicações e contextos de conceitos; em segundo lugar, é possível explorar os *links* para outros SOC que já foram mapeados para o Wikidata.

A tarefa de encontrar correspondentes semânticos entre o tesauro e os recursos de conjuntos de dados externos pode ser feita por meio de serviços de pesquisa que consulta a nuvem LOD para então derivar um rótulo semântico correspondente no tesauro local. Segundo Wetz et al. (2012), essa abordagem pode ser chamada de interligação ou alinhamento de entidades que trazem um benefício para o tesauro, pois “a estratégia de vincular entidades LOD a um tesauro local usa os conceitos de dados vinculados para expandir e enriquecer as informações armazenadas no tesauro, levando a uma base de conhecimento mais valiosa.” (WETZ et al., 2012, p. 7).

Binding e Tudhope (2016), por sua vez, consideram que embora haja um vínculo conceitual rico intra tesauros pode não haver um *link* semântico correspondente de um conceito para diferentes tesauros, ou seja, um conceito em um tesauro local talvez não possua um rótulo correspondente em vocabulários semânticos externos, o que tornaria a participação de especialistas humanos ainda mais necessária. Nesse contexto, um ambiente colaborativo para incentivar a participação de muitos atores sociais pode ser fundamental, pois ajuda a desenvolver consenso sobre os mapeamentos resultantes.

Assim, para garantir que as correspondências representem reais contribuições, a equipe de construção precisa selecionar primeiramente os conceitos no tesauro local, identificar os termos correspondentes e especificar ou refinar esses relacionamentos. Essa ação pode ocorrer com base na consulta de especialistas, pelo método empírico ou uma combinação desses métodos.

Apesar da abordagem de consulta de especialista extrair o máximo possível do consenso estabelecido sobre a relevância e os relacionamentos dos termos, o método empírico é o mais aplicado na publicação de tesauros em LOD devido ao grande número de ferramentas já desenvolvidas para esse fim (ÇAĞDAŞ; STUBKJÆR, 2015). Desse modo, técnicas simples de correspondência de *strings* são bastante apropriadas para fornecer *links* candidatos em uma estrutura de dados conectados, porém a avaliação humana deve confirmar o maior número de correspondências encontradas, principalmente em conjuntos de dados multilíngues (MORSHED et al., 2011).

Albertoni et al. (2014) enfatizam que a vinculação dos tesauros com recursos terminológicos externos garante uma ampla cobertura conceitual, fortalece o reconhecimento da dimensão cultural da organização do conhecimento e permite diferentes níveis de compreensibilidade e aplicabilidade para usuários com conhecimentos diferentes. A nuvem de



informações necessárias têm que estar presentes no conjunto de dados. Pode se referir aos esquemas de conceitos (nível ontológico), nível de valores ausentes em propriedades específicas do modelo de dados utilizado, ao nível de completude das vinculações ou interligações com diferentes fontes de dados contextuais.

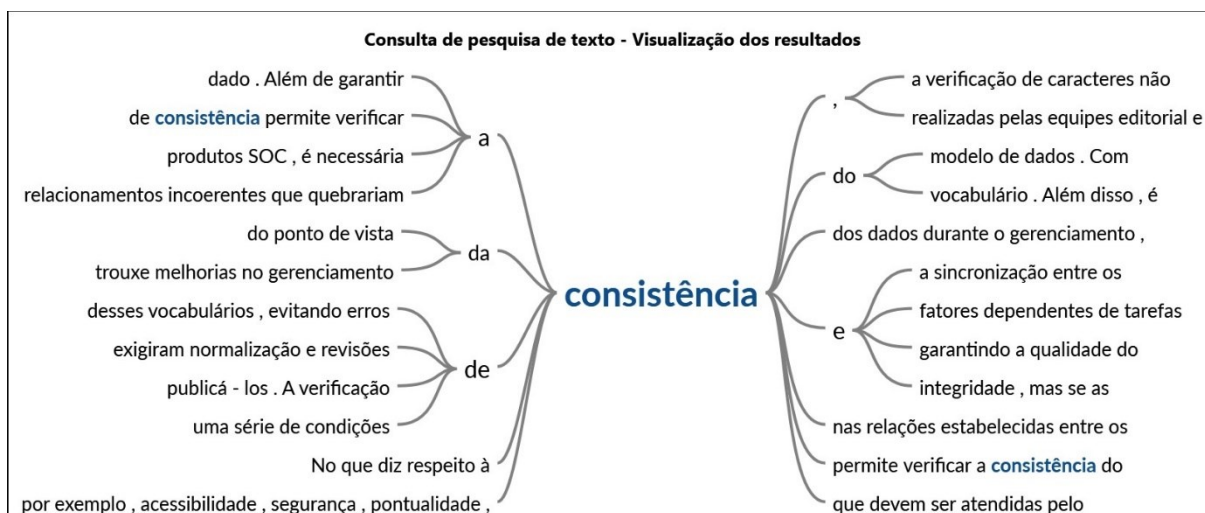
No tocante aos tesouros, a avaliação da qualidade dos dados é o princípio que assegura que o modelo de dados utilizado, bem como os relacionamentos entre os termos, foi estabelecido de maneira apropriada. Está relacionada à identificação e à correção de valores incoerentes, fragmentados ou incompletos dos vocabulários semânticos empregues. Os problemas mais comuns são os erros sintáticos, propriedades não declaradas ou o uso inapropriado de etiquetas lexicais (W3C, 2017). Contudo, é preciso esclarecer que essa avaliação pode ter alusão a duas categorias: a consistência dos dados e a integridade dos dados. Embora pareçam similares, há uma distinção bem clara entre elas.

Nos estudos analisados, tanto a consistência quanto a integridade dos dados foram os temas mais frequentes nos fragmentos de texto que receberam o código de “avaliação da qualidade”. Essa avaliação é mencionada nos estudos de (RAJBHANDARI; KEIZER, 2012; LARA, 2013; STELLATO et al., 2015; HARPRING, 2018; ZENG; MAYR, 2018; ZOU, 2018; DÍAZ-CORONA et al., 2019; FABEIRO; MASAGUÉ; ROVIRA JARQUE, 2019). A avaliação da integridade foi mencionada nos estudos de (PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012; PASTOR-SÁNCHEZ, 2013; MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2014, 2019b; ANIBALDI et al., 2015; ALBERTONI; DE MARTINO; PODESTÀ, 2018; DÍAZ-CORONA et al., 2019).

Para comprovar a qualidade dos dados é preciso executar uma verificação estrutural e semântica dos tesouros expressos e compartilhados como LOD. Do ponto de vista estrutural, ferramentas conhecidas como validadores são aplicadas para identificar violações nos esquemas de conceitos que possam comprometer a integridade e a consistência dos dados. De outra forma, problemas na estrutura dos *links* podem interferir nas pesquisas. Portanto, ferramentas de verificação devem ser operadas para gerenciar erros e garantir a qualidade dos tesouros (MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2014).

A árvore de palavras ilustrada na Figura 46 mostra que a consistência no contexto dos tesouros concerne ao nível de coerência dos relacionamentos entre os conceitos e a conformidade das representações com os modelos de dados utilizados. Verifica-se, ainda, que a avaliação da consistência está fortemente condicionada ao gerenciamento, no qual a responsabilidade sobre as revisões e as adequações recaem às equipes editoriais.

**Figura 46 -** Árvore de palavras da categoria consistência



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Sabe-se que para haver vinculações entre conceitos, isto é, organização, os dados dos vocabulários semânticos na *web* incorporam identificadores únicos e persistentes em seus registros de origem. A persistência desses identificadores também é requisito essencial de qualidade, visto que uma vez inalterados os recursos podem ser acessados de forma permanente. Do ponto de vista colaborativo, vários parceiros podem trabalhar em registros que contém informações para a mesma entidade (conceito). Por exemplo, uma coleção de conceitos está necessariamente relacionada a um esquema de conceitos se um dos profissionais adicionar ou suprimir termos nessa estrutura sem o devido cuidado. Desse modo, muito provavelmente a estrutura lógico-semântica do tesouro será comprometida, podendo resultar no aparecimento de termos órfãos ou subordinados indevidamente a uma classe.

Uma das características do SKOS é a possibilidade de que o mesmo conceito possa pertencer a um ou vários esquemas conceituais e até não pertencer a nenhum. Isso pode apresentar alguns problemas no gerenciamento de tesouros, quando é feito de maneira compartilhada por vários usuários. Para modelar nossa proposta, é necessário delimitar ou estabelecer certas restrições a esse respeito, a fim de garantir a integridade dos dados do sistema (PASTOR SÁNCHEZ, 2009, p. 166, tradução nossa).

Portanto, as contribuições devem resultar em registros unificados para cada entidade, apresentados claramente para os usuários finais. Quando os vocabulários são preparados para a publicação como LOD, essas vinculações exigem consolidação e normalização, limpando as inconsistências que surgem ao longo de períodos de manutenção manual. Assim como nas relações associativas dentro de um vocabulário, os vínculos entre vocabulários precisavam ser

lógicos e corretos. Assim, os registros devem permanecer consistentes para que outras aplicações possam reutilizá-los, sendo essa uma importante condição para assegurar a confiabilidade dos serviços que hospedam seu conteúdo.

Deve-se observar que, para produtos SOC, é necessária a consistência e a sincronização entre os bancos de dados originais e os repositórios RDF. Caso contrário, se a versão LOD de um SOC não for atualizada quando a fonte de dados original for atualizada, a qualidade desse produto se tornará questionável (ZENG; MAYR, 2018, p. 210, tradução nossa).

Logo, o gerenciamento dos níveis de qualidade deve ser contínuo e as atualizações das versões serem fornecidas periodicamente. Esses serviços de hospedagem funcionam como pontes de acesso unificado para o vocabulário e necessitam comprovar que os dados seguem padrões de qualidade. Algumas ferramentas já fornecem um serviço *online* para verificar a integridade e a consistência de tesouros em SKOS. Diante disso, podem ocorrer vários problemas relacionados às propriedades do modelo de dados aplicados, como conceitos não declarados ou o uso inapropriado de etiquetas lexicais. Algumas dessas falhas dá-se por erros sintáticos, enquanto outros parecem ser erros na compreensão do uso dos modelos pela equipe de construção do vocabulário.

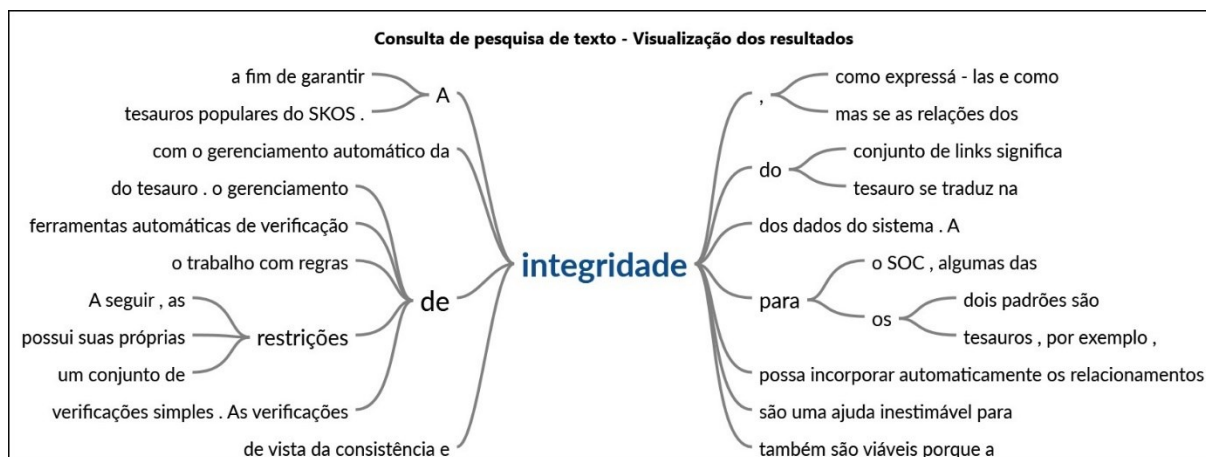
Esses aplicativos precisam gerenciar a **coerência semântica** dos relacionamentos criados no tesouro, impedindo automaticamente a introdução de possíveis relacionamentos incoerentes que quebrariam a **consistência do vocabulário**. Além disso, é desejável que uma ferramenta que lide com o gerenciamento automático da **integridade** possa incorporar automaticamente os relacionamentos que podem ser derivados daqueles já incluídos no tesouro. [...] As ferramentas automáticas de verificação de **integridade** são uma ajuda inestimável para a preparação e **manutenção** adequadas desses vocabulários, evitando erros de **consistência** e garantindo a qualidade do tesouro (MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2014, p. 6, tradução nossa).

As inconformidades são, muitas vezes, ocasionadas pelo mal uso de ferramentas, podendo serem corrigidas mediante procedimentos simples de aplicação automáticas, enquanto que outras requerem algum julgamento e conhecimento adicional das intenções da equipe de construção do vocabulário e ser sucedida por especialistas humanos. A automatização de correções incorporadas nos analisadores SKOS ou nas ferramentas de validação ajudam a evitar erros e tornam o reparo do vocabulário mais escalonável em relação à abordagem manual.

Concernente à categoria integridade, a árvore de palavras ilustrada na Figura 47 aponta que para inferir qualidade aos dados os aplicativos se baseiam em regras ou restrições de

integridade, que permitem realizar inferências automáticas sobre os dados do tesauro e derivar um conjunto de inconformidades que logo são encaminhadas para o devido reparo.

**Figura 47 -** Árvore de palavras da categoria integridade

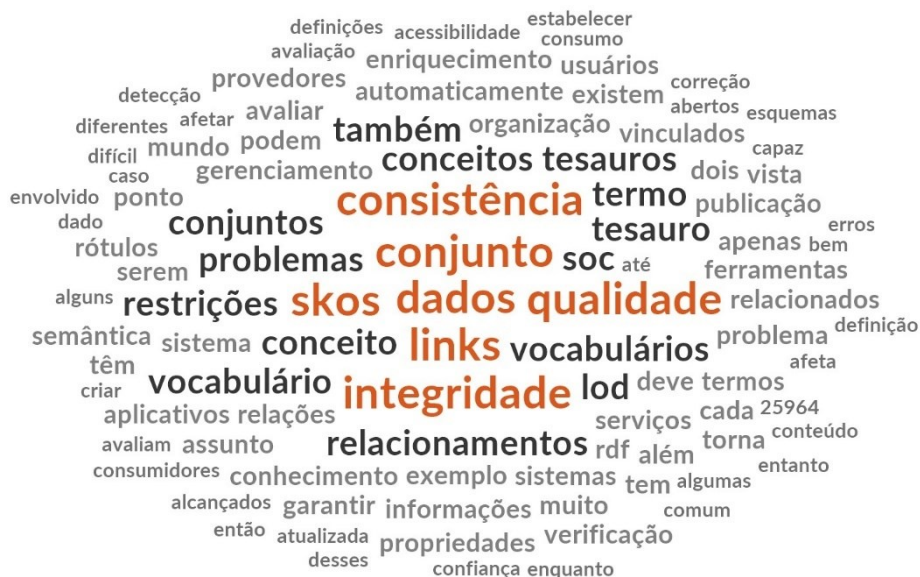


**Fonte:** Fornecido pelo *software* Nvivo (2020).

Muitos dos problemas elencados podem ser resolvidos com as ferramentas de validação. As ferramentas para o processamento de linguagem natural representam uma alternativa para sanar os problemas sintáticos, como o uso indevido de plurais ou grafias incorretas. Já os erros relativos à ausência de rótulos preferenciais, definições ou hierarquias incompletas, além dos validadores automáticos, os profissionais podem ajudar a identificá-los à medida que eles editam a estrutura de conceitos.

Relatórios de atividades também são uma boa opção, pois ajudam os gerentes a identificar os erros mais comuns, e podem ser compartilhados e discutidos com os demais membros da equipe de construção e evitar que se repitam no futuro.

**Figura 48** - Nuvem de palavras do princípio da Avaliação da Qualidade



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Os principais elementos que qualificam o princípio da avaliação da qualidade são apresentados na ilustração acima. Percebe-se o rigor que deve ser dado aos procedimentos básicos de avaliação, bem como a importância de desenvolver diretrizes ou ferramentas que assegurem a consistência e a integridade dos dados, mesmo nos cenários de colaboração, porque a participação de muitos atores sociais também pode interferir na qualidade do produto final.

#### **f) Acesso**

O princípio do acesso sugere que os projetos de publicação de tesauros como LOD ofereçam as condições necessárias para que os usuários possam encontrar e explorar seu conjunto de dados, os quais devem estar disponíveis de forma atualizada e proporcionar compreensibilidade tanto para humanos como para máquinas. Quanto mais compartilháveis e abertos, maior o valor potencial dos dados. Em vista disso, discussões quanto à necessidade de fornecer acesso aos dados foram observadas nos estudos de (QI; LUO; ZHU, 2010; CARACCILO et al., 2011; WETZ et al., 2012; LAPORTE; MOUGENOT; GARNIER, 2012; NEUBERT, 2012; AGUILAR, 2013; PASTOR-SÁNCHEZ, 2013; LAPORTE et al., 2014; WRIGHT; HARRISON; WATKINS, 2015; KEMPF; NEUBERT, 2016; ZENG; MAYR, 2018; FABEIRO; MASAGUÉ; ROVIRA JARQUE, 2019; MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019b), que versam a respeito das aplicações para o consumo humano e as ferramentas que tornam os dados processáveis por máquinas.

O uso de padrões, protocolos e modelos de dados é uma condição fundamental para tornar comum o acesso aos dados conectados. Os dados podem estar acessíveis de diversas formas: API ou um SPARQL Endpoint. No entanto, uma das premissas para a sua publicação como LOD é permitir que os usuários possam descarregar o conjunto de dados. A depender do volume, pode ser essencial fornecer acesso aos subconjuntos de dados, permitindo a recuperação ou o *download* de lotes menores.

As APIs são recomendadas para o acesso e a manipulação dos conceitos do tesouro, pois oferecem melhor capacidade de processamento de volumes de dados em tempo real. Em geral, as APIs para o acesso de dados dos tesouros na *web* são construídas utilizando o padrão REST (*Representation State Transfer*) – um padrão para arquitetura de *softwares* aplicado ao desenvolvimento de serviços na *web* – os quais assumem o encargo de consultar os termos mantidos nos sistemas simultaneamente com a navegação e a expansão do termo dentro de uma classe de conceitos de um tesouro. Caso este seja aplicado em conjunto com os sistemas de busca, permite que a pesquisa seja ampliada por recursos semânticos e de vocabulário baseados no modelo SKOS, por exemplo (QI; LUO; ZHU, 2010; CARACCILO et al., 2011).

Por conseguinte, buscadores semânticos também são requeridos. Eles devem seguir todas as recomendações do W3C relativas ao uso do protocolo SPARQL. O conteúdo interligado pode ajudar os usuários a expressar e usar de maneira mais rápida e melhor os tesouros. Para isso, é preciso dispor de um servidor de dados, representado por um ponto de extremidade SPARQL, que permita consultar e explorar bases de conhecimento em RDF. Segundo Caracciolo *et al.* (2011), a publicação de dados conectados oferece a vantagem de um único ponto de acesso usar linguagens de consulta padrão, sendo amplamente implementadas em aplicativos de computação.

A árvore de palavras lançada na Figura 49 exprime a relevância dos SPARQL Endpoint para o acesso aos dados do tesouro. Observa-se com a árvore que para estarem disponíveis na *web* os tesouros como LOD precisam oferecer pontos de acesso unificado e compartilhado, bem como mecanismos de pesquisa que possibilitem aos usuários navegar e explorar os conjuntos de dados. Nesse cenário, os tesouros passam de simples instrumentos de indexação a vocabulários de valor semântico capazes de representar notáveis bases de conhecimento.



**Figura 49 -** Árvore de palavras do princípio do acesso



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Um SPARQL *endpoint* combinado com uma resolução HTTP de suas entidades permite não apenas a publicação dos tesouros como dados conectados, assim como facilita a exploração e descarregamento para o uso em outras aplicações (PASTOR-SÁNCHEZ, 2013; MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019b). Apesar de representar uma condição necessária para o acesso aos dados no contexto do LOD, os pontos de extremidade SPARQL parecem não ser aplicáveis a usuários finais cujas tarefas não estão relacionadas às tecnologias semânticas. Dessa forma, Zeng e Mayr (2018) acreditam que os SOC na *web* podem ter variadas formas de utilização que vão além da função de um vocabulário controlado convencional.

Além de tornar disponíveis para a leitura de máquinas, o uso tradicional dos tesouros como instrumentos de indexação que auxiliam na recuperação da informação é essencial para que os projetos tornem os dados do tesouro disponíveis para a consulta de usuários que não têm familiaridade com as tecnologias da WS. Nesse sentido, quando habilitados para o ingresso por humanos devem estar disponíveis mediante interfaces de fácil utilização, com recursos de busca, navegação e visualização dos conceitos.

Para permitir que os usuários acessem as informações relevantes de cada conceito, era necessária uma interface do usuário acessível pela *web*. Essa interface precisava ser legível por humanos, exibindo claramente os rótulos dos conceitos, a definição e quaisquer relações com outros conceitos (internos

e externos ao tesouro) (WRIGHT; HARRISON; WATKINS, 2015, p. 108, tradução nossa).

No cenário dos tesouros multilíngues ou construídos como resultado da colaboração de muitos especialistas ou instituições, as plataformas capazes de agregar recursos compilados figuram como uma possibilidade de acessar sistematicamente os dados que residem em diferentes fontes (QI; LUO; ZHU, 2010). Nessa perspectiva, o acesso é também proporcionado pelos serviços de busca integrados que permitem explorar metadados associados a diversos recursos indexados pelos tesouros. Esses portais assumem a forma de agregadores de conteúdo, facilitando as buscas contextuais na rede e, geralmente, estão vinculados às coleções de dados de diferentes setores (bibliotecas, museus e arquivos nacionais).

Corroborando com a necessidade premente de fornecer acesso integrado aos recursos de tesouros, Zeng e Mayr (2018, p. 211) ressaltam a importância de portais que funcionem como ponto de conexão unificado para vocabulários SOC mantidos por diversas instituições, os quais podem fornecer a versão mais atual do vocabulário ou manter todas as versões. Funções adicionais podem estar disponíveis, além de pesquisar, navegar e exibir.

Assim, as plataformas desenvolvidas com esse propósito precisam dispor de interfaces dinâmicas para que o usuário final possa navegar e explorar o conjunto de dados do tesouro de maneira intuitiva. Os editores podem optar por disponibilizá-las em uma estrutura tradicional, publicadas via HTML Browser, com conceitos harmonizados hierarquicamente, ou, ainda, empregar um esquema de visualização gráfica na forma de árvore de conceitos. Esta última é comumente projetada como um desenho gráfico, permitindo uma maior interação dos usuários com as redes de conceitos interligados.

As ferramentas de visualização se mostram como recursos que ampliam as possibilidades de acesso e uso do tesouro como um instrumento de recuperação da informação. Nesse ponto, Fabeiro, Masagué e Rovira Jarque (2019, p. 20, tradução nossa) acreditam que a incorporação de uma visualização gráfica é capaz de conduzir o usuário “[...] a navegação entre os diferentes nós e explica os relacionamentos e *links* entre os termos de forma mais dinâmica e intuitiva”. Apesar dos pontos de extremidade SPARQL serem uma importante forma de acesso e busca semântica, eles per si não ajudam a cumprir a missão do tesouro de forma satisfatória, pois os tradicionais usuários demandam por ferramentas que eles conhecem e sabem usar.

Dessarte, os tesouros são estruturados como mapas de conhecimento que permitem aos usuários explorar campos semânticos e sua estrutura hierárquica é o principal aspecto a ser considerado quando se trata de visualização. Um tesouro que pode ser visto a partir de

relacionamentos entre os conceitos de domínio do conhecimento favorece a compreensão dos usuários (AGUILAR, 2013). Nesse sentido, Neubert (2012) acredita que a criação de uma interface terminológica comum para um domínio em específico permite que usuários especialistas explorem um conjunto de dados interligados sobre um único domínio.

Além de proporcionar o *download* ou a busca semântica dos dados, formas alternativas de apresentação oferecem oportunidades para acesso. O W3C recomenda que os aplicativos permitam aos usuários obterem uma visão mais concisa ou imediata dos dados, podendo estarem na forma de sumários analíticos ou tabelas que favoreçam-nos a compreender rapidamente o significado e incentivar ao reuso. A nuvem de palavras na Figura 50 possibilita identificar os principais temas em torno do princípio do acesso.

**Figura 50** - Nuvem de palavras do princípio do Acesso



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Observa-se que temas como pesquisa, navegação, consulta e recuperação estão entre os principais recursos de aplicações. Com isso, percebe-se que as questões de acesso aos dados do tesauro estão principalmente relacionadas às condições de leitura por humanos. Isso se deve à concepção tradicional do tesauro e sua função diante dos sistemas de recuperação.

Desse modo, fornecer condições de acesso é condição fundamental para que o tesauro possa ser reutilizado; proporciona que os usuários encaminhem *feedbacks* à equipe de construção. Assim, um componente de acesso em consonância a um componente de *feedback* contribui para que o usuário final avalie a usabilidade e utilidade percebida.

Além disso, oferecer formas de *feedback* pode ajudar a depreender como os usuários estão acessando e utilizando esses dados. Esta é uma funcionalidade que retorna para a equipe editorial informações de como esses atores sociais estão a explorar a rede de conceitos e se está atendendo as necessidades de busca e recuperação da informação. De posse desses *feedbacks*, a equipe pode aprimorar a linguagem de representação ou implementar novas formas de acesso.

### **g) Reuso**

O princípio do reuso diz respeito à capacidade dos projetos do LOD conceder metadados para tornar os dados compreensíveis por humanos e interoperáveis por computadores, aumentando as chances de reutilização. Compreende também os incentivos para gerenciar a publicação e atualização dos dados e acompanhar as demandas da comunidade que os consumirá.

Discussões sobre o reuso dos dados no contexto dos tesouros foram observadas em (ISAAC; MATTHEZING, 2007; ZHU; QIAO; SI, 2009; MORSHED et al., 2010; PASTOR-SÁNCHEZ; LÓPEZ-CARREÑO; MARTÍNEZ-MÉNDEZ, 2010; QI; LUO; ZHU, 2010; PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012; COYLE, 2012; LAPORTE; MOUGENOT; GARNIER, 2012; PASTOR-SÁNCHEZ, 2013; BACA; GILL, 2014; ANIBALDI et al., 2015; WRIGHT; HARRISON; WATKINS, 2015; COBB, 2016; ROSATI et al., 2017; VITI, 2018; ZENG; MAYR, 2018; DÍAZ-CORONA et al., 2019; FABEIRO; MASAGUÉ; ROVIRA JARQUE, 2019; MARTÍNEZ-GONZÁLEZ; ALVITE-DÍEZ, 2019b; MOCHÓN-BEZARES; MÉNDEZ-RODRÍGUEZ; LA FUENTE, 2020). Os estudos ressaltam principalmente a importância de utilizar modelos de dados interoperáveis, protocolos e padrões da WS, garantindo não apenas a interoperabilidade do tesouro, como também o aumento das possibilidades de reutilização. Portanto, os modelos de dados comuns e amplamente utilizados, assim como os diferentes formatos de serialização, expandem sua legibilidade por máquinas, fornecendo alternativas para o reuso econômico no futuro.

Uma vez expressos como dados legíveis por computador, estruturados e abertamente reutilizáveis, os sistemas de informação podem interpretá-los e usá-los para criar relacionamentos com outros conjuntos de dados conectados (COYLE, 2012; BACA; GILL, 2014). Considerados como ontologias leves, os tesouros têm se tornado vocabulários que ampliam a capacidade de recuperação da informação; constituem uma rica base de conhecimentos utilizados para indexar recursos ou documentos na *web*. À medida que são

estabelecidos os vínculos com outros vocabulários semânticos, eles têm o potencial de contribuir com a inferência e o processamento automático de dados.

A árvore de palavras na Figura 51 indica os contextos em que a palavra reutilização (palavra mais frequente nos fragmentos que receberam a codificação “reuso”) foi empregada nos estudos analisados. A reutilização de dados conectados está necessariamente relacionada à capacidade dos projetos descrever seus metadados e torná-los disponíveis livremente, facilitando a sua exploração por diversas comunidades do LOD.

**Figura 51 -** Árvore de palavras do princípio do reuso



**Fonte:** Fornecido pelo *software* Nvivo (2020).

A capacidade de importar e/ ou exportar tesauros representados em diferentes formatos é uma questão fundamental para alcançar a interoperabilidade. Como facilitar sua reutilização é um dos princípios do LOD, a escolha de formatos não proprietários para a leitura e troca de dados recai sobre o formato SKOS, sendo um padrão comum recomendado pelo W3C. O SKOS é amplamente aplicado pelos projetos e abrange todos os quesitos para a publicação de vocabulários controlados em LOD (ISAAC; MATTHEZING, 2007; LAPORTE; MOUGENOT; GARNIER, 2012; MOCHÓN-BEZARES; MÉNDEZ-RODRÍGUEZ; LA FUENTE, 2020).

Isto posto, é nesse cenário prático de reutilização que os dados conectados de tesouro marcam um passo importante para a organização da informação na *web* de forma colaborativa.

Uma vez interoperáveis, os dados podem ir além dos limites de unidades de informação especializadas e contemplar uma rede de recursos interconectados, saindo de uma esfera local para uma dimensão global de representação e uso (ANIBALDI et al., 2015).

De modo igual, Viti (2018) concorda que a conversão do tesauro em formatos de marcação adequados, como o SKOS, seguindo os padrões da WS e do LOD, melhora sua função como um “hub de interligação”. Os tesouros são capazes de facilitar a pesquisa e a organização de conteúdo por diferentes comunidades profissionais. Caso os dados estejam abertos e puderem ser descarregados em formatos como o RDF, eles poderão ser reutilizados por diversos projetos na *web*. Logo, à medida que os padrões são mundialmente aceitos e incorporados pelos projetos do LOD, reutilizar dados representa também um benefício econômico.

Atualmente, a padronização de alguns países desenvolvidos e grupos de pesquisa no exterior deve atrair mais atenção dos pesquisadores de vocabulário doméstico e pode ser considerada introduzida adequadamente para realizar o alto grau de reutilização do vocabulário e reduzir o custo do desenvolvimento de sistemas de *software* de processamento de informações (QI; LUO; ZHU, 2010, p. 526, tradução nossa).

Portanto, incentivar o reuso de dados não só favorece sua vinculação com recursos externos, como promove o desenvolvimento de aplicações que beneficiam o consumo compartilhado. Uma vez construídos para atender às necessidades de informação de um grupo específico, as ferramentas que oferecem o suporte prático para o acesso e consumo dos dados ajudam a incentivar sua reutilização também por diferentes setores. Nesse cenário destaca-se o desenvolvimento de API, que são serviços *web* voltados para os consumidores dos dados. Conforme as melhores práticas para a publicação e consumo de dados na *web*, as APIs devem fornecer uma documentação completa e de fácil compreensão, é quesito fundamental para sua qualidade e utilidade. Se a documentação não estiver clara, dificilmente um desenvolvedor utilizara-la.

Para Morshed *et al.* (2010), explorar a reutilização de dados conectados é uma metodologia cujas lições aprendidas podem ser aplicadas por outras pessoas com objetivos semelhantes. Daí a importância de colaborar com a abertura dos dados e o incremento de serviços *web*, como as API e os pontos de extremidade SPARQL. Esses recursos promovem não apenas a partilha, mas também a reutilização econômica e eficiente dos dados.

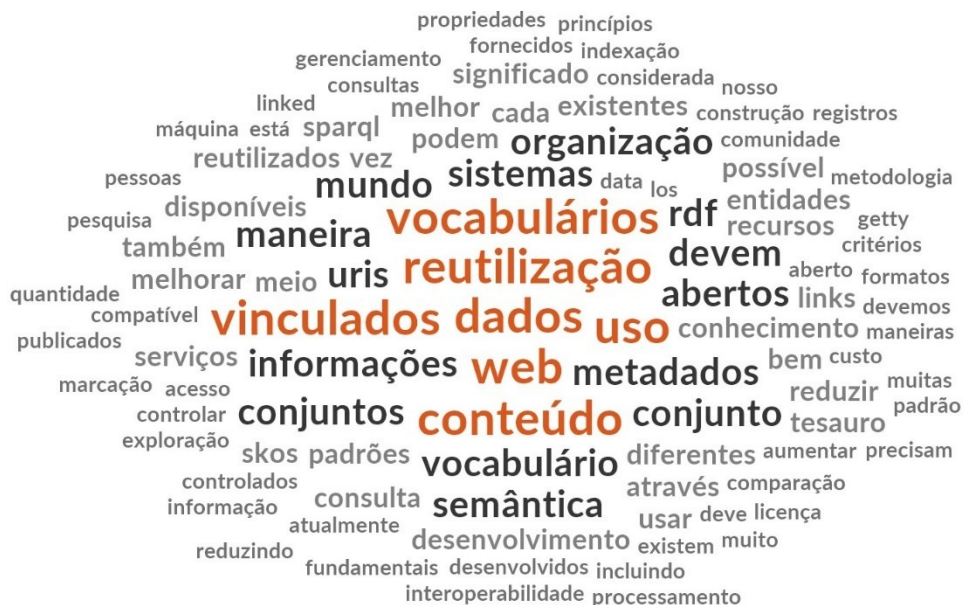
Para além de utilizar padrões de dados interoperáveis, assegurar sua livre reprodução na *web* passa pela necessidade de definir licenças de uso, as quais protegem legalmente os dados e controla seu uso, informando o que é permissivo e o que não é fazer com estes. Em se tratando

de tesouros abertos e tesouros conectados, a navegação e o acesso a todo o conjunto de dados devem ser disponibilizados gratuitamente, suportados por uma licença aberta, a fim de facilitar a reutilização ou adaptação mais ampla.

Em geral, os trabalhos analisados evidenciaram que a maioria dos tesouros fazem uso das licenças *Creative Commons* ou *Open Data Commons*. Apesar do significativo número de licenças padrão já existentes, os projetos podem disponibilizar dados fazendo uso de licenças próprias, desde que as especificações estejam devidamente claras para o usuário. As informações quanto à licença devem ser descritas empregando vocabulários específicos, como o Dublin Core, schema.org ou XHTML, para a facilitar a compressão de humanos e máquinas (PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012).

A nuvem de palavras na Figura 52 sinaliza as palavras mais frequentes em torno do princípio do reuso. Observa-se que não há como incentivar o reuso dos dados sem uma adoção mínima de metadados, padrões e protocolos. Entre eles destacam-se o modelo de dados SKOS, a linguagem RDF e o protocolo SPARQL.

**Figura 52** - Nuvem de palavras do princípio do reuso



**Fonte:** Fornecido pelo *software* Nvivo (2020).

Todavia, os projetos de tesouros como LOD precisam conhecer o público e as demandas de informação para que haja fomento ao consumo. Para Fabeiro, Masagué e Rovira Jarques (2019), conforme os tesouros são disponibilizados livremente na *web* pode-se aumentar sua relevância e observar tendências de uso dos dados. Construir a comunidade em torno dos

dados inclui também assegurar sua qualidade. Assim, é preciso possibilitar aos consumidores de dados o envio de *feedbacks* à equipe de construção. A partir deles é possível desenvolver ações para corrigir possíveis inconsistências e assegurar que a qualidade dos dados alcance o melhor nível.

Por mais que seja essencial a participação dos usuários finais, os conceitos compilados mediante ferramentas de *feedback* devem ser verificados pelos editores para evitar duplicação e garantir a conformidade. Assim sendo, Wright, Harrison e Watkins (2015) apontam o método denominado painel de especialistas do domínio como responsáveis pela decisão ou adequação de um conceito candidato à inclusão no tesouro, podendo o método fazer alterações necessárias nos rótulos, nos relacionamentos e na definição dos termos sugeridos.

De outra forma, recursos de interação podem ser utilizados como expansão interativa de consultas ou anotações sociais com vistas a verificar sua adequação à linguagem do usuário. O contato com colaboradores externos fomenta continuada manutenção, visto serem capazes de dar respostas significativas ao sistema, apontando debilidades e sugerindo alterações. Uma vez que o tesouro é disponibilizado ao público é essencial um mecanismo para permitir a adição ou a sugestão de novos conceitos, tanto pelos gerentes como pelos usuários finais. Partindo do pressuposto que os tesouros na *web* são instrumentos derivados de uma cocriação, as ferramentas aplicadas devem funcionar como pontos de interação, de modo a favorecer a comunicação entre os usuários e a equipe de gestão do tesouro.

#### **4.2.2 Considerações gerais sobre os princípios evidenciados na literatura**

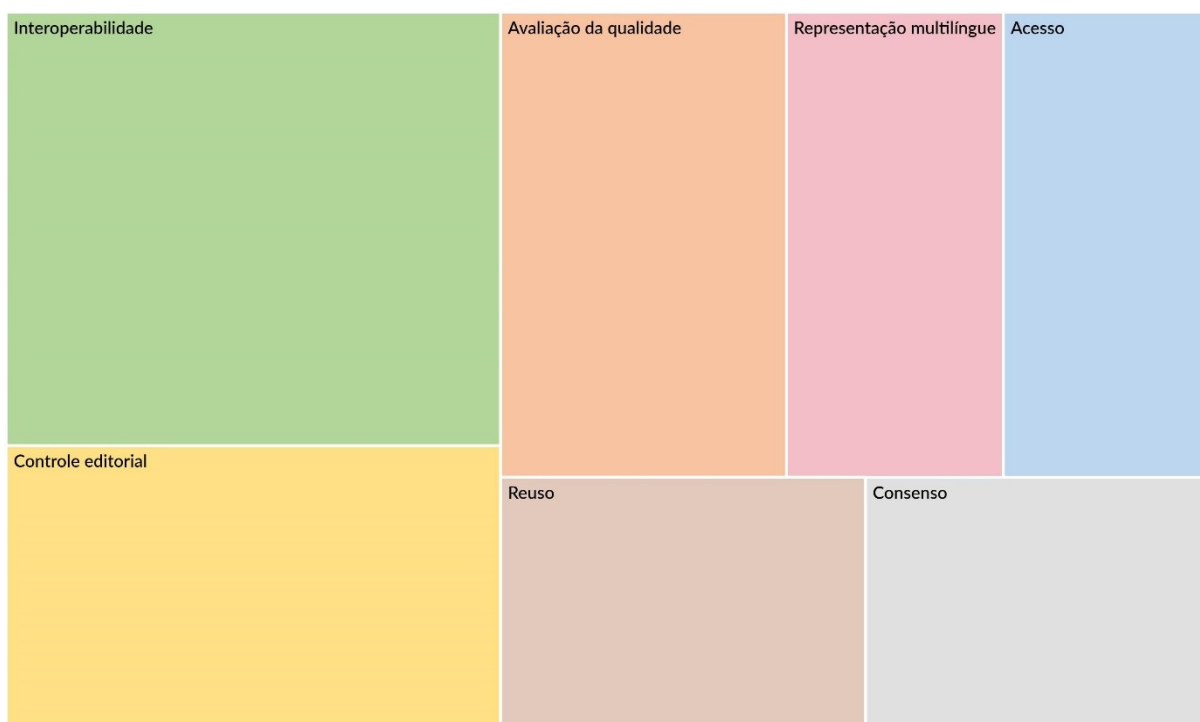
A análise da literatura permitiu que alguns padrões fossem identificados. Com a estruturação foi possível identificar as etapas de construção em que os princípios devem ser aplicados, as dimensões de colaboração e o papel dos atores sociais em cada procedimento. Portanto, os projetos de tesouros desenvolvidos no âmbito do LOD devem selecionar os recursos tecnológicos adequados em função desses princípios e das atividades de construção necessárias.

Após a codificação e sistematização dos dados percebeu-se o destaque dado ao princípio da interoperabilidade nos estudos analisados. A Figura 53 ressalta que a maioria das pesquisas apresentam referências às questões de interoperabilidade, como o uso de padrões e modelos de dados, tecnologias, mapeamento e interligação com outros conjuntos de dados. Isso se deve à necessidade de incorporar tecnologias da WS na construção de tesouros.



Além de promover a representação e a indexação unívoca de recursos na *web*, os tesouros têm se tornado um importante hub de ligação capaz de vincular recursos de diferentes fontes de dados. O impacto desse princípio nos estudos analisados pode estar relacionado à interdependência de alguns fatores, como a necessidade de operar modelos de dados e padrões para sua construção, sendo este também condição fundamental para inferir qualidade ou para torná-los acessíveis e próprios para o consumo e a reutilização.

**Figura 53** - Gráfico hierárquico dos princípios mais citados entre os estudos analisados



**Fonte:** Fornecido pelo *software* Nvivo (2020).

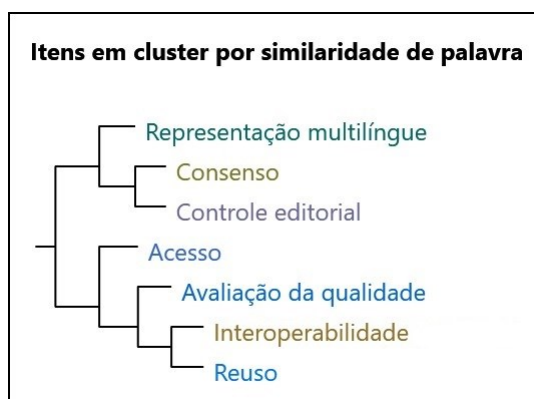
Concebeu-se, ainda, a relevância do princípio do controle editorial e da avaliação da qualidade. O controle editorial garante que o trabalho de construção possa ocorrer por via de práticas colaborativas sem trazer prejuízo para a qualidade do produto final e as atividades sejam distribuídas conforme os papéis de atuação de cada profissional envolvido. Estes formam os grupos de gestão do tesouro e têm como único objetivo representar o domínio do conhecimento escolhido. São guiados, *a priori*, por normativas e diretrizes estabelecidas que devem ser compartilhadas entre os especialistas envolvidos, de maneira que limites de decisão sejam conhecidos por todos para que o mínimo de erros possível possa ocorrer.

A avaliação da qualidade também apresentou um impacto relevante entre os estudos analisados. Esse princípio garante que os níveis de consistência e integridade do tesouro sejam

mantidos, posto que se o conjunto de dados mostra inconformidades em relação aos padrões de dados, muito provavelmente não será reutilizado por outras aplicações na *web*. Assim, garantir a qualidade dos dados é fundamental, pois do contrário o tesouro não cumprirá com sua função.

A Figura 54 representa a análise de cluster dos códigos. O diagrama agrupa os códigos selecionados com características similares. O cálculo é realizado com base na similaridade entre as palavras do *corpus* de análise e os resultados exibidos como um dendograma. Assim, foi possível qualificar a formação de dois principais agrupamentos que, segundo a disposição dos dados no *cluster*, podem ser interpretados como: o primeiro agrupamento de princípios orientados à construção de tesouros; e o segundo agrupamento de princípios orientados à publicação de tesouros.

**Figura 54** - Cluster por similaridade de palavras



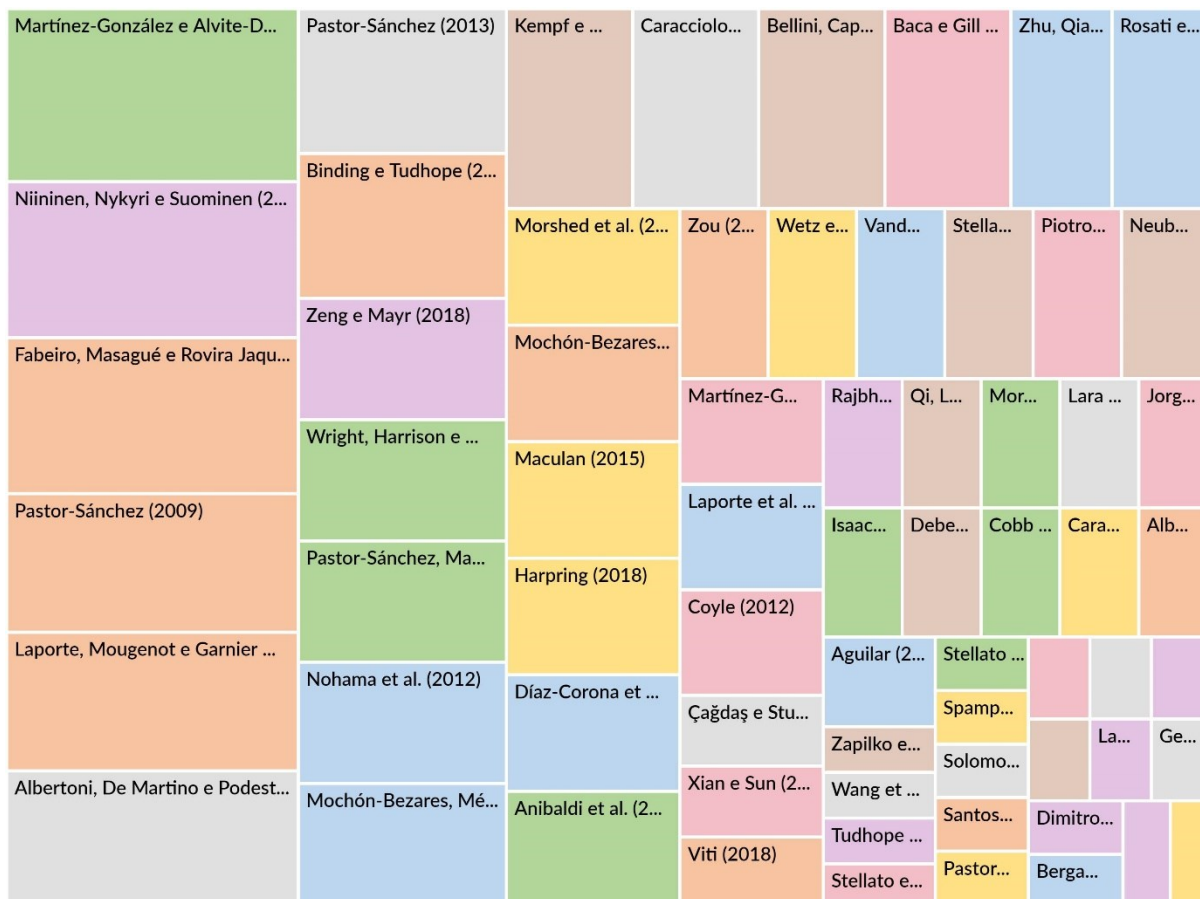
**Fonte:** Fornecido pelo *software* Nvivo (2020).

O primeiro agrupamento está representado pelo seguinte conjunto de princípios: representação multilíngue, consenso e controle editorial. Podem estar relacionados entre si, visto constituírem quesitos fundamentais aplicados na fase inicial de construção de tesouros; envolvem as etapas de planejamento, compilação e seleção de termos, inclusive em outros idiomas. Estas etapas são estabelecidas nas diretrizes e orientam os profissionais a executarem os procedimentos de forma padronizada.

Quanto ao segundo agrupamento, tem como conjunto de princípios: acesso, avaliação da qualidade, interoperabilidade e reuso. Estes princípios estão associados aos requisitos para a publicação dos tesouros no contexto do LOD; orientam o uso de padrões, protocolos, modelos de dados e ferramentas de aplicação que favorecem o consumo e a reutilização dos dados na *web*.

Considera-se importante destacar os estudos que apresentaram o maior número de princípios codificados. A Figura 55 ressalta os trabalhos de Martínez-González e Alvite-Díez (2019), Niininen, Nykyri e Suominen (2017) e Fabeiro, Masagué e Rovira Jaque (2019), Niininen, Nykyri e Suominen (2017) e Fabeiro, Masagué e Rovira Jaque (2019).

**Figura 55** - Gráfico hierárquico dos estudos que contemplam o maior número de princípios



**Fonte:** Fornecido pelo *software* Nvivo (2020).

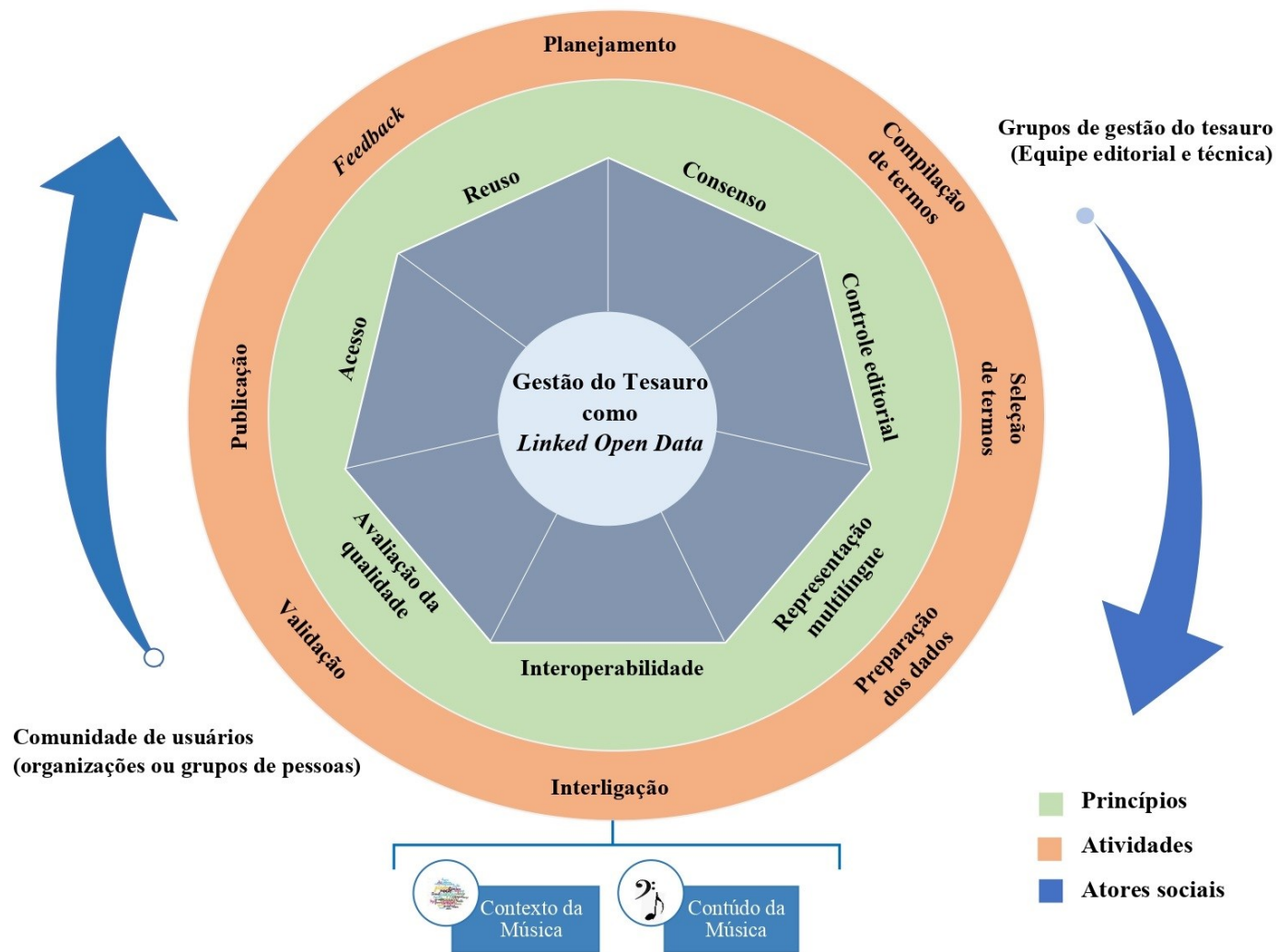
Na próxima seção apresentar-se-á a proposta do modelo desenvolvido, no qual utilizou-se dos princípios estruturados com base na literatura e nos relatos coletados com as instituições que participaram desta pesquisa.

## **5 MODELO COLABORATIVO PARA A CONSTRUÇÃO E A PUBLICAÇÃO DE TESAuros DO DOMÍNIO DA MÚSICA COMO *LINKED OPEN DATA***

Dada a natureza orgânica da construção e publicação de tesauros na *web*, os princípios para o seu gerenciamento são conduzidos por um conjunto de ações que levam os principais atores sociais envolvidos a cooperar sistematicamente. A proposta está fundamentada na concepção de que o processo colaborativo é resultado das atividades de interação entre os grupos desses atores, distribuídos em três dimensões: comunidade de usuários, grupos de gestão do vocabulário e a comunidade do LOD do domínio da Música.

Isto posto, os princípios evidenciados na literatura constituem o “núcleo” do sistema colaborativo, e a partir dele oito atividades essenciais são especificadas, quais sejam: planejamento, compilação de termos, seleção de termos, preparação dos dados para a publicação, interligação, validação, publicação e *feedback*. As dimensões colaborativas, os princípios fundamentais e as atividades essenciais são retratadas na Figura 56.

**Figura 56** - Modelo colaborativo para a construção e publicação de tesouros do domínio da Música como *Linked Open Data*



Fonte: Elaborado pelo autor (2021).

Na dimensão da comunidade de usuários, os tesouros são (re) utilizados por organizações ou por grupos de pessoas. Estes atores sociais contribuem com a construção do tesouro, mas não entre si, e sim à medida que: a) as organizações disponibilizam fontes de referência ou léxicos sobre o domínio; b) os usuários exploram a rede de conceitos e oferecem *feedbacks* à equipe de construção. Esta dimensão atua sob os princípios do acesso e do reuso, os quais, mediante atividades de publicação, permitem que os dados dos tesouros estejam prontos para o consumo e interligados com outros sistemas de conceitos, formando uma rede semântica que é empregada e retroalimentada pela comunidade.

A dimensão dos grupos de gestão do tesouro é concebida como um espaço particular de interação. Nele, o tesouro é estruturado por meio da colaboração e da negociação entre os membros da equipe editorial e da equipe técnica. A primeira equipe é constituída de um grupo de indivíduos que executam as atividades de edição do tesouro. São compostos por profissionais da informação, especialmente bibliotecários e editores de ontologias. Já a segunda é formada por indivíduos (especialistas do domínio, terminólogos e profissionais de área de tecnologia de informação) que auxiliam nas atividades de edição. Assim, para que sejam cumpridas as atividades de edição, as equipes devem dispor de ferramentas para apoiar o processo colaborativo de construção e manutenção do tesouro, bem como atuar sob os princípios do consenso, da representação multilíngue, do controle editorial e da avaliação da qualidade.

A dimensão da comunidade do LOD do domínio da Música oferece o suporte básico para que os vínculos com outros tesouros e com conjuntos de dados diretamente relacionados ao domínio sejam estabelecidos – é preciso considerar o mapeamento com conjuntos de dados externos para expandir as representações conceituais e enriquecer semanticamente o tesouro – interage com a dimensão dos grupos de gestão mediante atividade de interligação e atua sob a regência dos princípios da representação multilíngue e da interoperabilidade.

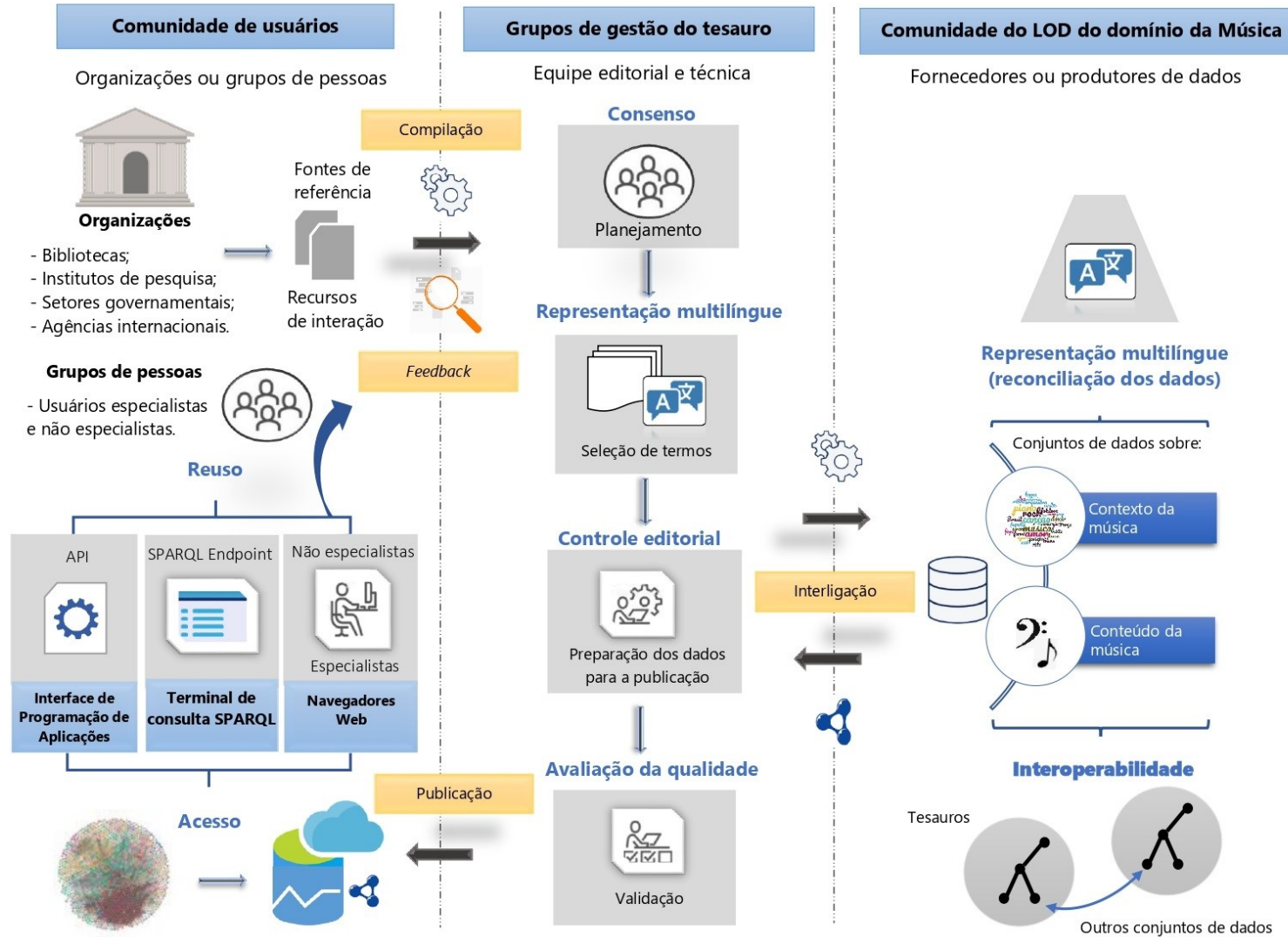
As oito atividades foram inspiradas nas atividades de planejamento, levantamento de termos candidatos, interoperabilidade e mapeamentos entre vocabulários arroladas nas normas para a construção de tesouros ISO 25964-1/2 (2011; 2013), bem como no guia de melhores práticas recomendadas pelo W3C (2017) para a publicação de dados como *Linked Data*. As atividades são desempenhadas neste modelo da seguinte forma:

- a) **Planejamento:** define objetivos, escopo e o público-alvo do tesouro. São levantadas questões sobre os recursos humanos e tecnológicos necessários para a sua construção;
- b) **Compilação de termos:** descreve os procedimentos automáticos e semiautomáticos para a compilação de termos candidatos à inclusão no tesouro;

- c) **Seleção de termos:** orienta as equipes de construção quanto à análise dos termos compilados, à construção de categorias e à definição dos critérios para a especificação dos termos preferenciais;
- d) **Preparação dos dados para a publicação:** estabelece diretrizes para realizar a modelagem e o mapeamento dos dados. Fornece orientações concernente à definição de padrões para URI, conversão e o gerenciamento dos dados em RDF;
- e) **Interligação:** compreende as etapas de mapeamento e seleção de rótulos semânticos correspondentes em conjuntos de dados externos relacionados ao domínio da Música;
- f) **Validação:** detecção e correção das inconsistências estruturais, conforme as regras de restrição do modelo de dados selecionado para a representação do tesouro;
- g) **Publicação:** compreende a descrição dos conjuntos de dados de maneira que humanos e máquinas possam acessá-los e reutilizá-los. Inclui a seleção de tecnologias para promover a busca, a exploração e a visualização dos dados;
- h) **Feedback:** disponibiliza mecanismos para que os usuários comuniquem suas necessidades de uso e os níveis de satisfação em relação ao consumo de dados.

Essas atividades são desenvolvidas em um fluxo contínuo de tarefas e cada dimensão é regida por um conjunto de princípios. A Figura 57 ilustra o diagrama geral do modelo, no qual é possível visualizar como as atividades e os princípios estão distribuídas nas três dimensões de colaboração.

Figura 57 - Diagrama geral do modelo colaborativo para a construção e a publicação de tesouros do domínio da Música como *Linked Open Data*

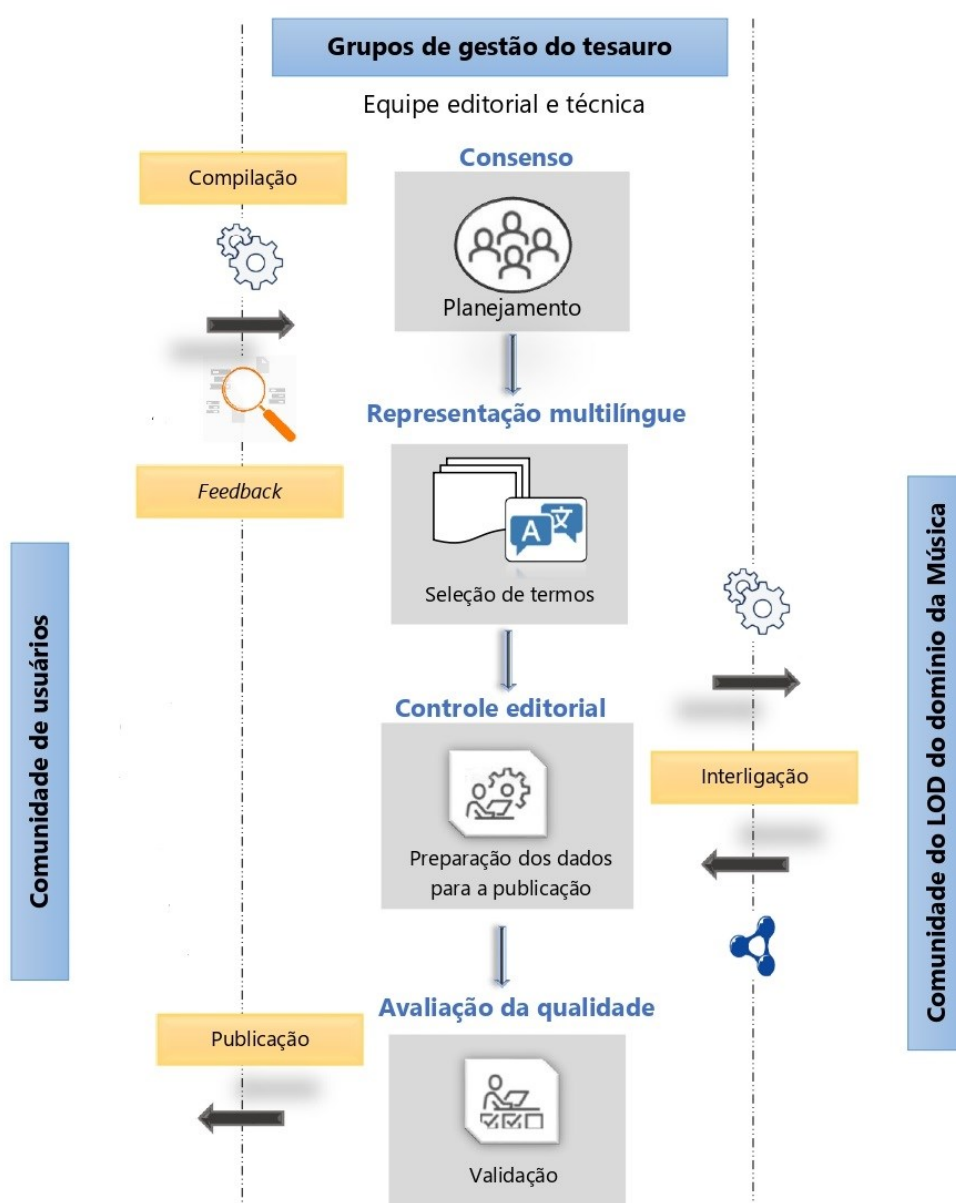


Fonte: Elaborado pelo autor (2021).



De posse dos termos candidatos, coletados a partir da atividade de aquisição, a equipe editorial, mediante ferramentas colaborativas de edição, dá início à seleção dos termos, à preparação dos dados para a publicação e a validação. A Figura 58 explicita como sucede a colaboração na dimensão dos grupos de gestão. As ferramentas de edição colaborativa têm a função de integrar as múltiplas contribuições de todos os membros da equipe. Seja qual for a plataforma escolhida para a manipulação do tesouro, ela deve oportunizar a execução dos princípios do consenso, do controle editorial e da avaliação da qualidade.

**Figura 58** – Dimensão Grupos de gestão do tesouro



**Fonte:** Elaborado pelo autor (2021).

É importante esclarecer que esta é uma dimensão técnica de atuação interprofissional, na qual a equipe editorial é formada por profissionais multidisciplinares (terminólogos, especialistas de domínio, profissionais da informação, especialistas em tecnologia da informação e pesquisadores). Os grupos de gestão interagem com a dimensão da comunidade de usuários através de dois canais de entrada (aquisição e *feedback*) e de saída (publicação). A equipe editorial recebe o *feedback* da comunidade de usuários mediante recursos de interação. No final do ciclo, um componente de avaliação autoriza que especialistas validem e reparem os dados identificados como inconsistentes, sendo este um requisito fundamental para a garantia da qualidade dos dados.

A colaboração na dimensão dos grupos de gestão do tesouro é caracterizada como formal e direta, cujas etapas de construção precisam ser guiadas por diretrizes estabelecidas, *a priori*, na etapa de planejamento. Essas etapas estão condicionadas fortemente ao fluxo de tarefas, compartilhamento de informações e a comunicação entre os membros da equipe editorial. Como envolve profissionais de múltiplas áreas do conhecimento, é essencial o diálogo permanente entre os membros da equipe para garantir a consonância de procedimentos.

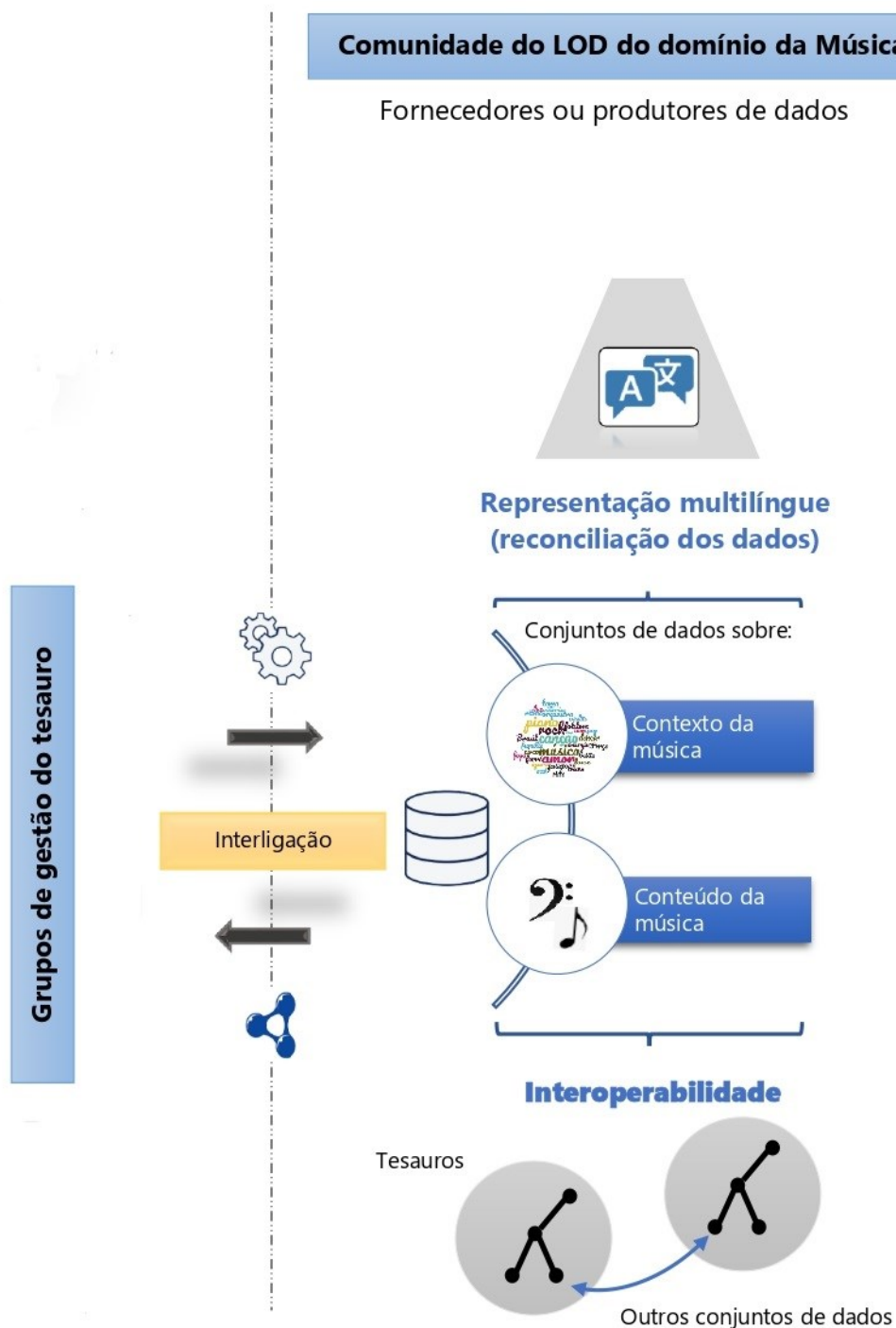
A comunidade do LOD do domínio da Música é demonstrada na Figura 59. A interação com a dimensão dos grupos de gestão ocorre mediante atividade de interligação, responsável pelo mapeamento de rótulos (ou termos que representam os conceitos) correspondentes em vocabulários externos, ampliando a vinculação com as demais entidades desse domínio.

As correspondências nos níveis lexicais e conceituais devem ser consideradas. Sob o ponto de vista do nível do idioma, estas são fundamentais para garantir que os tesouros superem as barreiras culturais e alcancem o nível mínimo de representação multilíngue. Logo, atender ao princípio da representação multilíngue em sua totalidade pode ser desafiador, pois exige uma análise minuciosa do idioma de destino em comparação com os termos do tesouro no seu idioma de origem. Assim, é preciso esclarecer que as traduções requerem negociação de elementos culturais.

Isto posto, decidir sobre quais idiomas os conceitos serão mapeados é uma responsabilidade da equipe de gestão, que deve observar todas as garantias: de uso, usuário, literária, cultural, estrutural e organizacional. É preciso identificar quais os tesouros do domínio da Música em outros idiomas são, de fato, relevantes para ampliar essa rede de termos multilíngues. Desse modo, a equipe de gestão necessita não apenas ter o conhecimento no idioma de destino, mas, também, a experiência com o domínio da Música para avaliar a

persistência e inclusão dos conceitos correspondentes. Outrossim, será fundamental introduzir na equipe um especialista em traduções.

**Figura 59** – Dimensão da Comunidade do LOD do domínio da Música



**Fonte:** Elaborado pelo autor (2021).

Na dimensão da comunidade de usuários, ilustrada na Figura 60, organizações e grupos de pessoas se apresentam como os principais atores sociais contribuindo com a construção do tesouro à medida que: a) as entidades disponibilizam publicamente seus recursos lexicais; b) a comunidade de usuários ou especialistas compartilham suas experiências por meio de fóruns públicos de discussão; c) usuários executam pesquisas e exploram os dados nas interfaces de busca. A abordagem considera a atividade de aquisição como o principal componente de entrada do modelo. Os termos candidatos são coletados de fontes de referências ou léxicos do domínio (tesauros, dicionários, enciclopédias eletrônicas, um *corpus* representativo da coleção da instituição/organização ou serviços documentais) e recursos de interação (fóruns de discussão e *logs* de transações).

Os usuários ou consumidores de dados podem acessar os dados do tesouro através de três recursos: Interfaces de programação de aplicações (API), terminais de consulta SPARQL e Navegadores *web*. O primeiro é um sistema que fornece um conjunto de rotinas e padrões para que outros aplicativos na *web* possam utilizar os serviços de dados. O segundo é um terminal HTTP para realizar consultas aos dados na linguagem RDF. E o terceiro intenciona explorar e visualizar dados na linguagem HTML. Os navegadores *web* devem disponibilizar motores de busca para que seus utilizadores realizem as buscas a partir de conteúdo, geralmente por meio de palavras-chave.

A colaboração na dimensão da comunidade de usuários é majoritariamente não formal ou indireta, porque as contribuições advindas dos atores sociais não estão condicionadas a um chamamento obrigatório, mas sim à participação espontânea da comunidade de usuários do sistema. Nesse cenário, o princípio do acesso e do reuso norteiam as ações que visam o compartilhamento e a cooperação permanente do público com o tesouro. É importante enfatizar que o LOD é um movimento aberto e comunitário e, portanto, todas as iniciativas para integrar seus diferentes públicos, especialistas ou não, e inclui-los na construção de tesauros trará benefícios não somente para tornar a construção da linguagem mais ágil, assim como para otimizar a recuperação da informação musical na *web*.

**Figura 60** – Dimensão da Comunidade de Usuários



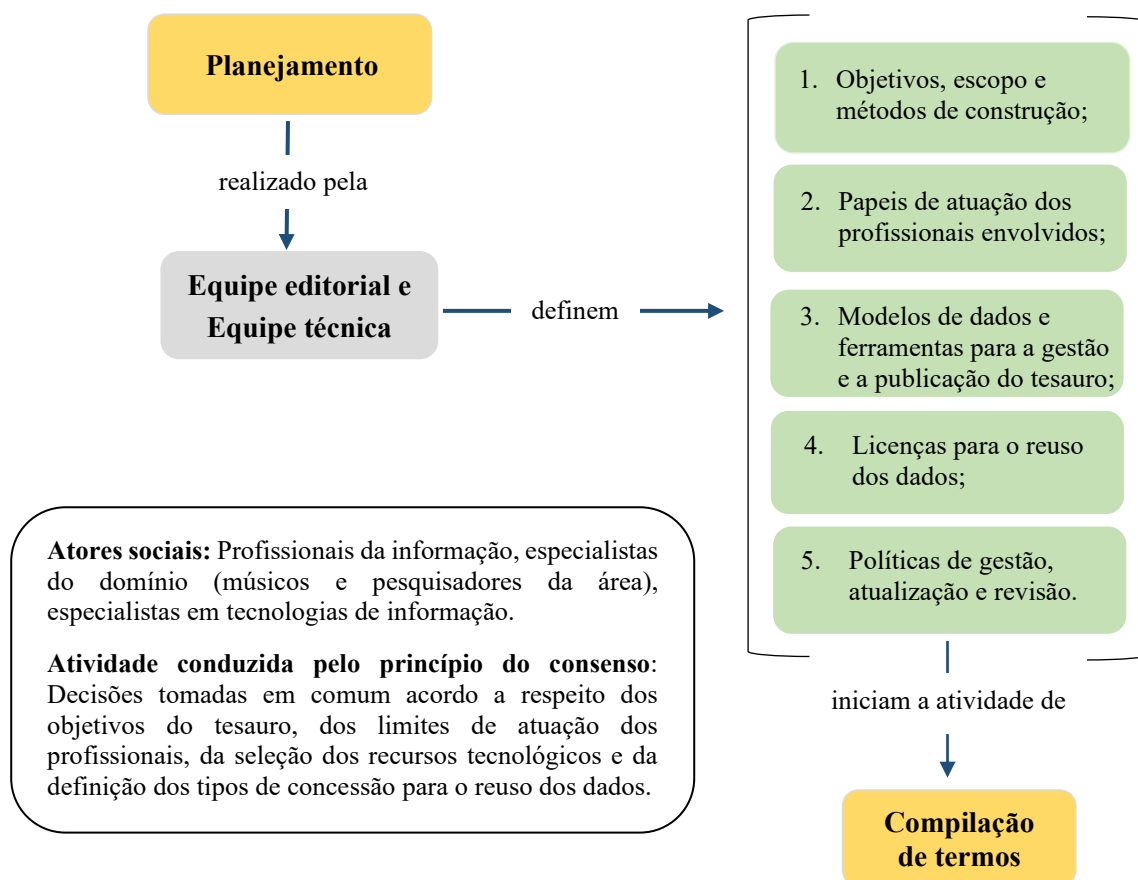
Fonte: Elaborado pelo autor (2021).

Nas próximas seções serão especificadas as atividades do modelo proposto.

## 5.1 PLANEJAMENTO

Antes de implementá-lo, os gerentes de projeto precisam estabelecer diretrizes sobre os procedimentos para a elaboração do tesauro. Isto se deve em razão da necessidade de controlar as tarefas e melhorar a produtividade do trabalho. Diante disso, o planejamento funciona como uma orientação para que as equipes de construção também conheçam os limites de atuação. Esta atividade se desenvolve conforme as etapas explanadas na Figura 61.

**Figura 61** – Etapas da atividade de Planejamento



**Fonte:** Elaborado pelo autor (2021).

Recomenda-se que os grupos de gestão do tesauro sejam formados por diversos profissionais, de modo que estes colaborem sob a supervisão de pelo menos um gerente de projeto. Assim, as decisões devem ser tomadas em conjunto, seguindo o princípio do consenso, para definir uma série de questões discutidas na sequência, como: os objetivos do tesauro, público-alvo, os papeis de atuação de cada profissional envolvido, os requisitos para a seleção de ferramentas de gestão do tesauro, as licenças abertas para o reuso dos dados e as políticas de gestão, atualização e revisão.

### 5.1.1 Definição dos objetivos, escopo e métodos de construção do tesouro

A construção de tesouros envolve a participação de profissionais com formação acadêmica ou técnica em diversos níveis de especialidade. A norma ISO:2594-1 (2011) estabelece que os diferentes profissionais envolvidos podem ter visões diferentes sobre a necessidade de estruturação do tesouro e recomenda que antes de iniciar sua elaboração a equipe editorial precisa definir objetivos, observando, principalmente, as características e necessidades da comunidade que irá utilizá-lo.

A abordagem de aplicação do modelo se concentra no domínio da Música, porém seja qual for a área de especialidade o desafio de elaboração pode estar relacionado à necessidade de o tesouro atender às demandas de um público múltiplo. No caso específico do domínio da Música, a comunidade de usuários é formada por músicos, docentes e pesquisadores, ou pelo público em geral, que buscam a informação musical para diferentes fins, sejam eles, acadêmicos e profissionais, sejam eles para o lazer.

Assim, no sentido de auxiliar às equipes de construção a compreender seu público-alvo e definir os objetivos do tesouro, alguns questionamentos são postos e precisam ser respondidos:

**a) Qual será o público-alvo?** É preciso realizar um levantamento das principais características do usuário e suas necessidades de informação. No domínio da Música, usuários especialistas e não especialistas podem lidar com o mesmo sistema de busca para recuperar a informação musical. São considerados usuários do tesouro: músicos, docentes, pesquisadores, discentes de escolas de Música com ênfase na teoria musical e na prática instrumental. Assim, precisam ser tomadas decisões sobre o nível de especificação dos termos;

**b) Qual será o nível de especialidade do tesouro?** É importante ter em mente que o tesouro pode não atender a todos da mesma maneira, pois os termos que compõem sua estrutura são definidos conforme o nível de especialidade do público-alvo. Portanto, alguns acordos no nível estrutural devem ser mantidos, como a inclusão de termos que não se apoiam totalmente na garantia do usuário ou literária, mas que são admitidos no vocabulário porque apresentam uma função basilar útil. Nessa metodologia, duas categorias de usuários poderão ser estimadas: usuários especialistas e não especialistas. Como se trata de um público diverso, a atividade de representação dos conceitos precisa se atentar, em alguns casos, aos níveis mais gerais de descrição e, igualmente, contemplar termos mais específicos e técnicos. Por isso, pode haver a necessidade de compatibilização da linguagem ou adaptação de termos.

**c) Por meio de quais fontes de referência os termos candidatos serão coletados?**

É preciso definir onde e como os termos que irão compor os tesouros serão recolhidos. O perfil do usuário é decisivo na escolha das fontes de referência. Os termos candidatos podem ser recolhidos de tesouros existentes, dicionários especializados, índice de publicações periódicas ou revistas técnicas. Recomenda-se empregar não apenas léxicos para compilar os termos candidatos, mas também os recursos de interação. Pressupõe-se que as expressões livres de busca são fontes de referência fundamentais para executar a garantia do usuário. Assim, a compatibilização da linguagem poderá ocorrer a partir da análise do comportamento de busca tanto de usuários especialistas como de usuários não especialistas.

É importante destacar que no contexto dos vocabulários já existentes pode haver conjuntos de dados de tesouros disponíveis para reutilização, podendo estes serem reaproveitados, desde que assumam os formatos exigidos pelo conjunto de melhores práticas do *Linked Data*. Assim, os gerentes de projetos devem selecionar os tesouros existentes em LOD, observando, principalmente, os níveis de qualidade dos dados e da terminologia desenvolvidas a fim de verificar se estes vocabulários abertos estão ajustados às necessidades do público-alvo.

**d) Quais serão os métodos ou as abordagens de construção utilizados?** Nessa fase do planejamento, é importante definir os métodos e as regras a serem adotados na construção. Estes influenciam diretamente na compilação, seleção e categorização dos termos que representam os conceitos do domínio. Logo, há diversas abordagens teóricas e metodológicas que podem influenciar na construção de tesouros. A ANI/NISO Z39.19-2005 (R2010) categoriza os métodos de construção em quatro tipos de abordagens: abordagens de Comitê; abordagem empírica; combinação de métodos; e assistência de máquinas.

Sugere-se a combinação de métodos dedutivos e indutivos para a compilação e seleção de termos. Estes métodos consideram tanto a literatura como o endosso do usuário, ou a experiência de especialistas do domínio para coletar termos que verdadeiramente representam os conceitos do domínio. Desse modo, a combinação dos dois métodos ainda contribui para estruturar o tesouro e estabelecer as relações lógico-semânticas entre os termos que irão compô-lo.

Assim, visto tornar a construção mais ágil, a equipe de gestão deve considerar todas as opções metodológicas disponíveis, conferindo rigor às etapas de compilação, seleção e especificação de termos. Destaca-se que os métodos semiautomáticos podem tornar a disposição menos onerosa e mais rápida, visto que usa máquinas para coletar termos representativos em amostras da coleção, assim como considera a participação de especialistas



humanos que analisam os termos compilados e executam as etapas de seleção e especificação dos termos do tesouro.

### 5.1.2 Papéis de atuação dos profissionais envolvidos

Os acordos produzidos durante a etapa de planejamento são documentados e comunicados a todos os membros da equipe. Cada profissional assume um papel importante durante a elaboração do tesouro e, por isso, pode ser necessário especificar ou delegar funções. Por conseguinte, para que haja um controle no processo editorial, os gerentes de projetos precisam definir, *a priori*, as atribuições de cada responsável na organização do tesouro e na manipulação dos dados.

Segundo a literatura analisada, os atores sociais podem assumir as seguintes funções durante a criação do tesouro:

**Quadro 43** - Funções dos membros da equipe editorial

<b>Atores sociais</b>	<b>Funções</b>	<b>Referências na literatura</b>
Administradores	Indivíduo responsável pelo gerenciamento do projeto de construção do tesouro. Supervisiona aspectos verticais de modelagem dos dados e autoriza as alterações realizadas em propriedades amplas ou específicas de classes de conceitos.	(QI; LUO; ZHU, 2010; BELLINI; CAPPuccio; NESI, 2011; STELLATO et al., 2015)
Lexicógrafos ou terminólogos	Especialistas que conhecem o repertório lexical de um domínio, os termos, seus usos e pares do discurso. Atuam na construção de vocabulários, em especial resolvendo questões de semântica e forma, especificidade e ambiguidade de termos.	(NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010; NOHAMA et al., 2012)
Editores de ontologias	Especialistas que conhecem as técnicas aplicadas à modelagem de dados, utilizando ontologias que melhoram as representações conceituais em tesouros.	(CARACCILO et al., 2013; PIOTROWSKI; COLAVIZZA; THIERY, 2014; ROSATI et al., 2017).
Validadores	Sujeitos que aprovam as alterações realizadas pelos lexicógrafos e editores de ontologia durante a construção. São especialistas com vasta experiência no domínio, os quais verificam a consistência semântica e estrutural dos	(CARACCILO et al., 2011; LAPORTE; MOUGENOT; GARNIER, 2012; ROSATI et al., 2017).

	conceitos dentro de suas hierarquias e os mapeamentos com outros vocabulários externos.	
--	---	--

**Fonte:** Elaborado pelo autor (2021).

Ao assumirem as funções especificadas anteriormente, os membros da equipe de construção reconhecem seus limites de atuação e só executam os procedimentos que lhes competem, contribuindo, dessa maneira, para a redução de erros no decurso da elaboração do tesauro. Não obstante, deve-se selecionar ferramentais que ofereçam o suporte necessário para que os procedimentos ocorram de forma colaborativa.

### 5.1.3 Seleção de modelos de dados e ferramentas para a gestão e a publicação do tesauro

Tradicionalmente criados de forma manual ou a partir de editores de texto, os tesauros, atualmente, precisam de ferramentas que atendam às necessidades de gerenciamento de vocabulários controlados no contexto da *Web Semântica*. Os recursos tecnológicos selecionados devem oferecer o suporte às atividades de construção e estarem em conformidade com os requisitos para o tratamento e publicação do tesauro como LOD.

Nesse sentido, reforça-se a importância de que as equipes de construção sejam constituídas por profissionais multidisciplinares. Por conseguinte, a atuação de profissionais da área de Tecnologia da Informação nesta etapa contribui para a seleção de ferramentas apropriadas ou até o desenvolvimento de ferramentas próprias. Isto posto, para selecionar as ferramentas apropriadas deve-se observar as funcionalidades de aplicação, a capacidade do sistema de manter a integridade do vocabulário, a interoperabilidade dos dados ou a integração dessas ferramentas com outros sistemas de informação.

Dessa maneira, com o intuito de auxiliar a seleção das ferramentas, os grupos de gestão do tesauro precisam responder aos seguintes questionamentos:

**a) Quais serão os modelos de dados utilizados para representar o tesauro?**

Deve-se dar atenção à compatibilidade com os modelos RDF e SKOS, os quais possibilitam que os tesauros sejam representados na *web* e publicados como LOD. Embora hajam outros vocabulários semânticos e de igual importância para a publicação de tesauros, o SKOS e seus modelos de extensão são os mais aplicados nos projetos analisados nesta pesquisa.

**b) Será necessário executar algum procedimento de extração de dados?**

Caso os tesauros estejam em outros formatos, como, por exemplo, na forma de registros de autoridades em bases bibliográficas, pode ser necessário o uso de extratores, que mapeiam o

conteúdo para diferentes padrões, incluindo o SKOS. Entre as ferramentas próprias para a extração de dados estão: *Helping Interdisciplinary Vocabulary Engineering*<sup>55</sup>, *Poolparty Semantic Search*<sup>56</sup>, *Poolparty Extractor*<sup>57</sup>, *Semantic Turkey*<sup>58</sup> e o *SKOSy*<sup>59</sup>.

c) **Quais ferramentas serão aplicadas para avaliar a qualidade dos dados?** É preciso avaliar a consistência e a integridade dos dados do tesauro. Para isso, são usadas ferramentas capazes de avaliar se o modelo de dados operado para a sua representação foi empregado de maneira correta. Esses aplicativos validam a qualidade da marcação e realizam a inferência e a verificação de restrições de integridade, funcionando como um teste de conformidade dos dados do tesauro para os padrões da *Web Semântica*. Entre as ferramentas possíveis estão: *qSKOS*<sup>60</sup>, *Poolparty SKOS Quality Checker*<sup>61</sup>, *ASKOSI*<sup>62</sup> e *Skosify*.

d) **Qual será a forma de apresentação dos dados do tesauro?** Podem ser apresentados como uma lista alfabética ou hierárquica. Seja qual for a diagramação escolhida, se mostrado na forma hierárquica tradicional ou mediante árvores de conceitos, deve permitir que os usuários facilmente recuperem, visualizem e naveguem pela estrutura de conceitos. Além da forma de apresentação, as tecnologias para a exploração de dados em linguagem RDF precisa atender aos requisitos do *SPARQL Protocol and RDF Query Language* (SPARQL). Os navegadores são aplicativos que aprovam a visualização ou a exploração de dados nos formatos da *Web Semântica* ou relacionados ao LOD, especialmente os vocabulários controlados expressos no modelo SKOS. Algumas ferramentas dão suporte para o desenvolvimento de interfaces de exploração e visualização dos dados, como: *ASKOSI*, *Callimachus*<sup>63</sup>, *Finnish Ontology Library Service ONKI*<sup>64</sup>, *Rhizomer*<sup>65</sup>, e *SKOS-Reader*<sup>66</sup>.

A definição de padrões de dados influencia diretamente na forma como estes devem ser armazenados e editados. Assim, diante da necessidade de selecionar ferramentas compatíveis com os modelos RDF e SKOS, os aplicativos selecionados precisam suportar o armazenamento e a recuperação de dados na forma de triplas que podem ser consultadas a partir

<sup>55</sup> <https://cci.drexel.edu/mrc/research/hive/>

<sup>56</sup> <https://www.poolparty.biz/semantic-search/>

<sup>57</sup> <https://www.poolparty.biz/poolparty-extractor/>

<sup>58</sup> <http://semanticturkey.uniroma2.it/>

<sup>59</sup> <https://code.google.com/archive/p/skosify/>

<sup>60</sup> <https://github.com/cmader/qSKOS/>

<sup>61</sup> <https://qskos.poolparty.biz/login>

<sup>62</sup> <http://askosi.org/>

<sup>63</sup> <http://callimachusproject.org>

<sup>64</sup> <https://onki.fi/>

<sup>65</sup> <http://rhizomik.net/html/rhizomer/>

<sup>66</sup> <https://labs.mondeca.com/skosReader.html>

de terminais SPARQL. Nesse aspecto, as ferramentas se dividem conforme as seguintes funções:

- (I) **armazenamento triplestore:** dão suporte ao armazenamento e à recuperação de dados RDF. Destacam-se as ferramentas: *AllegroGraph*<sup>67</sup>, *Apache Marmotta*<sup>68</sup>, *Linked Media Framework*<sup>69</sup>, *Mulgara Semantic Store*<sup>70</sup> e o *OpenLinkVirtuoso*<sup>71</sup>;
- (II) **editores:** ferramentas próprias para a edição de dados do tesauro nos padrões da Web Semântica. São exemplos de editores: *Poolparty*<sup>72</sup>, *SKOSed*<sup>73</sup>, *Tematres*<sup>74</sup>, *ThManager*<sup>75</sup>, *Vocbench*<sup>76</sup>;
- (III) **conversores:** ferramentas para alterar o formato do vocabulário semântico. Por exemplo, um tesauro elaborado como uma ontologia OWL pode ser convertido para o modelo SKOS ou vice-versa. Entre os aplicativos que cumprem este papel estão: *OWLtoSKOS*<sup>77</sup>, *SKOS Play!*<sup>78</sup>, *SKOS2OWL*<sup>79</sup> e o *Skosify*.

As ferramentas especificamente voltadas para a edição de dados dos tesouros são projetadas para assegurar os níveis de colaboração e guiar as etapas de construção em um fluxo de trabalho contínuo. Funcionam como espaços virtuais onde são conduzidos os acordos e as decisões que incidem sobre o tratamento e a modelagem dos dados.

À vista disso, o Sistema de gerenciamento escolhido deve combinar e configurar uma série de etapas de manipulação e análise de dados em sequências segmentadas e coerentes de instruções, de modo que cada etapa assume a forma de um componente em um serviço de *software*. Portanto, devem ser considerados os principais requisitos: a) oferecer mecanismos para que todos os membros da equipe, mediante ferramentas síncronas e assíncronas de comunicação, trabalhem de forma integrada; b) facilitar a troca e o compartilhamento de informações; c) conduzir as operações de maneira uniforme por fluxo de tarefas.

---

<sup>67</sup> <https://allegrograph.com/>

<sup>68</sup> <https://marmotta.apache.org/>

<sup>69</sup> <https://www.openhub.net/p/lmf>

<sup>70</sup> <http://docs.mulgara.org/>

<sup>71</sup> <https://virtuoso.openlinksw.com/>

<sup>72</sup> <https://www.poolparty.biz/poolparty-thesaurus-manager/>

<sup>73</sup> <https://code.google.com/archive/p/skoseditor/>

<sup>74</sup> <https://www.vocabularyserver.com/>

<sup>75</sup> <http://thmanager.sourceforge.net/>

<sup>76</sup> <http://vocbench.uniroma2.it/>

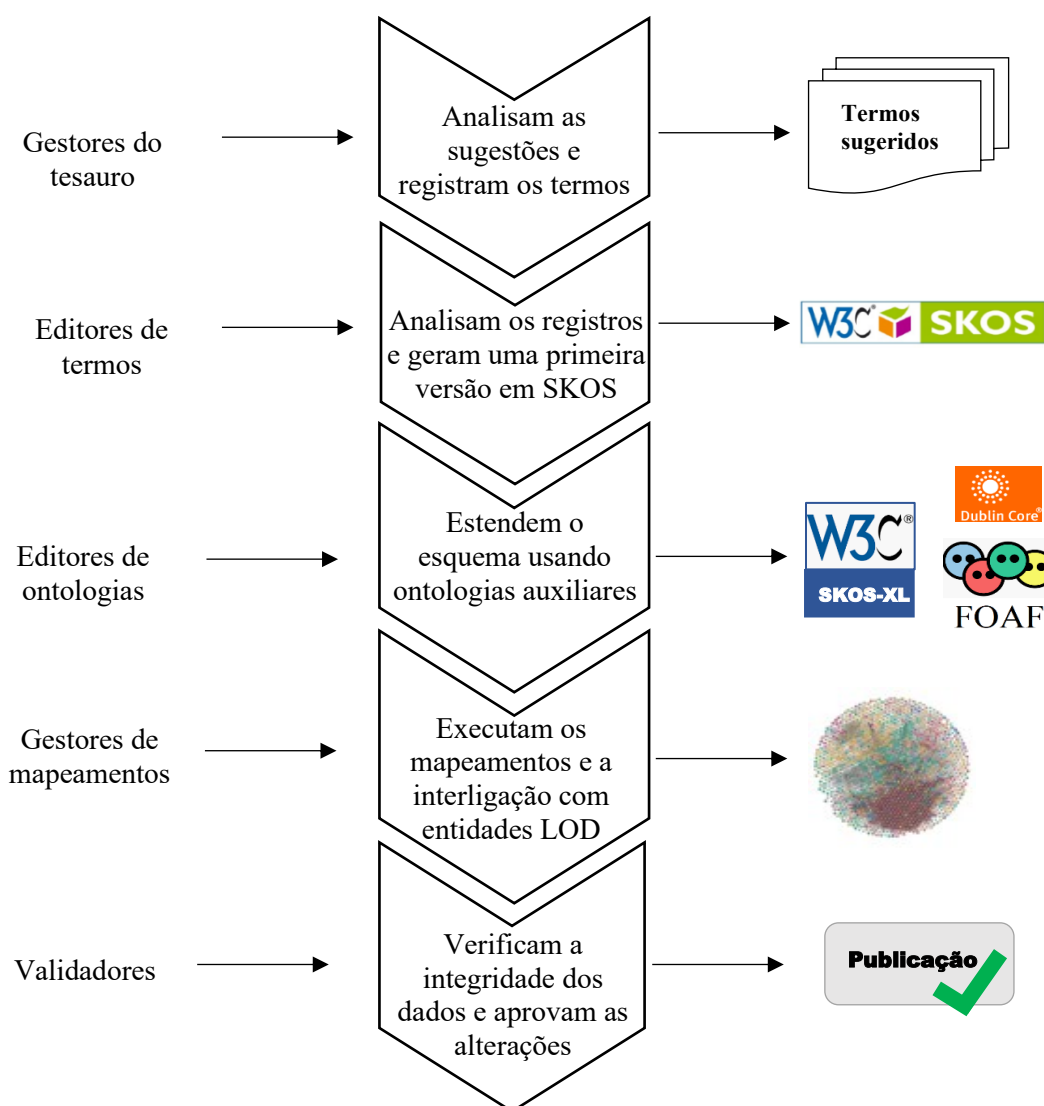
<sup>77</sup> <https://code.google.com/archive/p/owltoskos/>

<sup>78</sup> <http://labs.sparna.fr/skos-play/>

<sup>79</sup> <http://www.ebusiness-unibw.org/tools/skos2owl/>

A Figura 62 explica a visão geral das principais tarefas executadas pela equipe editorial conforme os papéis de atuação de cada profissional, definidos na seção anterior. Deve-se observar se as ferramentas de edição dão o suporte para o cumprimento de cada uma das etapas a seguir. A construção tem início com a análise dos termos selecionados e registrados no sistema de gerenciamento, suscitando uma primeira lista classificada de tópicos. As permissões são estabelecidas segundo a especialidade de cada membro da equipe e configurada através de um mecanismo de controle de acesso baseado em funções nos sistemas de gestão.

**Figura 62** - Visão geral do fluxo de tarefas na edição dos tesouros



**Fonte:** Elaborado pelo autor (2021).

A natureza colaborativa da edição faz com que o princípio do controle editorial ocupe um lugar de destaque diante da seleção das ferramentas, porquanto os sistemas devem assegurar que os dados sejam editados conjuntamente por diversos membros da equipe de construção. Igualmente, de modo que a estrutura semântica do tesauro não seja comprometida ao final da edição. Assim, os dados dos tesouros são preservados, mesmo quando especialistas executarem alterações simultâneas.

A maioria dos projetos de elaboração de tesouros como LOD, é formada por uma equipe de especialistas que executam modificações nos conjuntos de dados, os quais podem estar geograficamente dispersos e executarem operações simultaneamente (LAPORTE; MOUGENOT; GARNIER, 2012; CARACCILO et al., 2013; ABECKER et al., 2015). Os sistemas de gestão de tesouros são sistemas transacionais, ou seja, operam transações em bancos de dados. A maior parte dos sistemas transacionais precisa executar transações sincronicamente para atender aos requisitos de desempenho. Logo, sem controle de concomitância, esses sistemas não podem fornecer resultados corretos e sequer manter seus conjuntos de dados de forma consistente.

Segundo Zarzour, Al-Ayyoub e Jararweh (2018), para editar colaborativamente os armazenamentos de grandes conjuntos de dados conectados, o sistema de gestão deve fornecer, de forma automática, um suporte flexível para acompanhar ou compartilhar as modificações executadas pelos demais membros da rede e garantir o gerenciamento de possíveis conflitos. O Quadro 44 descreve algumas abordagens possíveis para o controle de simultaneidade.

**Quadro 44** - Abordagens para o controle de simultaneidade nas ferramentas de edição dos tesouros

<b>Abordagens</b>	<b>Descrição</b>
<i>last in wins</i>	quando dois ou mais usuários editam o mesmo conjunto de dados ao mesmo tempo, as alterações mais recentes são salvas.
<i>Optimistic concurrency control</i> (controle de simultaneidade otimista)	caso um usuário tente atualizar um conjunto de dados que está sendo atualizado por outro usuário da rede, sua atualização falhará e poderá ocorrer erro. Neste caso, esse sujeito pode descartar suas alterações ou conferir a nova sequência de dados antes de fazer uma segunda alteração. Essa abordagem pode ser chamada também de “ <i>first in wins</i> ”.
<i>Semi-optimistic concurrency control</i> (controle de simultaneidade semi-otimista)	bloqueia operações apenas para os casos em que pode haver violações de algumas regras e, em outras situações, não bloqueia, mas pode atrasar a verificação de regras até o final da transação, como é feito na abordagem otimista.
<i>Pessimistic concurrency control</i> (controle de simultaneidade pessimista)	os conjuntos de dados são bloqueados desde o momento que são recuperados até o momento que são

	atualizados. Enquanto um usuário está editando uma sequência de dados, nenhum outro pode lê-los.
--	--

**Fonte:** Elaborado a partir de Zarzour, Al-Ayyoub e Jararweh (2018) e Watson (2006).

As estratégias de controle de simultaneidade apresentam vantagens e desvantagens e suas aplicações dependem do tipo de sistema que se pretende implementar (WATSON, 2006). Conforme Zarzour, Al-Ayyoub e Jararweh (2018), o controle otimista promove uma consistência fraca ou eventual dos dados, pois permite acesso às réplicas de dados sem que haja anteriormente uma sincronização. Já o controle pessimista permite uma consistência forte à medida que impõe um esquema de acesso sequencial. No entanto, de acordo com Ibáñez et al. (2013), a abordagem pessimista pode não ser adequada para sistemas distribuídos, visto que muitos usuários podem gerar um número significativo de atualizações.

Concernente à aplicação das abordagens no âmbito dos sistemas de gestão de tesouros, a estratégia de controle otimista e do tipo “*last in wins*” foi observada em algumas ferramentas de edição, como o *PoolParty*, por exemplo.

Seja qual for a abordagem, a plataforma empregue para conceber o tesouro precisa oferecer um tipo de controle de simultaneidade que garanta atualizações realizadas pela rede de grupos temáticos e evitar que operações coocorrentes possam alterar a estrutura lógico-semântica do tesouro. Os sistemas e gestão que têm ambientes colaborativos para a construção de tesouros como LOD, são essencialmente de base *web* e adaptados ao modelo de computação em nuvem.

Algumas ferramentas são do tipo *open-source*, isto é, propiciados com uma licença de código aberto, o que permite utilizá-los livremente. A presença de *softwares* comerciais no mercado também é significativa. Logo, optar por *softwares* livres ou comerciais fica a critério de cada instituição, pois questões como disponibilidade orçamentária podem estar envolvidas, todavia deve-se observar primordialmente se as ferramentas contemplam os requisitos mínimos. Após a etapa de seleção das ferramentas para a gestão e publicação do tesouro, a equipe editorial prossegue com a definição dos tipos de licenças abertas para o reuso dos dados.

#### **5.1.4 Definição de licenças abertas para o reuso dos dados**

A publicação de dados abertos implica na disponibilização de informações na *web* de forma completa, permitindo que qualquer pessoa ou sistema possa acessá-las livremente e reutilizá-las para qualquer finalidade. Todavia, os produtores de dados precisam definir as condições de reuso que devem ser especificadas mediante uma licença aberta, na perspectiva

de esclarecer aos consumidores/usuários o que pode ser feito, ou não, com os dados. Dessa forma, a adoção de licenças abertas são meios de viabilizar o reaproveitamento de dados no âmbito dos projetos para a publicação de tesouros como LOD.

Pode haver diferenças quanto ao entendimento da necessidade de publicar de forma aberta os dados dos tesouros na *web*, sobretudo em razão da natureza das instituições, que variam entre pública ou privada. Vale ressaltar que a escolha de licenças abertas é opcional e a decisão sobre o modo de publicação aberta também depende das políticas institucionais de informação, da natureza dos dados e da legislação vigente (ASSUMPCÃO, 2018). Assim, para seleção ideal do tipo de licença é recomendado que as instituições respondam a algumas questões: Os dados poderão ser alterados? Os dados poderão ser usados para fins comerciais? Serão mantidos os direitos dos dados? Ou, ainda, serão distribuídos sob a mesma licença?

É preciso esclarecer que os tesouros são ferramentas que disponibilizam informações sobre termos que representam conceitos; organizados como um vocabulário controlado; desenvolvidos para dar suporte às atividades de indexação e recuperação da informação, auxiliando, inclusive, na difusão do conhecimento. Por isso, são naturalmente reconhecidos como dados de caráter aberto. Ainda, não são constituídos por dados pessoais sensíveis que possam infringir a lei de proteção de dados pessoais. Porém, se a instituições optarem por manter os direitos concernentes aos dados deve-se estimar aspectos legais para fins de sua reprodução, especialmente quanto à Lei de Direitos Autorais.

A Lei de Direitos Autorais<sup>80</sup> prevê a proteção de bancos de dados que constituem a criação intelectual, cabendo ao titular do direito patrimonial decidir autorizar a reprodução total ou parcial, entre outras questões relativas à modificação das informações, como tradução, adaptação e distribuição (BRASIL, 1998). Assim, as instituições podem decidir se os consumidores de tesouros poderão utilizar os dados como um todo, se novos dados lograrão ser construídos a partir deles, ou se estes serão reproduzidos para qualquer finalidade.

De modo geral, percebe-se na literatura que os tesouros como LOD, são criados e mantidos por setores governamentais ligados à área de patrimônio cultural, como Bibliotecas, Arquivos e Museus, ou por organizações não-governamentais, como institutos de pesquisa e agências internacionais, a exemplo as agências das Nações Unidas UNESCO e *Food and Agriculture Organization* (FAO). Todavia, se o tesouro estiver associado a algum setor governamental no Brasil, uma licença aberta deve observar a legislação brasileira não só no âmbito da Lei de Direitos Autorais, mas, principalmente, na Lei de Acesso à Informação.

---

<sup>80</sup> [http://www.planalto.gov.br/ccivil\\_03/leis/19610.htm](http://www.planalto.gov.br/ccivil_03/leis/19610.htm)



A Lei de Acesso à informação<sup>81</sup> afirma que o Estado deve oferecer acesso todas as informações que produz ou detém, a não ser que estas sejam consideradas de caráter sigiloso (BRASIL, 2011). Apesar de o Brasil não definir o tipo de licença que deve ser usado na publicação de dados abertos, Pires (2015) ressalta que alguns parágrafos da lei podem ajudar na formulação de uma licença aberta compatível com os princípios do LOD, em particular o Art. 8 § 3º:

Art. 8º É dever dos órgãos e entidades públicas promover, independentemente de requerimentos, a divulgação em local de fácil acesso, no âmbito de suas competências, de informações de interesse coletivo ou geral por eles produzidas ou custodiadas. [...]

§ 3º Os sítios de que trata o § 2º deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:

I - conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão;

II - possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações;

**III - possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;**

**IV - divulgar em detalhes os formatos utilizados para estruturação da informação;**

**V - garantir a autenticidade e a integridade das informações disponíveis para acesso;**

**VI - manter atualizadas as informações disponíveis para acesso; [...]**  
(BRASIL, 2011, grifo nosso).

O acesso automatizado e a legibilidade por máquinas são quesitos fundamentais para a sua inserção no contexto do LOD, pois uma vez disponíveis na *web* os dados poderão ser reutilizados tanto por *softwares* quanto por usuários humanos. Além disso, os projetos de dados abertos devem manter as informações atualizadas, oferecer informações acerca da autenticidade dos dados e especificar quais formatos são operados na estruturação da informação.

Nesse cenário, a ampla adoção de licenças abertas permite que os dados dos tesouros sejam distribuídos, copiados, reutilizados e modificados na maioria dos casos, desde que mencionadas a origem. Com base na legislação, os desenvolvedores podem criar suas próprias licenças ou utilizar uma licença existente. Entre as licenças existentes, Rautenberg et al. (2018) destacam que alguns tipos que podem ser utilizadas, especialmente as que concernem ao domínio público, atribuição e compartilhamento sob a mesma licença:

---

<sup>81</sup> [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2011-2014/2011/Lei/L12527.htm#art1](http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm#art1)

- a) Domínio público: licenças em que todos os direitos são concedidos, ou seja, o autor renuncia todos os direitos sobre os dados;
- b) Atribuição: quando a origem dos dados deve ser mencionada, isto é, os dados são creditados ao publicador original;
- c) Compartilhamento pela mesma licença: permite a publicação sobre as mesmas licenças dos dados originais.

Não há na legislação vigente brasileira qualquer diretiva a respeito da necessidade de restrição de acesso ou reuso de dados de tesouros na *web*, nem tampouco a definição ou recomendação de uso de uma licença específica. O que se sabe, segundo Pastor-Sánchez, Martínez-Méndez e Rodríguez-Muñoz (2012), é que as licenças mais aplicadas para publicar os conjuntos de dados dos vocabulários controlados como LOD, são geralmente as do tipo domínio público ou de atribuição, sendo as definições mais comuns: *Open Data Commons* e *Creative Commons*. O Quadro 45 destaca os principais tipos de licenças abertas.

**Quadro 45** - Tipos de licenças abertas para a publicação de tesouros como LOD

Organizações	Tipos de licenças abertas
<i>Creative Commons</i>	<p><b>Domínio público:</b> CC0 1.0: Universal – Declaração ao Domínio Público: permite que os licenciados renunciem a todos os direitos e coloquem os dados disponíveis no domínio público.</p>
	<p><b>Atribuição:</b> CC BY: Atribuição: considerada a mais flexível, consente a distribuição, remixagem, adaptação e criação a partir dos dados de origem, seja para fins comerciais ou não, desde que o crédito seja atribuído à fonte original dos dados.</p>
	<p><b>Compartilhamento sob a mesma licença:</b> CC AS: concede que os dados sejam alterados e adaptados, desde que sua distribuição utilize a mesma licença.</p>
<i>Open Data Commons</i>	<p><b>Domínio público:</b> Licença de dedicação de domínio público (ODC-PDDL): assente compartilhar, modificar e usar dados sem nenhuma restrição. Nenhum direito em relação ao dado será reservado ao autor.</p>
	<p><b>Atribuição:</b> <i>Open Data Commons Open database license</i> (ODbL): libera a distribuição e adaptação de dados, desde que o proprietário seja mencionado.</p>
	<p><b>Compartilhamento sob a mesma licença:</b> <i>Open Data Commons Attribution License</i> (ODC-By) 1.0: licença aberta para bancos de dados que inclui uma atribuição e requerimentos de compartilhamento.</p>

**Fonte:** Elaborado pelo autor (2021) a partir de (OPEN KNOWLEDGE FOUNDATION, 2008; CREATIVE COMMONS, 2018)

Além dos tipos mais comuns acima anteriormente, a licença de atribuição do *The Open Government Licence* (OLG) também pode ser utilizada. Ela foi criada, especificamente, para os setores públicos e permite que os consumidores de dados possam copiar, distribuir ou adaptar os dados de forma comercial e não comercial. É importante esclarecer que os tipos de licenças mais vantajosos são aqueles considerados flexíveis e amplamente aceitos pela comunidade LOD, e que a decisão final sobre o uso de licenças existentes deve prezar os objetivos do tesouro.

### 5.1.5 Políticas de gestão, atualização e revisão

No tocante às políticas de gestão, atualização e revisão do tesouro é preciso considerar não apenas as diretrizes para a sua elaboração, bem como as formas de uso pela comunidade do domínio. Nesse sentido, deve-se pensar como será realizada a manutenção da terminologia nos quesitos:

- a) admissão de sugestões de inclusão por usuários e indexadores;
- b) revisão das sugestões;
- c) monitoramento do uso de termos preferenciais;
- d) formas de divulgação das atualizações;
- e) alinhamento com outros vocabulários na *web*.

A atualização contínua torna a atividade terminológica mais eficaz. A ocorrência de novos termos e suas definições deve ser objeto de discussão permanente entre os responsáveis pelo tesouro. Novos termos podem surgir com a evolução natural da terminologia do domínio ou a partir das contribuições do público-alvo, ao observar a existência de termos que não traduzem suas necessidades de informação, podendo sugerir alterações junto à equipe de construção. Esta, por sua vez, recebe o *feedback* do usuário e, então, realiza uma série de revisões nos termos e nas estruturas hierárquicas de subordinação e superordenação.

Tanto os usuários como os indexadores devem dispor de mecanismos para sugerir a inclusão de termos no tesouro, os quais podem ser realizados mediante formulários eletrônicos, cujos colaboradores devem indicar o termo sugerido, incluir uma definição do conceito, indicar as fontes de referência aplicadas para fundamentar a sugestão e indicar formas alternativas (como sinônimos, por exemplo). Pode ser preciso indicar se o termo sugerido apresenta alguma relação com outros termos existentes no tesouro. Todas essas informações podem ser registradas no formulário, no qual pode ser inserida uma justificativa para a sua inclusão.

Regularmente, os editores precisam revisar as sugestões solicitadas. Para os casos de incerteza ou que exigem uma decisão mais acurada sobre a inclusão dos termos, estes podem ser submetidos à avaliação, no qual os novos termos são admitidos no sistema de gerenciamento e testados pelos indexadores. Havendo consenso, eles podem ser admitidos no tesauro.

Vale ressaltar que as alterações no tesauro podem influenciar diretamente na recuperação da informação. Quando um novo termo é introduzido, este fica disponível para a utilização, tanto na indexação como na busca. Todavia, os documentos indexados com o termo anterior podem não ser recuperados. Embora alguns sistemas implementem uma espécie de atualização automática, em alguns casos pode ser difícil efetivar correções automáticas retrospectivas. Assim, é preciso ponderar ou manter o equilíbrio entre conservar a estabilidade do tesauro ou atender completamente às novas atualizações em decorrência da evolução da terminologia.

Concernente ao alinhamento do tesauro com outros conjuntos de dados na *web*, é preciso considerar a necessidade de realizar revisões e melhoramentos constantes nos mapeamentos entre vocabulários. Pode ocorrer alterações significativas nos vocabulários de destino e isto prejudicar a qualidade dos vínculos com o vocabulário local. Diante disso, recomenda-se empregar tecnologias que possam identificar e sanar, de modo automático ou semiautomático, erros nos vínculos com rótulos de conjuntos de dados externos.

As revisões devem ser constantes e sucedidas conforme as necessidades de atualização da terminologia. Esses requisitos são considerados fundamentais para evitar a obsolescência do tesauro, porque se os usuários perceberem a ineficiência nas atividades de representação ou de recuperação da informação, podem abandonar o seu uso. Com isso, os tesauros passam a ser pouco (re) utilizados, seja por usuários humanos, seja por sistemas que utilizam-no como base para o desenvolvimento de aplicações na *web*.

Sob o ponto de vista da governança dos projetos de tesauros, alguns pontos sensíveis são apresentados por Hodge e Bedford (2018), tais como a participação de atores sociais envolvidos (organizações e grupos de pessoas) e o uso de normas. Por conseguinte, é preciso envolver a comunidade de usuários, disponibilizando canais de comunicação para que estes deem *feedbacks* precisos sobre o uso da terminologia. Também pode se estabelecer parcerias com organizações que (re) utilizam o tesauro, em especial tencionando otimizar os serviços de mapeamento entre vocabulários. Sobre o uso de métodos, o campo da Ciência da Informação apresenta diversas abordagens teóricas e metodológicas para a construção de tesauros, porém elas sempre serão interpretadas e adaptadas a um contexto de aplicação específico que pode

variar conforme os recursos disponíveis para a sua construção e as necessidades das organizações.

Após a atividade de planejamento, dar-se o início aos procedimentos de levantamento e seleção de termos que irão compor o tesauro. Nesta fase, todas as etapas estão bem definidas e são de conhecimento de todos os membros da equipe editorial, que devem ser acompanhados pelos gerentes de projeto, responsáveis também por realizar intervenções caso os profissionais não cumpram efetivamente seus papéis. Assim, para que o tesauro seja construído, a equipe deve usar da produção científica da área, de léxicos do domínio e da ajuda de especialistas e ferramentas tecnológicas para realizar o levantamento de termos candidatos à inclusão no tesauro. Tais procedimentos são discutidos na atividade de compilação de termos.

## 5.2 COMPILAÇÃO DE TERMOS

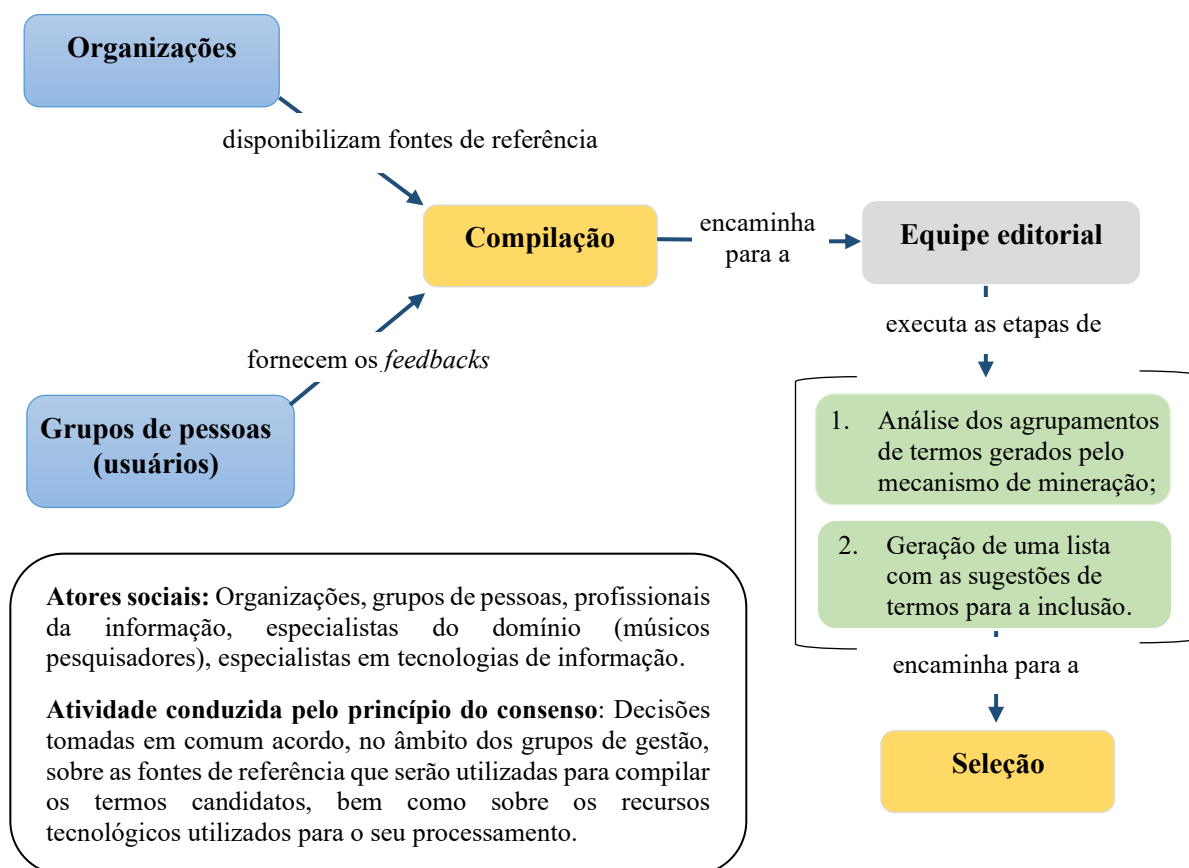
A atividade de compilação de termos usa técnicas para extrair palavras representativas em um *corpus* textual. Essas palavras constituem os termos candidatos à inclusão no tesauro. Há no mercado ferramentas de *software* que analisam documentos, extraem os principais tópicos do assunto e os agrupa conforme o número de ocorrência. Embora um tesauro criado automaticamente por computador não tenha a mesma qualidade de um tesauro preparado de forma manual, pressupõe-se que os procedimentos automáticos podem acelerar o processo intelectual dos especialistas.

Nesta oportunidade, sugere-se uma atividade de aquisição que busca extrair os termos candidatos a partir de fontes de referência e recursos de interação. São consideradas fontes de referência vocabulários, léxicos e publicações científicas da área, como tesouros, dicionários e enciclopédias especializados existentes e amostras da coleção de documentos das instituições, no qual o tesauro será utilizado na indexação, resumos de periódicos ou quaisquer outras fontes atualizadas em que a produção técnica e científica da área possa ser acompanhada.

Já os recursos de interação são caracterizados pelos relatórios de atividades de interação do usuário com as interfaces de busca do tesauro ou quaisquer outros meios em que se possa compreender como a comunidade está explorando e consumindo os dados do tesauro. Algumas estratégias podem ser lidadas, como analisar os *logs* de transação dos usuários do sistema e disponibilizar fóruns de discussão ou formulários eletrônicos para coletar *feedbacks*.

A Figura 63 demonstra como os principais atores sociais e os processos se articulam para o cumprimento da atividade de compilação de termos.

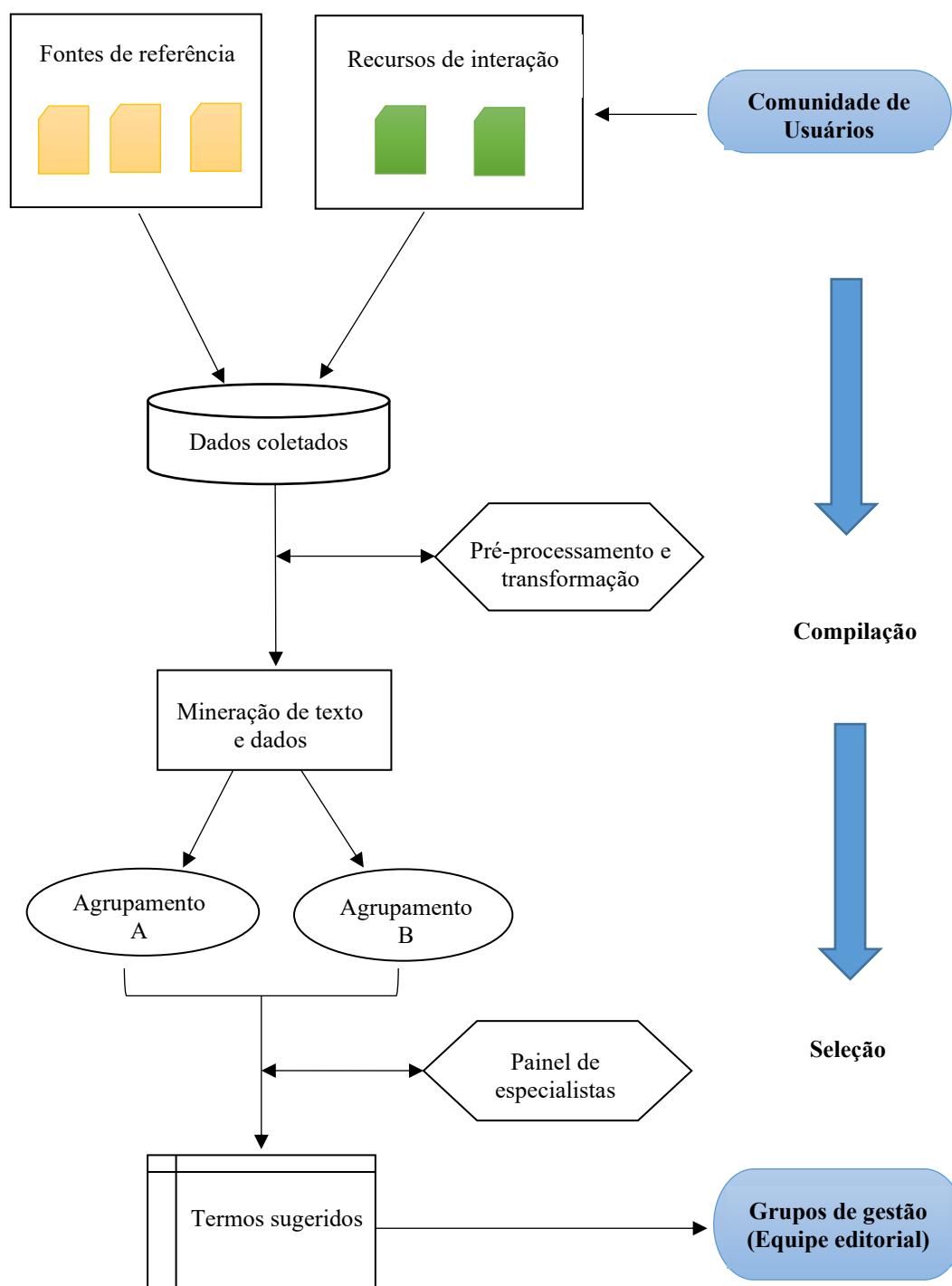
**Figura 63** - Etapas da atividade de compilação de termos



**Fonte:** Elaborado pelo autor (2021).

Essa etapa é fundamentada em uma sequência de procedimentos, conforme ilustrada acima, que mescla abordagens manuais, automáticas e semiautomáticas para extrair desses recursos uma lista contendo sugestões de termos. Tanto as fontes de referência como os recursos de interação têm origem na dimensão da comunidade de usuários, sendo os primeiros fornecidos por organizações e os segundos por pessoas ou grupos de pessoas.

**Figura 64** - Fluxograma da atividade de compilação de termos candidatos



**Fonte:** Elaborado pelo autor (2021).

No topo do fluxo, constam procedimentos de coleta, pré-processamento e transformação dos dados, os quais têm a função de preparar os dados para submetê-los a um mecanismo de mineração. Ao final, a metodologia de painéis de especialistas possibilita que, mediante atividade intelectual, os profissionais analisem os resultados produzidos pela ferramenta de mineração e construam uma lista inicial de termos sugeridos.

### 5.2.1 Coleta de dados

A etapa de coleta dos dados consiste no acesso aos recursos e no recolhimento de fragmentos textuais ou de dados que possam fornecer termos atualizados sobre o domínio de aplicação do tesauro. Essa etapa pode ser compilada de maneira manual ou automática e semiautomática. O método mais tradicional é do tipo manual, no qual são sucedidos *downloads* ou cópias de conteúdo em páginas eletrônicas na *web*, ou, ainda, recortes de conteúdos em fontes impressas. Mas, atualmente, há, ferramentas de *softwares* que simulam uma navegação humana para extrair informação específica de páginas na *web* usando técnicas de HTML *scraping* ou *web Crawler*.

Os procedimentos automáticos de coleta trazem benefícios para a equipe de construção dos tesouros, uma vez que utilizam de tecnologias para obter, de maneira rápida, conteúdos relevantes a respeito do domínio com custos de manutenção relativamente baixos.

Antes de executar a etapa de coleta, é preciso definir o que coletar e onde coletar. A aplicação de ferramentas automáticas para colher grandes volumes de dados em páginas eletrônicas pode ocasionar na sobrecarga de *sites*. Contudo, quando se trata de dados para a construção de tesouros, a reunião de termos candidatos geralmente é realizada em pequenos lotes. Neste modelo, três fontes de coleta podem ser aplicadas como fontes de referência:

- a) **Dados abertos:** entidades ou organizações que possuem léxicos especializados no domínio, bases de dados ou repositórios de conhecimento contendo publicações científicas e dados de pesquisa. Estes são disponibilizados por organizações governamentais ou não governamentais, ligadas aos setores de patrimônio cultural ou até universidades. As informações nessas fontes geralmente estão completas e em formato de arquivo convenientemente acessível e modificável, cujos recursos representam o *corpus* necessário para a execução da garantia literária.
- b) **Coleção da instituição:** são representados pelos documentos que fazem parte do acervo institucional. Vale ressaltar que o tesauro será operado como um instrumento para indexar os documentos que compõem os acervos das instituições. Que esses documentos podem dispor detalhes de como as principais classes de assuntos do domínio devem ser representadas e estruturadas no tesauro para atender às solicitações de busca do seu público principal.
- c) **Logs de transação:** são os relatórios de consultas realizadas pelos usuários em sistemas de informação eletrônicos. Os tesouros podem estar integrados aos catálogos bibliográficos, aos sistemas de gerenciamento de documentos ou



quaisquer outros recursos de busca e recuperação da informação na *web*. A estratégia de análise de *logs* de transação realiza um levantamento das atividades de interação dos usuários no sistema, especificamente dos percursos percorridos e de termos empregados por estes enquanto navegam e recuperam os conteúdos. A partir dos relatórios é possível, inclusive, identificar os tipos de usuários e as expressões de busca mais comuns. Vale sublinhar que a forma mais tradicional e fácil de coletar os *logs* de transação ocorre quando aqueles sujeitos acessam os sistemas de forma controlada, ou seja, quando precisam registrar sua entrada no sistema mediante *login* e senha. Essa abordagem é válida para os sistemas fechados ou que dispõem de algum mecanismo que identifica o perfil do usuário antes dele começar a explorar a interface de busca, como, por exemplo, a disponibilização de um formulário conciso no qual seja possível assinalar de forma rápida o tipo de usuário e o interesse de busca. Contudo, há outras possibilidades de analisar o comportamento daqueles indivíduos em sistemas de busca na *web*, como pelo uso da ferramenta *Google Trends*. A partir do *Google Trends*, é possível identificar a frequência com que um termo é buscado em dada página na *web*.

- d) **Fóruns de discussão:** páginas de tesouros na *web* devem dispor de canais de comunicação para que os usuários possam trocar experiências com a comunidade do domínio e estabelecer o diálogo com os responsáveis pela sua construção. A partir destes, é possível coletar informações sobre a qualidade dos termos, suas aplicações nas atividades de indexação ou de recuperação da informação. As interações dos usuários mediante fóruns públicos de discussão podem constituir os recursos necessários para que a equipe de construção execute a garantia do usuário.
- e) **Páginas *web*:** conteúdos de diretórios na *web*, como as enciclopédias eletrônicas disponíveis por meio das plataformas *Wikis*, representam um importante recurso para coletar dados, tendo em vista constar informações atualizadas sobre diferentes tópicos. Na maioria dos casos, o conteúdo dessas páginas é fornecido pela comunidade e, por isso, tais ferramentas também podem ajudar a executar a garantia do usuário. No domínio da Música, algumas plataformas colaborativas *Wiki* podem ser usadas, como o *International Music Score Library Project (IMSLP)*<sup>82</sup>, além da própria Wikipédia, que se destaca pela vasta quantidade de verbetes cobrindo diversos domínios.

---

<sup>82</sup> [https://imslp.org/wiki/Main\\_Page](https://imslp.org/wiki/Main_Page)

Considerando que usuários não especialistas ou especialistas no domínio usarão do mesmo ambiente para solicitar as buscas, os diversos perfis destes podem ser analisados a partir dos *logs* de transação. Logo, a convergência dos diversos tipos de comportamentos de busca representa uma alternativa para enfrentar a dificuldade de atender às demandas de um público diversificado, como os utilizadores do domínio da Música.

Apesar dos benefícios, os mecanismos automáticos de coleta também apresentam algumas limitações. É comum encontrar na *web* informações redundantes sobre um tópico, o que pode prejudicar a fase de processamento por computador, uma vez que a técnica de mineração de dados textuais opera algoritmos para calcular as ocorrências de palavras. Assim, caso os dados duplicados sejam submetidos aos mecanismos de mineração, naturalmente o número de ocorrência de determinados tópicos pode ser elevado.

Questões como volatilidade das páginas também são decisivos quanto à escolha das fontes de referência, pois o conteúdo da *web* é dinâmico, ou seja, pode ser atualizado com frequência ou até descontinuado. Diante disso, os usuários podem se deparar com páginas HTML com sintaxes incorretas ou erros de resposta do tipo 404. Isto posto, sugere-se um tipo de coleta criteriosa e supervisionada por humanos quando sucedida de modo semiautomatizada, para que o coletor *web* realize o levantamento de dados apenas em páginas confiáveis e seguras. Após a coleta dos *corpora* de texto e dados, uma fase posterior de pré-processamento e transformação é requerida.

### 5.2.2 Pré-processamento e transformação dos dados

Os dados brutos selecionados a partir de fontes de referência ou de recursos de interação são submetidos a uma fase de pré-processamento e transformação. Nesta fase ocorre a preparação dos documentos, de modo que estes possam ser submetidos ao mecanismo de mineração. No caso dos recursos de interação, é possível executar a transformação dos arquivos dos *logs* de transação em outros formatos. Isto ocorre porque, geralmente, os arquivos de *logs* são registros com extensão do tipo *.log*, e para que estes possam ser processados por máquinas ou até para a leitura de humanos precisam assumir formatos de texto ou de planilhas eletrônicas.

É importante sublinhar que o tipo de tratamento empregado varia conforme o tipo de conteúdo coletado. Quando se trata de fragmentos textuais mais extensos, como resumos de publicações científicas ou definições de verbetes coletados nos léxicos do domínio, faz-se essencial executar uma atividade de limpeza dos dados, cuja etapa busca eliminar ruídos e valores ilegítimos nos fragmentos de texto, ou seja, palavras sem significado. Para isso, são

utilizadas técnicas de Processamento de Linguagem Natural (PLN) que ajudam a converter a linguagem humana natural em representações formais, de modo que estas informações possam ser compreendidas por máquina.

A limpeza dos dados é descrita por Baeza-Yates e Ribeiro-Neto (2013) em três etapas:

- a) *Análise léxica*: interpreta o texto, transformando-o em termos, ou seja, converte uma sequência de caracteres em uma sequência de palavras. Fase conhecida também como *tokenização* ou *normalização*;
- b) *Eliminação de stopwords*: palavras desnecessárias que são muito frequentes no *corpus* textual, como, por exemplo, artigos, preposições e conjunções ou verbos, advérbios e adjetivos. Estes geralmente são excluídos;
- c) *Stemming*: quebra de palavras em porções. São removidas as partes que restam após a remoção de afixos (prefixos e sufixos). Os termos em tesouros são constituídos por substantivos ou grupos de substantivos com pouca variação sintática (plurais, gerúndios e sufixos);

A preparação de dados pode consumir mais tempo que a mineração de dados e possivelmente apresentar desafios semelhantes, ou até maiores. Por isso, Zhang, Zhang e Yang (2003) defendem a importância de executar de forma criteriosa esta etapa, na perspectiva de que os dados coletados possam ser impuros e apresentar muitos ruídos, logo precisam de sistemas de mineração que ofereçam padrões e desempenho de alta qualidade.

Quando se trata dos relatórios de *logs* de transação, é preciso compreender que estes podem ser constituídos de dados estruturados, ou seja, seu conteúdo não se apresenta organizado de maneira linear ou lógica como, a exemplo, nos resumos ou definições de verbetes. A forma de apresentação dos arquivos de *logs* varia segundo a configuração do servidor e a técnica usada para coletá-los. Estes arquivos podem registrar o número do IP da máquina operada para acessar o sistema, data, horário, as seções da página em que o usuário teve acesso, as expressões de busca e as fontes consultadas (LAIPELT, 2015). Desse relatório podem ser extraídos dados que ajudam a comparar os termos de consulta com os tipos de usuário do sistema.

Diferentes perfis de usuários podem empregar diferentes palavras ou frases para recuperar a mesma informação. Os *logs* ajudam a identificar como e quando as expressões de busca são aplicadas e como variam conforme as categorias de usuários. Um usuário especialista pode usar termos mais técnicos, ao passo que usuários não especialistas recorrem a expressões populares ou jargões. Nesse ponto, ferramentas de mineração de dados são úteis para realizar uma

análise estatística desses relatórios e identificar a frequência com que determinado termo aparece nas consultas realizadas pelos usuários aludidos.

Como os *logs* de transação são compostos por dados estruturados e apresentam sequências textuais bem menores, pode ser mais adequado realizar uma transformação nos formatos *Comma-Separated Values* (CSV) e arquivo de texto (formato de extensão txt). Estes formatos são de fácil leitura e podem ser manipulados por diferentes aplicativos. Depois de modificados, os registros são submetidos ao mecanismo de mineração. Mas, se alguma limitação tecnológica surgir há a alternativa de encaminhar para a análise de especialistas humanos.

### 5.2.3 Mineração de texto e dados

Após realizar a limpeza dos dados, são executados procedimentos semiautomáticos para identificar padrões úteis e desconhecidos no *corpus* coletado. Neste sentido, tanto a mineração de texto como a mineração de dados cumprem este papel, e a escolha de qual técnica será aplicada depende da natureza dos dados que serão analisados. Assim, no âmbito da recuperação da informação, em especial na construção de tesouros, a mineração de texto e de dados pode ajudar a identificar, classificar, agrupar e sumarizar um conjunto termos representativos.

Recomenda-se aplicar a técnica de mineração de dados (do inglês *data mining*) em dados estruturados, como os dados provenientes dos arquivos de *logs* de transação, pois geralmente se apresentam na forma de tabelas e bancos de dados, o que facilita o seu processamento. As ferramentas de mineração de dados lidam com algoritmos de mineração para extrair padrões até então desconhecidos, porém relevantes para o domínio. Estes algoritmos são responsáveis por analisar dados e gerar os *clusters* ou agrupamentos. Esta técnica exploratória de dados emprega medidas de similaridade para calcular a frequência com que os termos aparecem nos *corpus* coletado e tenciona agrupar entidades com características semelhantes observando os padrões de regularidades nesses dados (CORRÊA; MARCACINI; REZENDE, 2012).

Já a mineração de texto (do inglês *text mining*) se aplica a dados não-estruturados, como em fragmentos de texto. Os mecanismos de mineração de texto podem extrair termos, frases e entidades nomeados de informações não estruturadas usando modelos de conhecimento combinados com métodos de aprendizado de máquina, que lidam com tarefas de agrupamento

e classificação e objetivam organizar um conjunto de entidades similares intragrupos, e, igualmente, reconhecer entidades dissimilares entre grupos diferentes (EVERITT et al., 2011).

Para o cálculo de similaridade, os algoritmos são desenvolvidos com base em regras de associação para produzir a análise de coocorrência de pares de palavras, de modo a estabelecer índices que ajudam a identificar a força das associações entre as palavras presentes no *corpus*. Nesta etapa, consoante com Granada, Vieira e Lima (2012), ocorre uma análise das ocorrências de primeira ordem, ou ocorrência direta, quando dois ou mais termos aparecem em contextos semelhantes. O resultado dessa análise é a geração de agrupamentos de termos em classes temáticas<sup>83</sup>. O emprego dessas técnicas facilita a avaliação das relações entre os termos e a decisão da equipe editorial sobre sua relevância e inclusão no tesouro.

A presença de termos ambíguos no *corpus* coletado é um fenômeno que pode dificultar a atividade de mineração para os propósitos de construção de tesouros, visto que uma vez baseado no cálculo de frequência de palavras, os mecanismos podem agrupar termos semelhantes, mas que possuem significados diferentes. Assim, para ajudar as ferramentas a diferenciar conceitos que compartilham o mesmo significado, podem ser fornecidos os limites e contextos de aplicação de um termo executando a tarefa de Reconhecimento de entidades nomeadas (do inglês, *Named Entity Recognition*).

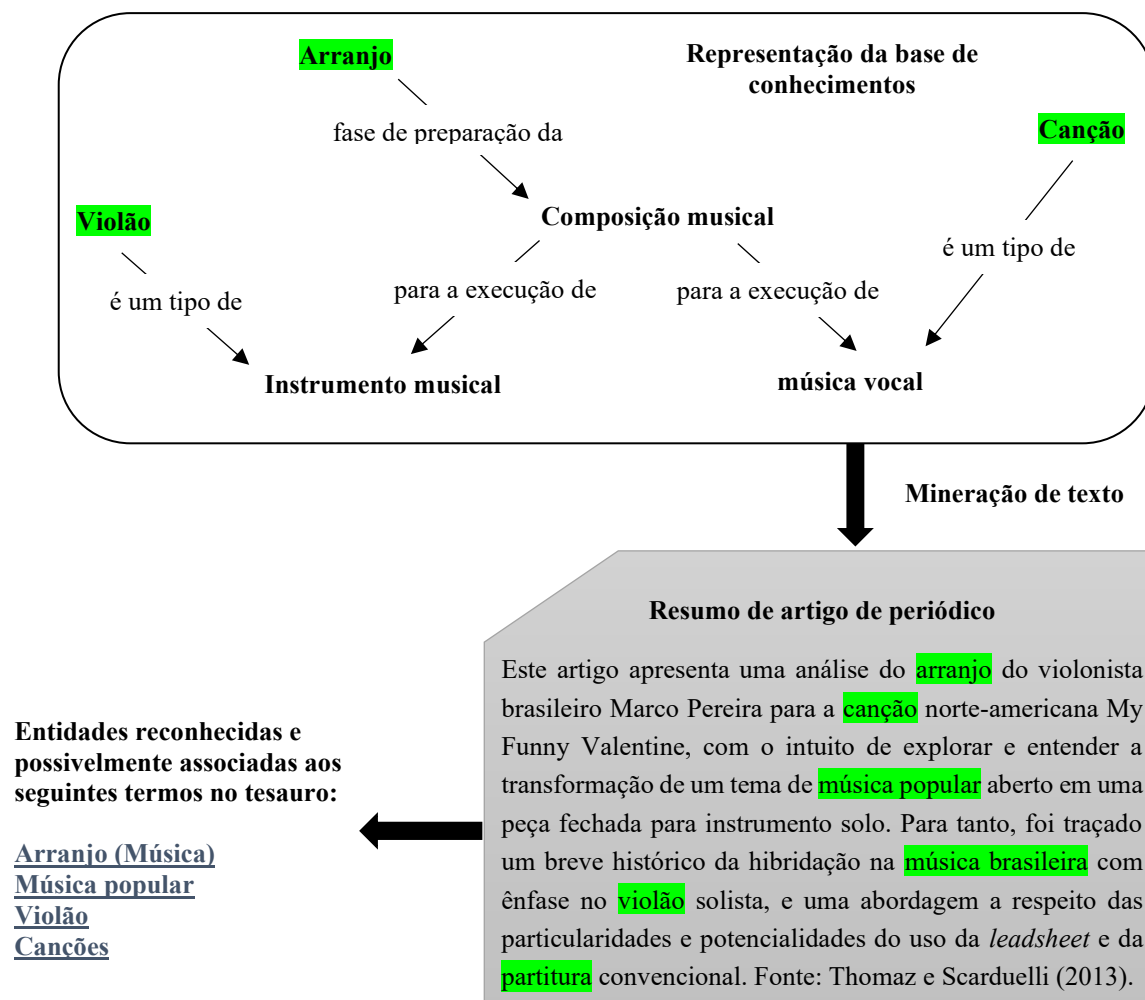
Esse Reconhecimento é um ramo da PLN também aplicada à mineração de texto que consiste em reconhecer, extrair e classificar elementos mencionados nos fragmentos de textos escritos em linguagem natural (SOTOMAYOR; VELOZ, 2017). A abordagem aplica bases de conhecimento ou uma camada semântica, como ontologias, taxonomias ou até mesmo de tesouros já criados para recuperar nos fragmentos de texto as entidades que apresentam algum grau de similaridade com os recursos dessa camada semântica, apresentando diversos rótulos possivelmente correspondentes.

A Figura 65 simboliza o seguinte exemplo: se um mecanismo de mineração encontra a palavra “Arranjo” no *corpus* coletado, ao utilizar uma camada semântica como componente de desambiguação, o mecanismo pode ajudar a identificar o contexto de aplicação da palavra. Neste caso, o fragmento concerne à preparação da composição musical para a execução ou a um artigo de decoração, como arranjo floral

---

<sup>83</sup> Está além do escopo da tese recomendar o uso de modelos estatísticos ou computacionais para executar o cálculo de coocorrência ou a análise de similaridade de termos. A comprovada eficácia de determinados modelos está condicionada à realização de testes e pode variar conforme tamanho e características linguísticas do *corpus*.

**Figura 65** - Exemplo de utilização de bases de conhecimento para o reconhecimento de entidades nomeadas em fragmentos textuais



**Fonte:** Elaborado pelo autor (2021).

Com isso, os fragmentos são colocados em contextos mais ricos e a desambiguação facilitada. O exemplo anterior mostra que o termo arranjo no fragmento textual concerne à fase de preparação da composição musical, portanto ligado ao domínio da Música. Além de identificar o contexto de aplicação, pode-se encontrar novos termos candidatos à inclusão, possivelmente relacionados aos termos que já existem no tesouro. Vale ressaltar que a abordagem de reconhecimento de entidades nomeadas pode ser mais útil para os casos em que os tesouros já foram construídos e, por isso, a aplicação da técnica de mineração de texto baseada nessa abordagem funcionaria apenas como um recurso de atualização do tesouro.

As técnicas de mineração de texto e de dados são abordagens que facilitam a atividade de aquisição de termos candidatos. Vale lembrar que elas não substituem o trabalho humano e uma fase posterior de seleção de termos é requerida. Na sequência, os especialistas prosseguem

com as etapas de análise dos termos sugeridos, com a definição dos termos preferenciais e a estruturação destes conforme as disciplinas do domínio de aplicação.

### 5.3 SELEÇÃO DE TERMOS

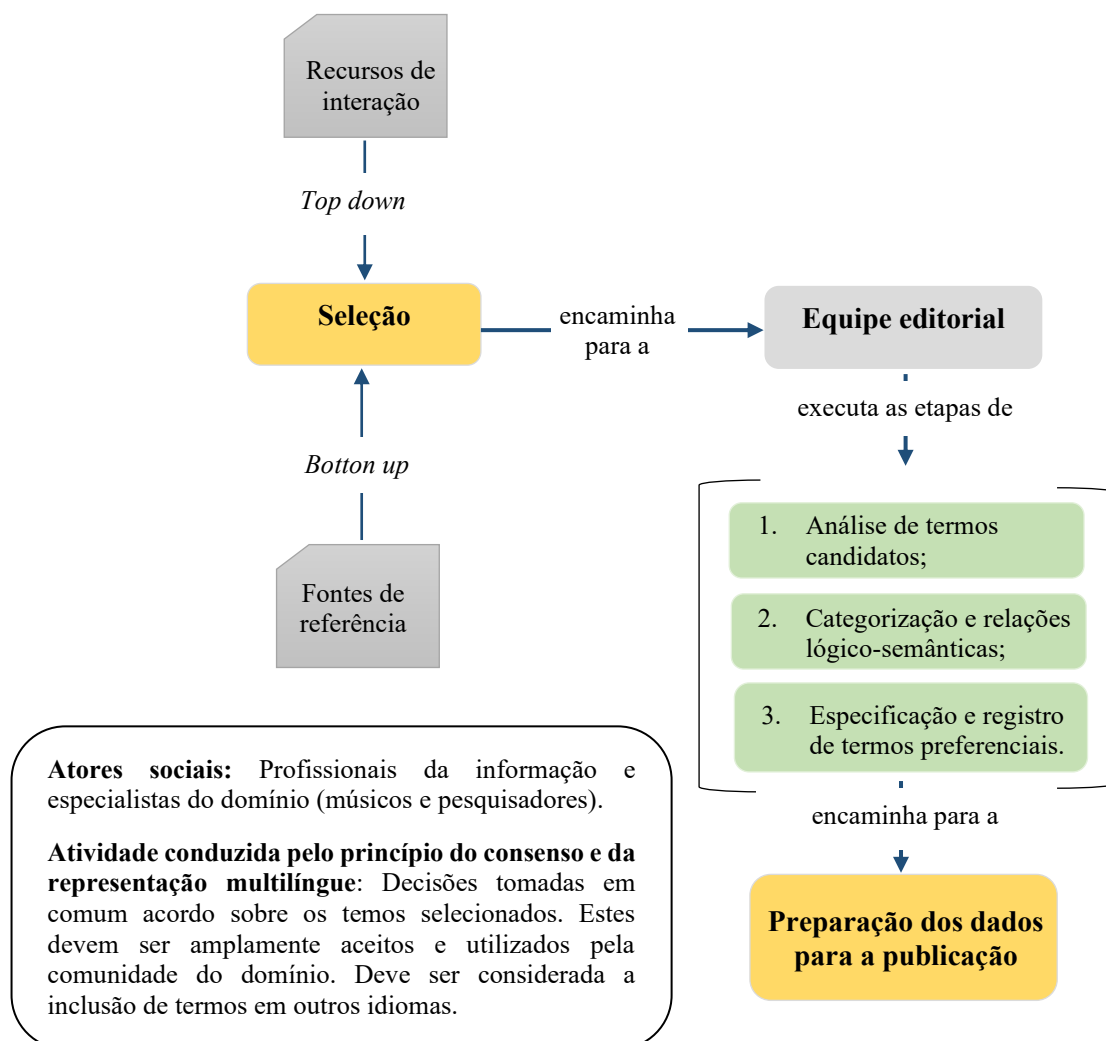
O processo de seleção de termos para a inclusão no tesauro envolve a análise de agrupamentos ou de listas de palavras geradas a partir da técnica de mineração de texto e de dados. Alguns aspectos precisam ser destacados, como: o domínio no qual o vocabulário será aplicado, a garantia literária, do usuário e estrutural e o nível de especificidade dos termos (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010).

Há dois métodos de seleção de conceitos (representados por termos) para a inserção no tesauro: o *top down* e o *botton up* (BARITÉ et al., 2010). O *top down* emprega um processo dedutivo que se apoia no endosso do usuário. Os conceitos são agrupados com o auxílio de especialistas que têm conhecimento acerca das principais classes ou facetas do domínio, os quais seguem uma divisão lógica orientada por classificações científicas ou especializadas conhecidas e compartilhadas pela comunidade do domínio. O *botton up*, por sua vez, aplica um processo indutivo cujos todos os conceitos que exigem descrição são coletados na literatura e os agrupamentos definidos posteriormente. Se apoia na coleta e análise de casos que aparecem na documentação especializada.

Uma abordagem híbrida que usa os dois métodos também pode ser aplicada, uma vez que assegura a organização mais correta do domínio à medida que estima tanto o endosso do usuário como a literatura. A justaposição desses dois métodos ajuda a compreender como a linguagem é aplicada na prática cotidiana da comunidade da área. As etapas da atividade de seleção são especificadas na Figura 66.

Recomenda-se adotar uma abordagem híbrida, na qual primeiramente são observados como os conceitos estão estruturados nas fontes de referência, visto haver alguns léxicos disponíveis do domínio da Música. Na sequência, são destacadas as demandas dos usuários, tendo em conta que o comportamento de busca pode indicar o uso mais apropriado de alguns termos. Com isso, a equipe de construção garante o maior nível de conformidade do tesauro com a linguagem lidada pelo usuário no momento da busca. Como resultado, os termos que aparecem com frequência na literatura são confrontados com aqueles seguidamente aplicados pela comunidade de usuários.

**Figura 66** - Etapas da atividade de seleção de termos



**Fonte:** Elaborado pelo autor (2021).

Decidir acerca da incorporação deles é uma tarefa complexa que deve ser alcançada mediante discussões com a equipe de gestão. Somente devem ser ponderados os termos em que todos os especialistas e profissionais da equipe editorial assintam com a inclusão. É importante lembrar que os tesouros podem não responder ao propósito de todos os usuários, mas a estrutura do domínio define seus limites de aplicação.

### 5.3.1 Análise de termos candidatos

Os termos candidatos são analisados e submetidos a uma série de procedimentos e acordos. A análise de coocorrência de primeira ordem, executada durante a etapa de mineração



dos dados, é mais efetiva para identificar relações sintáticas entre palavras. Uma exploração mais acurada desses agrupamentos pode ser efetuada por humanos mediante abordagem de painéis de especialistas.

Diante disso, e no sentido de definir os limites do domínio, algumas áreas temáticas podem ser estimadas, *a priori*. A construção primária de um quadro de referência ou uma árvore do domínio contendo as principais disciplinas do domínio da Música pode ser necessária. Segundo Rondeau (1984), a árvore do domínio representa o conjunto notacional do domínio a ser estudado. É usado como referência para que a equipe editorial possa identificar se os termos coletados na atividade de aquisição se enquadram em uma das áreas temáticas do domínio.

As principais disciplinas podem ser recolhidas em linguagens de indexação, empregues pela comunidade científica ou bibliotecária, como terminologias de assuntos e sistemas de classificação. Os termos são formulados para significar restritamente o domínio do tesouro e são criados à medida que as hierarquias são desenvolvidas. Uma lista inicial com os eixos temáticos, ou um plano de classificação geral, funciona como ponto de partida para que as hierarquias sejam construídas.

Sugere-se que os gerentes de projetos distribuam os especialistas do domínio em grupos de trabalho conforme suas áreas de especialidade, garantindo que a seleção de termos seja administrada por grupos temáticos competentes. Algumas das principais disciplinas do domínio da Música são especificadas no Quadro 46.

**Quadro 46** – Disciplinas e subdisciplinas do domínio da Música<sup>84</sup>

<b>Principais disciplinas</b>	<b>Descrição e exemplos de subdisciplinas</b>
<b>Teorias da Música</b>	Estruturas ou sistemas que formam a música nos aspectos da melodia, harmonia e ritmo. <b>Exemplos:</b> Teoria geral da música; Harmonia; Contraponto; Melodia; Teoria e formas da música nacional (vários países e povos); Teoria e formas da música da antiguidade.
<b>Composição musical</b>	Composição características e tipos de obras musicais, inclusive formas musicais, métodos de trabalho, equipamentos. <b>Exemplos:</b> Rondó; Canção; Sonata <i>allegro</i> ; Tema e variações; Concerto; Cópia de música; Reprodução de música.

<sup>84</sup> Não pretendeu-se elaborar uma lista extensiva, mas sim apontar algumas áreas temáticas do domínio da Música que constam nos índices da Terminologia de Assuntos da Biblioteca Nacional e no Sistema de Classificação Decimal Universal (CDU).

<b>História da música</b>	Compreende os períodos e fases da história da música, escolas, estilos e influências. <b>Por exemplo:</b> Música do período antigo. Música antiga; Música do período medieval; Música do período barroco; Música do período clássico; Música do período romântico. Música do período moderno.
<b>Profissionais da música</b>	Ocupações e atividades ligadas à música. <b>Exemplos:</b> Compositores; Maestros; Orquestras; Regentes de bandas; Bandas. Intérpretes; Professores de música; Produtores musicais.
<b>Tipos de execução ou apresentação musical</b>	Formas e métodos de reprodução de peças musicais. <b>Exemplos:</b> Concertos; Interpretação musical; Improvisação musical.
<b>Instrumentos musicais e conjuntos instrumentais</b>	Tipos de instrumentos e conjuntos musicais. São classificados conforme o som que é produzido. <b>Exemplos:</b> Instrumentos de cordas; Instrumentos de cordas e arco; Instrumentos de sopro; Instrumentos de percussão; Instrumentos de teclado; Orquestras.
<b>Gêneros da música</b>	Categorias de músicas que contemplam elementos semelhantes, como instrumentação, texto, função, estrutura e contexto. <b>Exemplo:</b> Música dramática. Ópera; Música popular. Música folclórica; Música de igreja. Música sacra. Música religiosa; Música vocal; Música instrumental.
<b>Música e tecnologia</b>	Aplicação de tecnologias para a gravação ou reprodução de registros sonoros. <b>Exemplos:</b> Música por computador; Música eletrônica; Composição musical por computador; Gravação e reprodução do som; Gravação musical; Mixagem.
<b>Música e sociedade</b>	Aplicação da música na indústria, comércio, ou na vida quotidiana. <b>Exemplos:</b> Música no teatro; Música no cinema; Filmes musicais; Indústria musical; Música e antropologia; Música em anúncios.

**Fonte:** Elaborado pelo autor (2021) a partir da Terminologia de Assuntos da Biblioteca Nacional e CDU.

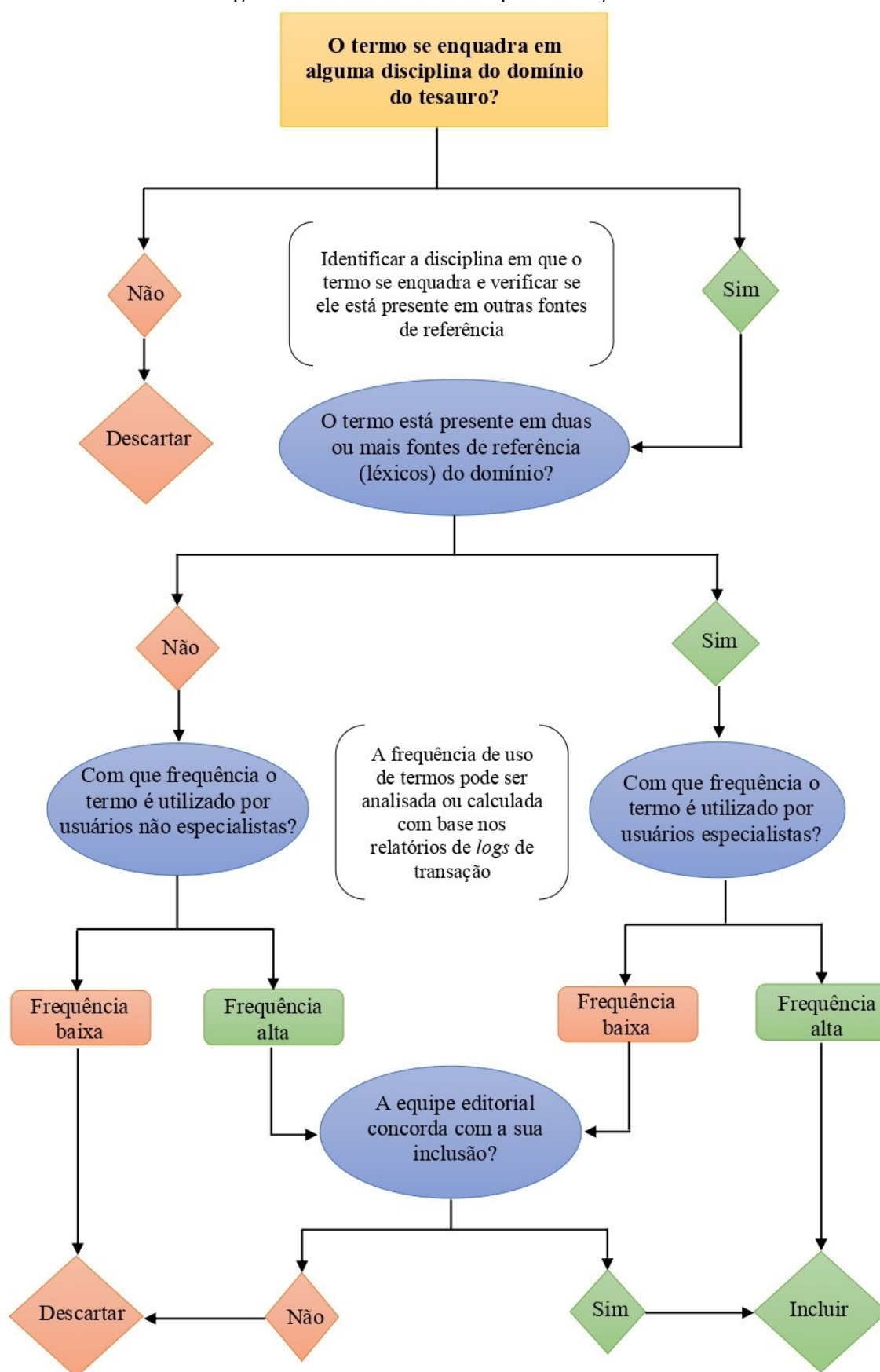
O princípio do consenso nesta etapa é garantido quando a seleção de termos se apoia em fontes de referência especializadas amplamente lidadas pela comunidade do domínio. A legitimidade dos termos compilados também é atestada ao passo que acordos incidem sobre a inclusão de termos pouco frequentes. Esses acordos refletem o consentimento de todos os especialistas acerca da necessidade de incorporá-los ao tesouro. Discussões a respeito da pertinência dos termos são essenciais nesta fase. Para isso, as equipes editoriais devem definir bem os critérios de escolha.

Nesse contexto, para algumas áreas em que as diferenças no uso de termos por especialistas e não especialistas for significativa, é recomendado que as equipes de construção decidam quanto à inclusão observando tanto a frequência com que aparecem na documentação quanto ao uso e efetividade na recuperação da informação. Para subsidiar tal processo, propõe-se que a equipe editorial use técnicas para facilitar a produção de acordos. Isso pode ser feito mediante uma árvore de decisão, como mostra o exemplo na Figura 66.

Nesta etapa, os especialistas precisam, basicamente, responder aos seguintes questionamentos:

- a) O termo se enquadra em alguma das disciplinas do domínio do tesouro?
- b) O termo está presente em duas ou mais fontes de referência (léxicos) do domínio?
- c) Com que frequência o termo é utilizado por usuários não especialistas?
- d) Com que frequência o termo é utilizado por usuários especialistas?
- e) A equipe editorial concorda com sua inclusão no tesouro?

Figura 67 - Árvore de decisão para a seleção de termos



Fonte: Elaborado pelo autor (2021).

Quando o termo se enquadra em alguma disciplina do domínio da Música, o passo seguinte é verificar se ele está presente em duas ou mais fontes de referência do domínio. Quando não for possível encontrá-lo nessas fontes, provavelmente ele pode ser empregado com frequência por usuários não especialistas. Caso um termo não encontrado em fontes de referência for muito aplicado como expressão de busca por usuários não especialistas, a equipe editorial deve consensuar acerca de sua inserção ou exclusão. Vale ressaltar que as ferramentas operadas na seleção precisam dar suporte à equipe de construção, a fim de promover a interação e a articulação de todos os membros da equipe para que as decisões possam ser tomadas em conjunto.

Assim, antes de iniciar a atividade de tratamento de dados para a publicação do tesauro como LOD, é necessário categorizar os termos selecionados tentando refinar e melhor estabelecer suas relações lógico-semânticas. Isso pode ser estabelecido, principalmente, com base na análise de definição dos conceitos do domínio, ou, ainda, nas principais áreas ou disciplinas que compreendem-no.

### **5.3.2 Categorização e relações lógico-semânticas**

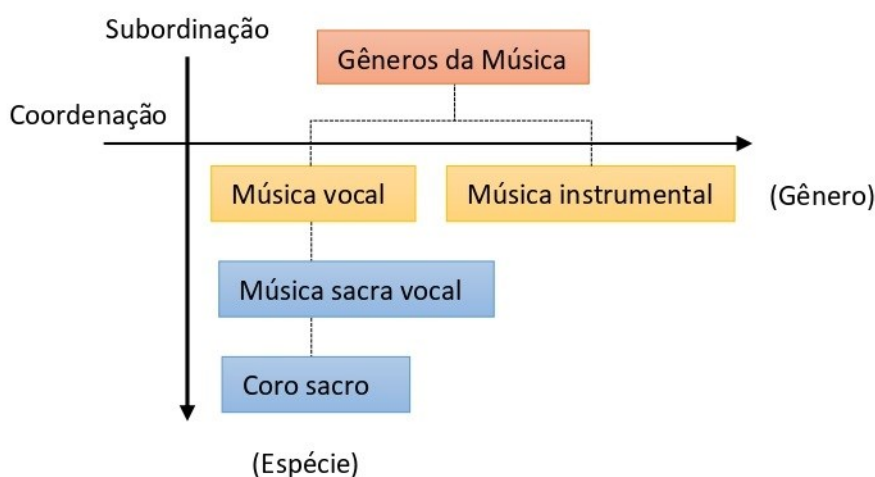
Esta etapa objetiva refinar a lista inicial de termos e estruturá-los conforme as principais disciplinas e subdisciplinas do domínio. Mediante categorização, os termos são organizados nos principais eixos temáticos e, logo em seguida, são estabelecidas as relações lógico-semânticas que há entre os conceitos (representados por termos). Estes podem ser organizados por facetas (ex.: ação, processo, propriedade) ou por temas (principais disciplinas do domínio). Ainda, são classificados, ou seja, reunidos aqueles que se assemelham ou que fazem parte de uma mesma classe. Posteriormente, são indicadas as relações entre os termos, estruturando-os de acordo com os relacionamentos semânticos, que podem ser dos tipos hierárquicos, associativos ou de equivalência.

Os tesauros compreendem uma estrutura sistemática elaborada na forma de categorias de assuntos que se desmembram em níveis hierárquicos menores. As macroestruturas podem ser organizadas na forma de árvore de categorias para melhorar a compreensão acerca dos principais eixos temáticos do domínio. Embora as relações semânticas indiquem os relacionamentos mais específicos ou mais abrangentes de um conceito, as categorias de assunto apontam para um contexto maior do domínio especificado.

Assim como explicitado na abordagem relacional de Dahlberg (1978), os relacionamentos entre os conceitos podem ser estabelecidos após a análise de suas definições. Considera-se, então, que ao analisar a definição dos conceitos que integram o sistema é possível identificar suas características principais e melhor identificar e estruturar as relações que existem entre eles. O método proposto pela autora em realce envolve três elementos básicos: o conceito, a análise conceitual e a definição do conceito.

Considerando a definição do conceito, os relacionamentos podem ser estabelecidos ao se observar os níveis de coordenação e subordinação. A coordenação é conhecida por considerar os relacionamentos sintagmáticos entre as classes, em que o conceito é significado por uma ideia ou gênero maior, que expressa uma intenção, enquanto que a subordinação estima os relacionamentos paradigmáticos, no qual observa-se o conjunto de ideias relativas ao gênero maior, validadas como uma extensão das classes principais. Assim, as hierarquias são organizadas tendo como referência um conjunto de ideias, especialmente em função da finalidade ou extensão de uma ideia principal. Para uma melhor compreensão, a Figura 68 ilustra como podem ser identificados os níveis de coordenação e subordinação dos conceitos.

**Figura 68** - Coordenação e subordinação de conceitos



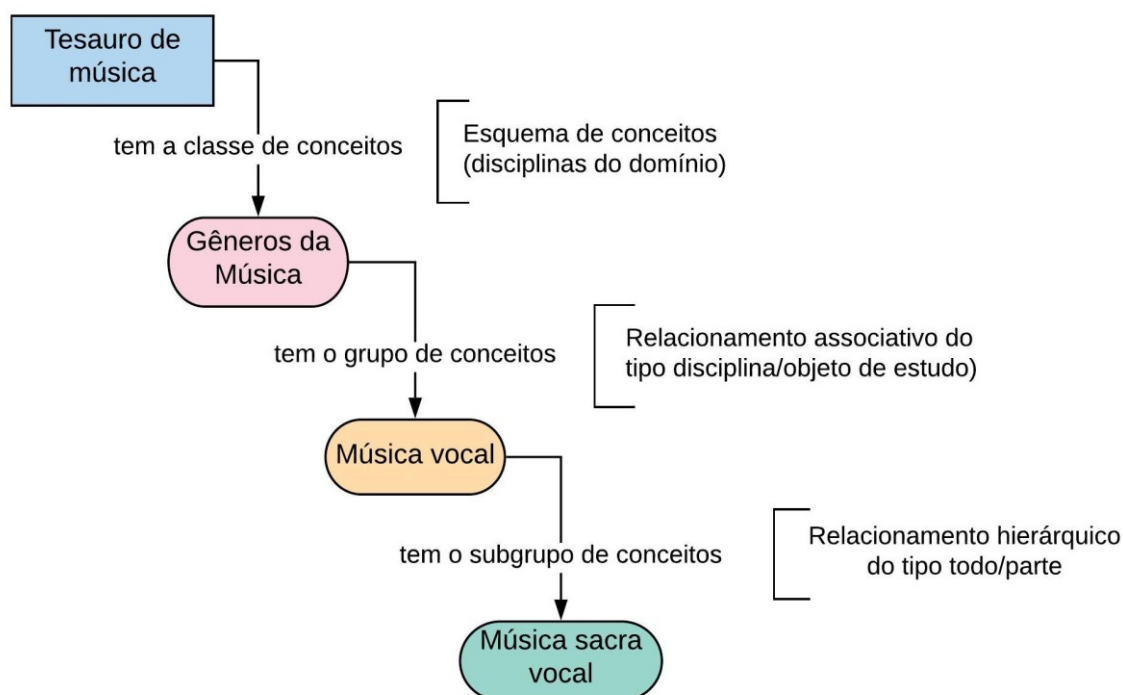
**Fonte:** Elaborado pelo autor (2021) com base em (DODEBEI, 2002).

Quanto ao domínio da Música, considera-se que as definições dos conceitos exigem a inclusão de outro conceito necessário à sua explicação, e que estes podem também revelar suas delimitações em termos de abrangência e especificidade. Por exemplo, “Gêneros da Música” ou “Gênero musical” pode ser definido como categorias de peças musicais que contemplam elementos semelhantes, como instrumentação, texto, função, estrutura e contexto. As peças

musicais, por sua vez, atribuídas à execução de instrumentos musicais ou de vozes (música vocal). Esta última pode ser descrita pelo número de intérpretes (solistas e conjuntos vocais) ou em função do texto (sacro e profano). Isto posto, admite-se, então, que o conceito de “Música sacra vocal” está relacionado ao conceito de “Música vocal”, e o conceito de “Coro sacro” associado ao conceito de “Música sacra vocal”.

Outra possibilidade é a organização dos termos conforme as disciplinas do domínio. A Figura 69 ilustra a composição de uma macroestrutura organizada em grupos de principais disciplinas e suas subdivisões em temas específicos da Música. Além da habitual divisão temática com as principais disciplinas do domínio (para o agrupamento das categorias de assunto e seus desdobramentos), o tesouro necessita estar organizado de modo a conter subgrupos de conceitos. Estes subgrupos especificam a relação entre os termos das subdisciplinas com as respectivas disciplinas, permitindo que todos os termos estejam vinculados a uma classe superior de conceitos. Isto facilita a manutenção da estrutura lógico-semântica.

**Figura 69** – Exemplo de grupos e subgrupos de conceitos



**Fonte:** Elaborado pelo autor (2021).

Igualmente, pode ser necessário acrescentar grupos de conceitos pouco observados nas fontes de referência, como, por exemplo, áreas associadas à aplicação da música na vida cotidiana (atividades comerciais, industriais e culturais). Com isso, acredita-se que as

necessidades dos usuários não especialistas também poderão ser contempladas. Elucidando, se um usuário se interessa por trilhas sonoras de filmes cinematográficos, a busca necessariamente aplicar-se-á à música no Cinema. Caso seja por música para uso em uma atividade física ou de lazer (música dançante, música para meditação, etc.), ele fará alusão aos gêneros musicais ou ao ritmo e melodia da música. Tais classes de assuntos são associadas às disciplinas Gêneros da música ou Música e sociedade.

Para atender ao princípio de representação multilíngue durante a fase de construção, a participação de especialistas em tradução se torna fundamental. Neste quesito, deve-se estimar a correspondência dos rótulos no nível semântico. Portanto, os terminólogos precisam se ater à completude dos termos selecionados no nível conceitual, tendo em conta a definição em todos os idiomas do tesouro. Caso não haja uma compatibilidade semântica é recomendado não incluir o termo no outro idioma. Este é o maior desafio para a construção de tesouros multilíngues, visto que todos os termos precisam estar alinhados semanticamente. De outra forma, deve-se prezar pela representação multilíngue no nível dos relacionamentos dos termos com outros tesouros ou vocabulários existentes, atendendo, assim, ao princípio da interoperabilidade.

A análise e seleção de termos não deve se resumir ao processo automático de compilar palavras frequentes na literatura, mas também suas formas de uso e aplicações no cotidiano. Como o domínio da Música compreende contextos de uso diversos, aspectos multiculturais são apreciados, tal como as formas variantes de um termo em detrimento aos aspectos sociolinguísticos de cada região ou país. De igual modo, o *corpus* de coleta consegue apresentar diversos termos, podendo haver alguns que correspondem ao mesmo conceito. Todavia, apenas alguns conjuntos de termos devem ser selecionados e designados como preferências.

### **5.3.3 Especificação e registro de termos preferenciais**

Uma vez selecionados e categorizados, os termos mais representativos do domínio na etapa anterior precisam ser verificados quanto à conformidade com os padrões estabelecidos para rótulos preferidos. Estes padrões são consensuados com os especialistas do domínio, que analisam e alteram as representações conforme for necessário.

As normas para a construção de tesouros especificam a forma mais apropriada de selecionar os rótulos preferidos para os conceitos, as quais consideram, principalmente, as variantes lexicais mais usadas pela comunidade. As formas variantes não selecionadas, apesar de não representarem o uso habitual do conceito, podem ser inclusas no vocabulário como



rótulos alternativos. Assim, são elaboradas referências cruzadas ou remissivas para termos sinônimos ou quase sinônimos que podem ser representados no tesauro como não preferenciais.

Quando a equipe editorial opta pelo uso de uma expressão, as preferências da comunidade devem ser analisadas e, à medida do possível, atendidas. Os especialistas devem observar os seguintes critérios durante a seleção de termos preferenciais:

**Quadro 47** - Critérios para a definição de termos preferenciais do tesauro

Critérios	Especificação
<b>Aspectos linguísticos e culturais</b>	Os termos podem apresentar formas ortográficas que variam de acordo com os aspectos linguísticos ou culturais de cada região, que devem ser escolhidas em razão do público-alvo. Exemplo: <b>Termo preferido:</b> Acordeão <b>Usado para</b> Sanfona <b>Usado para</b> Gaita
<b>Grafias</b>	Caso várias grafias existam e se são reconhecidas pela comunidade. A grafia mais aceita deve ser adotada para termos preferenciais. As formas variantes têm que ser inseridas no tesauro como termos não preferenciais, e nos registros constar as referências estabelecidas reciprocamente. Exemplo: <b>Termo preferido:</b> Acordeão <b>Usado para</b> Acordeon
<b>Homógrafos</b>	Homógrafos são palavras que apresentam a mesma grafia, mas podem ter significados diferentes. O uso de homógrafos como termos requer esclarecimentos quanto ao seu significado. Para isso, aplica um qualificador que especifica o domínio do significado ao qual o termo pertence, tornando-o inequívoco. Exemplo: Arranjo (Música) <b>Diferente de:</b> Arranjo (Decoração floral)
<b>Termos em outros idiomas</b>	Termos em outros idiomas podem ser admitidos se amplamente aceitos pela comunidade de usuários. As referências recíprocas também devem ser estabelecidas entre os dois termos. Exemplo: <b>Termo preferido:</b> Música - Execução <b>Usado para</b> Performance musical
<b>Neologismos, gírias e jargões</b>	Devem ser eliminados gírias ou jargões, visto serem transitórios. Alguns neologismos podem ser usados, porém são frequentemente cunhados em resposta a fatores emergentes e somente devem ser adotados quando considerados úteis para a recuperação da informação. Exemplos: <b>Termo preferido:</b> Música popular

	<b>Usado para Música pop</b>
<b>Nomes comuns e nomes comerciais</b>	<p>Alguns produtos podem ser conhecidos pelos nomes comerciais. Contudo, havendo um nome comum adequado ele deve ser adotado como o termo preferido, e o nome comercial admitido como um termo não preferido. Desse modo, se um produto genérico é vendido sob vários nomes comerciais concorrentes e os usuários desejam distinguir entre as diferentes marcas, o nome comum pode ser adotado como termo preferido, e os nomes comerciais empregados como termos mais restritos.</p> <p>Exemplo:  <b>Termo preferido:</b> Fonógrafo  <b>Usado para</b> Gramophone</p>
<b>Nomes populares e nomes científicos</b>	<p>Por ventura, se um nome popular e um nome científico concernem ao mesmo conceito deve-se dar preferência à forma mais provável de ser procurada pelos usuários do tesouro. Referências recíprocas precisam ser feitas nesses casos.</p> <p>Exemplo:  <b>Termo preferido:</b> Fonógrafo  <b>Usado para</b> Toca-discos</p>
<b>Abreviações e acrônimos</b>	<p>Abreviações e acrônimos são geralmente ambíguos, pois podem se referir a mais de um conceito. Portanto, a forma completa do nome normalmente deve funcionar como o termo preferido, com uma referência recíproca da forma abreviada. Excepcionalmente, abreviações ou acrônimos podem ser usados como termos preferenciais quando são amplamente conhecidos, inequívocos e facilmente compreendidos dentro do domínio do tesouro. Ainda, devem ser feitas referências recíprocas entre o termo completo e sua abreviação.</p> <p>Exemplo:  <b>Termo preferido:</b> MIDI (Normas)  <b>Usado para</b> Musical Instrument Digital Interface</p>
<b>Nomes próprios</b>	<p>Pode haver formas variantes para nomes de pessoas, lugares ou instituições. As formas mais conhecidas pela comunidade do domínio devem ser admitidas. Os nomes próprios costumam ser úteis em um tesouro especializado, todavia é possível tornar a estrutura extensa caso hajam muitos candidatos. Em se tratando do domínio da Música, a inclusão tem o benefício adicional de que vínculos hierárquicos ou associativos podem ser estabelecidos entre termos de assunto e nomes próprios relevantes, como, por exemplo, quando o último representa instâncias do primeiro.</p> <p>Exemplos:  <b>Grupo de termos:</b> Festivais de música  <b>Termo preferido:</b> Coachella Fast  <b>Usado para</b> Coachella Valley Music and Arts Festival</p>

**Fonte:** Elaborado pelo autor (2021) com base em (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010; INTERNATIONAL STANDARD ORGANIZATION, 2011, 2013)

Dada a variedade e ambiguidade da terminologia da Música, é inegável ser preciso uma linguagem padronizada para alcançar um alto nível de consistência na descrição dos recursos de informação musical. O campo da Música pode ser qualificado por uma justaposição de diferentes escolas e modelos teóricos. Pode haver significativas diferenças no significado de termos que parecem similares, ou que são lidos erroneamente por membros da comunidade de usuários não especialistas.

Para ilustrar esse fenômeno, tome-se como exemplos os termos Coro ou Canto coral. O coro concerne a um grupo de cantores agrupados em naipes (tipo de vozes). Já o canto coral tange ao conjunto de atividades ligadas ao coro, também conhecida como orfeão (termo originário da escola francesa). Nesse contexto, também podemos observar um caso de homonímia quando nos referimos a esses grupos corais. A palavra coral é, igualmente, usada para aludir à classe de animais marinhos. Dessa forma, é essencial esclarecer o contexto em que um termo é usado. No tesouro, isso pode ser feito no nível linguístico, aplicando um rótulo preferido não ambíguo ou atribuindo um qualificador para deixar claro o domínio de aplicação.

Nesta fase de seleção, os especialistas devem registrar as possibilidades de aplicação dos termos como numa espécie de registros em fichas terminológicas eletrônicas. Para a etapa de registro são utilizados *softwares* aplicativos desenvolvidos notadamente para a gestão de tesouros. Geralmente, os termos são armazenados em bases de dados relacionais que permitem que os dados sejam extraídos e manipulados. Apesar da existência de ferramentas dedicadas à construção de tesouros nos diferentes modelos de dados adotados para a *Web Semântica*, é preciso definir previamente como esses dados devem ser modelados, independente do aplicativo empregado para a construção. Com isso, a equipe editorial inicia a atividades de tratamento dos dados do tesouro para publicação como LOD.

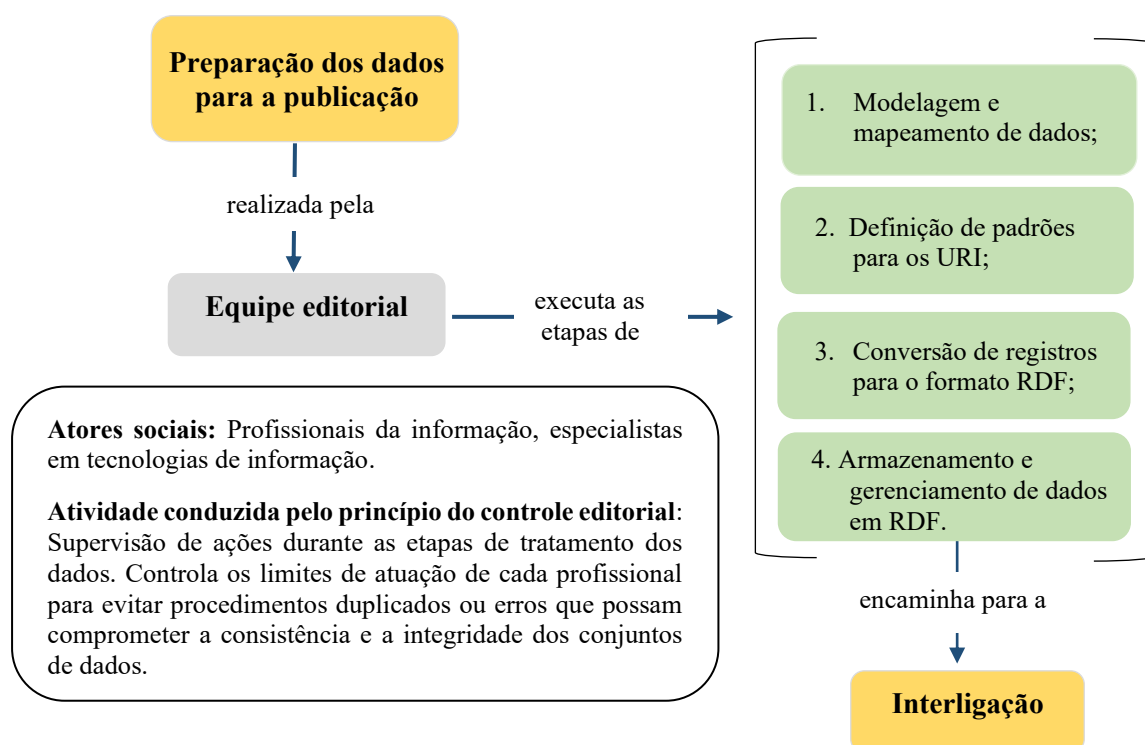
#### 5.4 PREPARAÇÃO DOS DADOS PARA A PUBLICAÇÃO

Posterior à seleção dos termos, é necessário aplicar os requisitos técnicos definidos na etapa de planejamento e preparar o tesouro para assumir modelos do *Linked Data*. A atividade de preparação dos dados para a publicação, ilustrada na Figura 70, consiste nas seguintes etapas: modelagem e mapeamento, conversão dos registros no formato RDF, definição dos padrões dos URI e armazenamento e gerenciamento dos dados RDF.

É importante frisar que pode haver tesouros já construídos e armazenados em bancos de dados relacionais, como em sistemas de gerenciamento bibliográfico. Estes sistemas não

suportam o gerenciamento de dados no formato RDF e, por essa razão, os metadados utilizados para descrever os tesouros precisam ser mapeados para os modelos de dados que apresentam sintaxe RDF. De igual modo, os desenvolvedores podem iniciar a elaboração aplicando as ferramentas de edição de tesouros que já dispõem de recursos tecnológicos para a publicação como *Linked Data*, então não haveria a necessidade de realizar o mapeamento.

**Figura 70** - Etapas da atividade de preparação dos dados para a publicação



**Fonte:** Elaborado pelo autor (2021).

Nessa etapa, todos os atores sociais envolvidos estabelecem os requisitos principais que reflitam os interesses dos usuários do tesouro, ou seja, aqueles que irão consumir os dados na *web*. Podem ocupar papéis distintos durante o tratamento dos dados, mas é preciso conhecer os limites de atuação de cada profissional antes de se iniciar os procedimentos. Todas as etapas são especificadas na sequência.

#### 5.4.1 Modelagem e mapeamento dos dados

A modelagem dos dados envolve a representação de objetos de dados e como eles são relacionados. Inclui a modelagem conceitual responsável por ilustrar uma mostra de alto nível

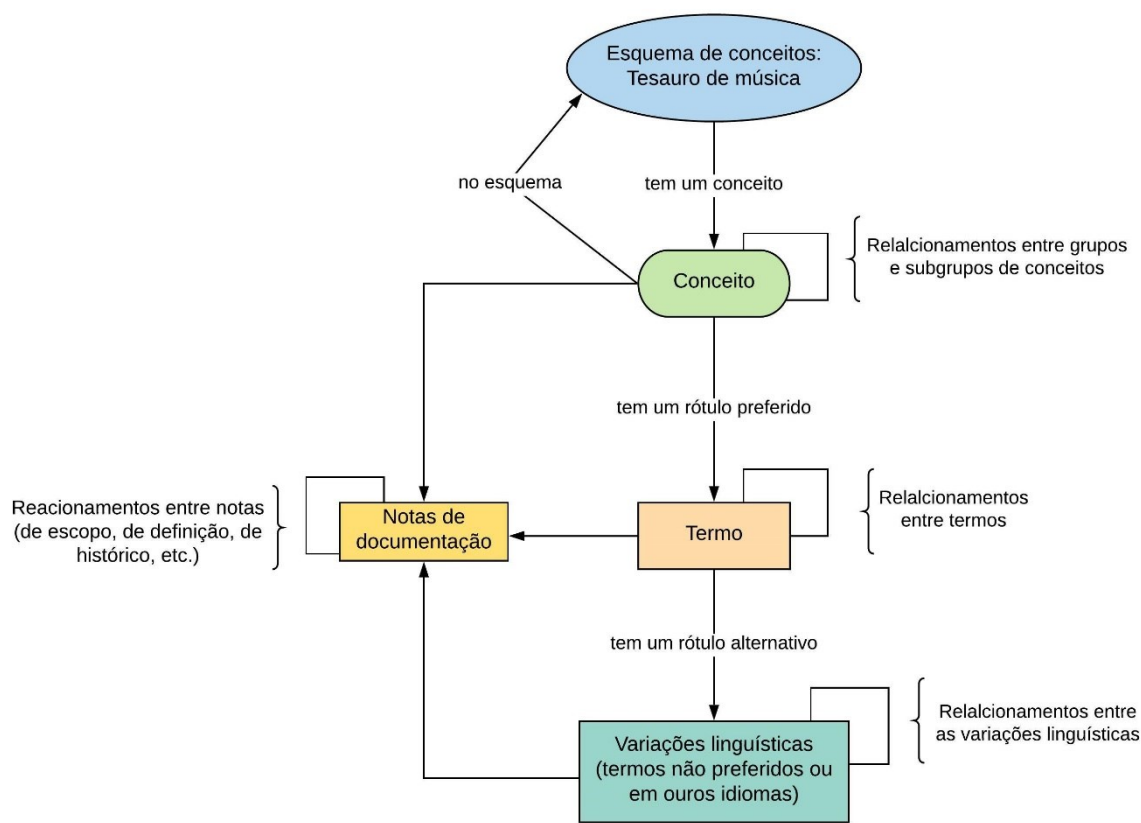
que expressa uma realidade ou o ponto de vista do criador dos dados. Ainda, é composto de um diagrama que demonstra as relações entre as entidades e atributos dessas representações (CHEN, 1976).

Conforme o W3C (2014d), a modelagem pode envolver a conversão de uma representação tabular dos dados em uma explicação baseada em gráfico. No decurso da modelagem, as partes interessadas devem descrever como os objetos estarão relacionados, usando vocabulários semânticos padronizados e reutilizáveis, pois estes auxiliam na inclusão e expansão dos dados na *web*.

Em se tratando de tesouros, o trabalho da equipe editorial, nessa fase, consiste em, após a aquisição e seleção de termos candidatos, representar o tesouro aplicando modelos de dados recomendados pelo W3C. Isso implica, também, na escolha de um sistema de gerenciamento de vocabulários que apoie todo o processo colaborativo de edição usando tecnologias e padrões do *Linked Data*.

A proposta de modelagem conceitual, ilustrada na Figura 71, está fundamentada na estruturação de disciplinas e subdisciplinas do domínio da Música. Os conceitos são distribuídos em grupos e subgrupos de conceitos, em que cada um deve ter um termo preferencial. Os termos, por sua vez, apresentam rótulos alternativos que correspondem às variações linguísticas definidas como termos não preferenciais ou em outro idioma. Conceitos, termos e variações linguísticas têm notas de documentação onde são registradas as informações sobre o escopo, definições ou histórico de alterações dos conceitos.

**Figura 71** - Visão geral da proposta de modelagem conceitual do tesouro



**Fonte:** Elaborado pelo autor (2021).

**(i) relacionamentos entre os grupos e subgrupos de conceitos:** são estabelecidas propriedades para descrever os vínculos entre grupos e subgrupos de conceitos ou entre conceitos, individualmente. Estes são reconhecidos como microtesauros do esquema de conceitos “Tesouro de Música”. Assim, um conceito pode estar vinculado a um grupo superior ou subordinado de conceitos. Como exemplo, ilustra-se o conceito Música erudita, que dispõe de duas classes de conceitos subordinados <temSubgrupo> “Música sacra” e “Música secular”. A primeira diz respeito à categoria da música de tradição religiosa, e a segunda não religiosa. Estas, no que lhe tocam, podem se desmembrar em demais classes de conceitos, cujos relacionamentos são designados por propriedades que definem o tipo de vínculo <temMembro> ou <membroDe>. As duas classes de conceitos podem ser afins à classe superior de conceitos “Música erudita” a partir de uma propriedade inversa que as designa como <temSupergrupo>.

**(ii) Relacionamentos entre os termos:** Os Conceitos, membros de uma classe de conceitos, têm rótulos preferenciais que estão relacionados a outros conceitos dentro de suas classes. Os relacionamentos hierárquicos e associativos entre termos são estabelecidos pelo uso de propriedades que designam os termos gerais <temTermoGeral>, termos específicos

<temTermoEspecífico> e termos relacionados <temTermoRelacionado> a um conceito. Por exemplo, o termo “Hinos religiosos” é considerado um termo específico da “Música sacra”.

**(iii) Relacionamentos entre as variações linguísticas:** Os termos podem apresentar variações linguísticas, expressos pelas relações de equivalência, especialmente representados por variações lexicais (como palavras com grafias diferentes que possuem o mesmo significado) ou variações semânticas (a exemplo palavras semelhantes com significados diferentes). Os relacionamentos entre as variações linguísticas podem se estabelecer pelo uso de propriedades do tipo <temTermoAlternativo>, <temFormaLiteral>, <temSinônimo> ou <temTradução>.

**(iv) Relacionamentos entre as notas:** Cada termo ou conceito pode apresentar uma nota que especifica seu escopo, definição ou histórico de atualização. Os relacionamentos entre as notas podem ser estabelecidos aplicando algumas propriedades do tipo <temTradução> (para especificar o relacionamento entre notas em dois ou mais idiomas, próprios dos tesouros multilíngues) ou <temOutraDefinição> (quando a definição do conceito está associada à outra nota de definição), por exemplo.

Para formalizar a modelagem dos dados, o W3C recomenda empregar vocabulários existentes e abertos. O *Linked Open Vocabulary*<sup>85</sup> é uma plataforma que pode auxiliar na escolha do vocabulário apropriado. Nesta pesquisa, optou-se por trazer exemplos de representação no modelo SKOS em razão do amplo uso pela comunidade do LOD. Todavia, dependendo da área, escopo e objetivos do tesouro, o esquema sob o qual este é construído pode se estender usando ontologias complementares. Com isso, recomenda-se os seguintes vocabulários abertos:

- a) SKOS<sup>86</sup>: vocabulário de propriedades lidado para a modelagem dos SOC, permitindo que estes possam ser representados e compartilhados e vinculados na *web*;
- b) SKOS-XL<sup>87</sup>: vocabulário de propriedades que define uma extensão para o *Simple Knowledge Organization System*, fornecendo suporte adicional para descrever e vincular entidades lexicais;
- c) ISO-THES<sup>88</sup>: propõe novas classes para os tipos de construções baseadas nas atualizações da norma ISO 25964 e estruturas de tesouros que não apresentam uma

<sup>85</sup> <https://lov.linkeddata.es/dataset/lov/>

<sup>86</sup> <https://www.w3.org/2009/08/skos-reference/skos.html>

<sup>87</sup> <https://www.w3.org/TR/skos-reference/skos-xl.html>

<sup>88</sup> <http://pub.tenforce.com/schemas/iso25964/skos-thes/>

equivalência direta no SKOS, tendo como tal grupos de conceitos e a inclusão de relações entre eles;

- d) UNESKOS<sup>89</sup>: pequeno conjunto de propriedades que estende o modelo SKOS e a extensão ISO-THES. Foi projetado para atender às necessidades do Tesauro da UNESCO e também funciona como um vocabulário que melhora a conexão entre os arquivos RDF;
- e) *Dublin Core*<sup>90</sup>: especificação dos termos de metadados mantidos, incluindo propriedades, esquemas de codificação de vocabulário, esquemas de codificação de sintaxe e classes;
- f) *RDF Schema*<sup>91</sup>: fornece um vocabulário para a modelagem de dados RDF. O esquema RDF é uma extensão do vocabulário básico de RDF;
- g) FOAF<sup>92</sup>: ontologia usada para a vinculação de pessoas e informações na *Web* mediante propriedades e classes nomeadas aplicando a tecnologia RDF.

Um exemplo de modelagem para tesouros empregando os modelos de dados citados anteriormente é apresentado na Figura 72<sup>93</sup>. A estrutura de dados é formada por três elementos principais: propriedades de rotulagem, propriedades de documentação e propriedades semânticas.

---

<sup>89</sup> <https://skos.um.es/TR/uneskos/>

<sup>90</sup> <https://dublincore.org/specifications/dublin-core/dcmi-terms/#>

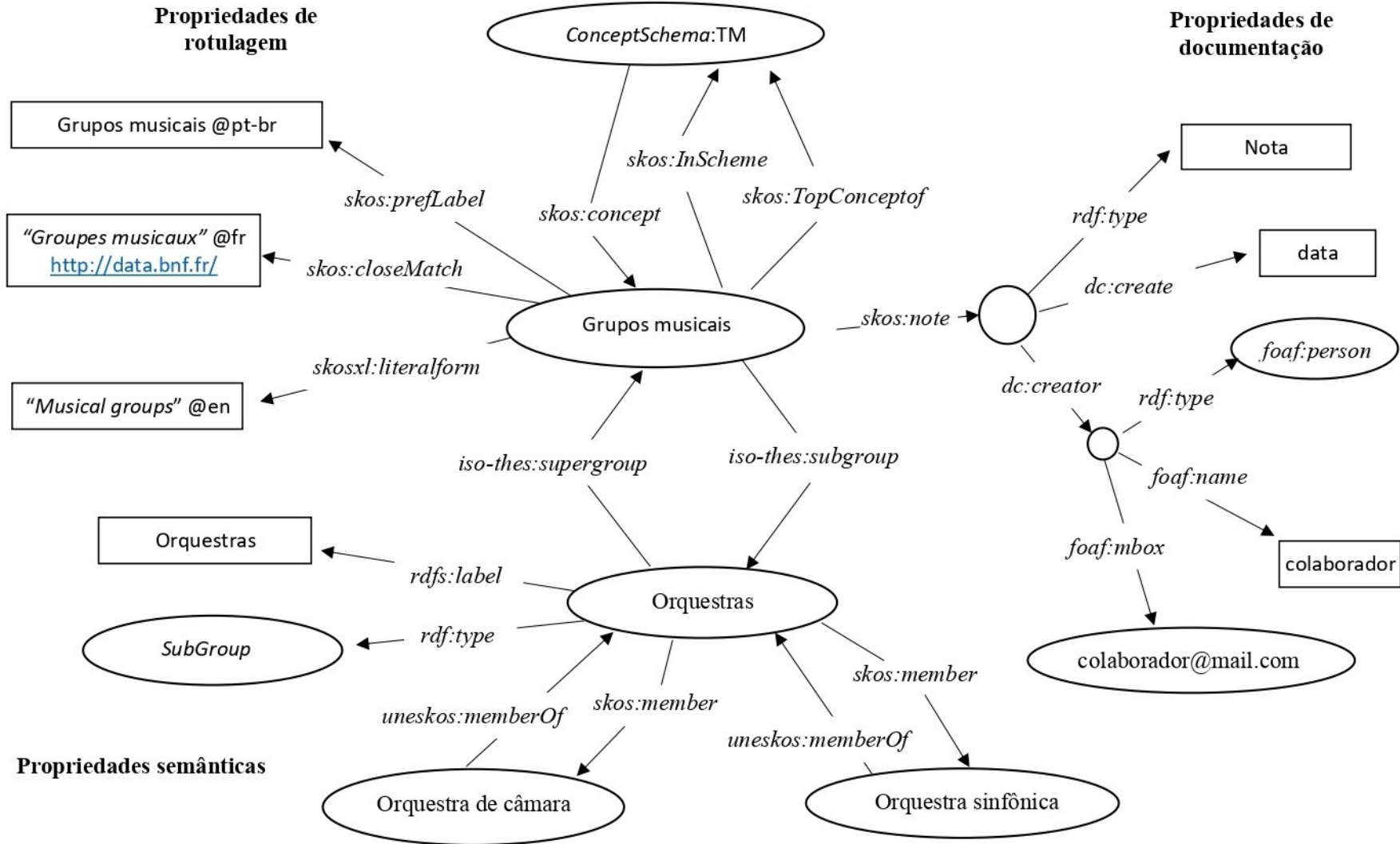
<sup>91</sup> <https://www.w3.org/TR/rdf-schema/>

<sup>92</sup> <http://xmlns.com/foaf/spec/>

<sup>93</sup> Não se pretende com o exemplo esgotar as possibilidades de uso dos vocabulários semânticos mencionados, mas sim apresentar as principais propriedades para a estruturação da macroestrutura do tesauro; em se tratando, especificamente, dos grupos e subgrupos de conceitos.



Figura 72 - Exemplo de modelagem para os dados dos tesouros usando os modelos SKOS, SKOS-XL, ISO-THES, UNESKOS, RDFS e FOAF



Fonte: Elaborado pelo autor (2021).

As propriedades da documentação são usadas para representar os metadados administrativos. A documentação é elaborada como uma descrição relacionada ao recurso que pode ser caracterizada por propriedades do Dublin Core (*dc:create* e *dc:creator*) e FOAF (*foaf:name* e *foaf:mbox*). O exemplo acima, detalha o colaborador responsável pela edição do conceito “*Grupos musicais*” (expressa pela propriedade *skos:note*). Nestas propriedades são registradas notas, datas de criação, nome e *e-mail* do colaborador.

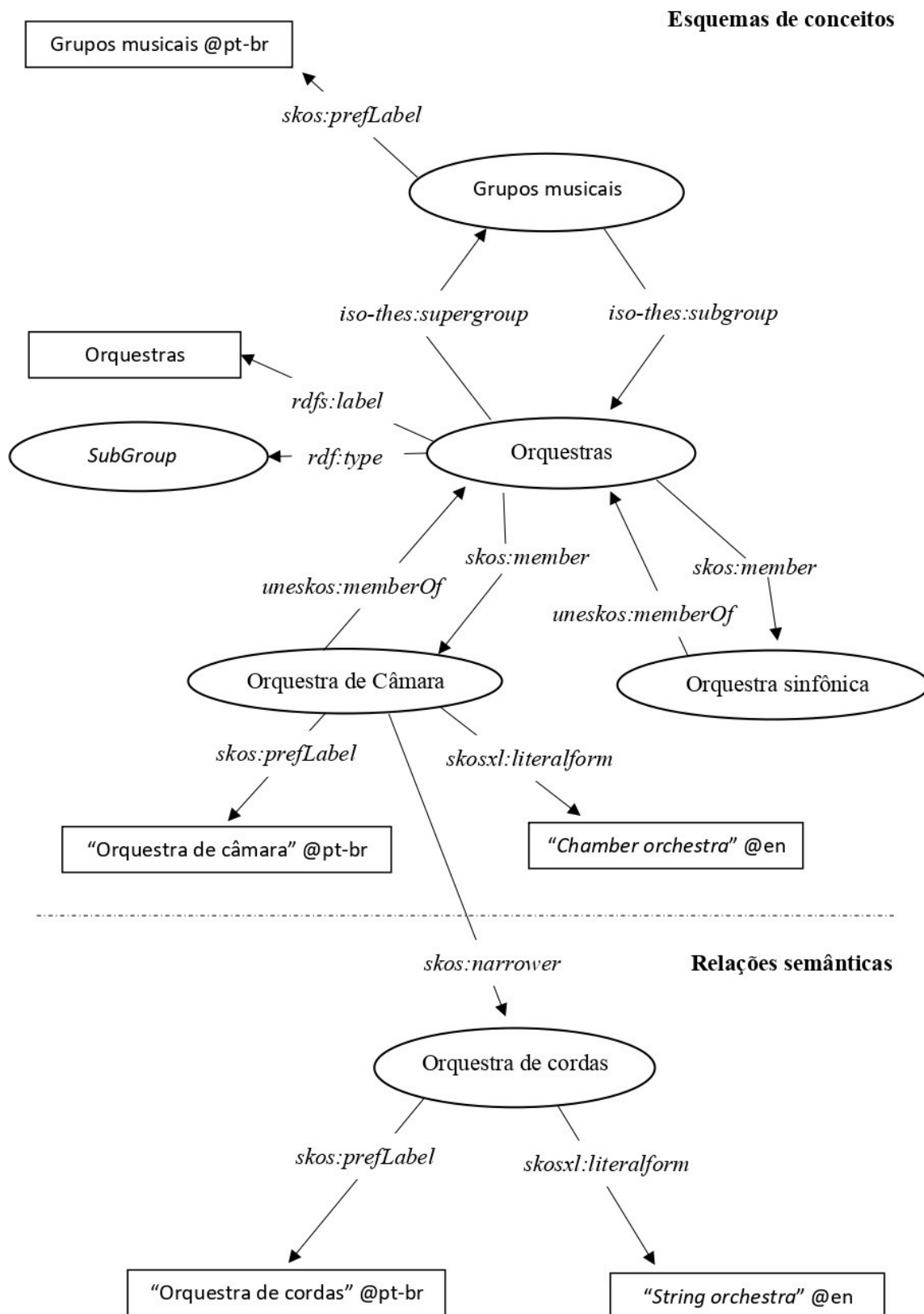
As propriedades de rotulagem são usadas para representar o nível lexical dos conceitos. Mediante propriedades *skos:prefLabel*, *skos:closeMatch* e *skosxl:literalForm* são atribuídas etiquetas léxicas preferenciais, de equivalência e alternativas ligadas ao conceito. No exemplo acima, as propriedades do SKOS-XL são operadas para rotular a forma literal dos termos em outros idiomas.

Um conceito está subordinado ao esquema de conceitos (*skos:InScheme*) do tesouro, podendo se desdobrar em subgrupos de conceitos (*iso-thes:subgroup*) e, na sequência, apresentar outras classes de conceitos subordinadas. Para facilitar o entendimento, a Figura 73 ilustra a significação dos conceitos “Orquestra de Câmara” e “Orquestra sinfônica”, os quais se apresentam como membros (*uneskos:memberOf*) da coleção de conceitos rotulada como “Orquestras”.

No nível dos relacionamentos semânticos, estão as propriedades *skos:broader*, *skos:narrower*, *skos:related*. Os conceitos mais amplos ou mais específicos podem estar associados aos conceitos de um subgrupo de conceitos (*iso-thes:subgroup*). Um exemplo de como os relacionamentos semânticos são estabelecidos é ilustrado na Figura 66, que mostra o termo “Orquestra de cordas” vinculado ao subgrupo de conceitos “Orquestra de câmara” pela propriedade *skos:narrower*, pois representa um conceito específico dentro deste esquema.

É prudente sublinhar que os conceitos organizados em grupos e subgrupos de conceitos devem ser estruturados de maneira coerente, com declarações mais amplas ou mais específicas de um conceito ou ser declarados como recursos independentes. Assim, os editores de ontologias precisam operar adequadamente as propriedades para não comprometer a estrutura lógico-semântica do tesouro.

**Figura 73** - Exemplo de modelagem de dados (esquemas conceituais e relações semânticas)



**Fonte:** Elaborado pelo autor (2021).

É preciso lembrar haver diferentes maneiras de registrar e gerenciar os dados do tesouro, que podem ser mantidos por instituições bibliotecárias, ou não, e assumirem diferentes formatos e padrões de metadados. Em se tratando de bibliotecas, os vocabulários controlados, como os tesouros, são mantidos em bases bibliográficas, compreendidos como registros de autoridade e representados por padrões de metadados amplamente difundidos pela comunidade bibliotecária, sendo o mais empregado o padrão MARC21. Todavia, este padrão é aplicado nesta abordagem para exemplificar como deve suceder a transformação de registros para o modelo SKOS.

Em consonância com as recomendações de Zapilko et al. (2013), a modificação dos dados do tesouro no formato SKOS pode ser dividido em três etapas: análise da estrutura dos dados, mapeamento e conversão. Primeiro, é efetuada uma análise da estrutura do tesouro, incluindo sua extensão e os tipos relacionamentos contidos entre termos. Isso pode ser feito a partir de uma versão XML do tesouro, capaz de fornecer uma visão completa dos dados e suas restrições. Nesta versão poderão ser visualizados os tipos de hierarquias, notas, termos preferenciais e não preferenciais. Com isso, a equipe pode optar por selecionar um pequeno conjunto de termos e realizar os testes de conversão.

Na sequência, deve-se efetivar o mapeamento dos metadados<sup>94</sup>, que auxilia na conversão dos registros mantidos nas bases de dados relacionais para os esquemas dos vocabulários semânticos escolhidos. Nesta etapa, são analisados os metadados dos tesouros para identificar as propriedades correspondentes no vocabulário SKOS. O Quadro 48 apresenta os tipos mais comuns de dados representados no MARC 21 para os registros de autoridades do tipo assunto: cabeçalhos tópicos principais ou de equivalências<sup>95</sup>, notas editoriais, notas de escopo e remissivas.

**Quadro 48** - Exemplo de mapeamento de metadados

<b>Registro de autoridades em MARC21</b>	<b>Característica/Função</b>	<b>Propriedades</b>
150 \$a	Cabeçalho tópico	<i>skos:prefLabel</i>
260 \$a	Nota de escopo	<i>skos:scopeNote</i>
450 \$a	Remissiva	<i>skos:altLabel</i> <i>skos:hideLabel</i>

<sup>94</sup> Procedimento adicional para os casos em que a instituição já possua um tesouro ou vocabulário controlado mantido em bases de autoridades ou em quaisquer outros tipos bancos de dados relacionais.

<sup>95</sup> O estabelecimento de equivalências liga posições do vocabulário controlado local com vocabulários controlados ou tesouros externos. Para os registros de autoridades em MARC21, a LC recomenda utilizar os códigos da ISO 25964-2 para especificar o tipo de relacionamento: EQ (Equivalência), BM (mapeamento mais amplo), NM (mapeamento mais restrito), RM (mapeamento relacionado).

550 \$w g \$a	Referência – Cabeçalho tópico Termo mais amplo	<i>skos:broader</i>
550 \$w h \$a	Referência – Cabeçalho tópico Termo mais restrito	<i>skos:narrower</i>
670 \$a	Fonte positiva - dados originais encontrados (citação da fonte)	<i>skos:editorialNote</i>
750 \$a \$4 EQ	Estabelecimento de Equivalência - Cabeçalhos Tópicos (EQ = equivalência exata ou aproximada)	<i>skos:closeMatch</i> <i>skos:exactMatch</i>
750 \$a \$4 BM	Estabelecimento de Equivalência - Cabeçalhos Tópicos (BM = mapeamento mais amplo)	<i>skos:broadMatch</i>
750 \$a \$4 NM	Estabelecimento de Equivalência - Cabeçalhos Tópicos (NM = mapeamento mais restrito)	<i>skos:narrowMatch</i>
750 \$a \$4 RM	Estabelecimento de Equivalência - Cabeçalhos Tópicos (RM = mapeamento relacionado)	<i>skos:relatedMatch</i>

**Fonte:** Elaborado pelo autor (2021).

A base para a transformação de um tesauro em formato SKOS cria uma análise detalhada deste, cujos registros são incorporados às ferramentas no formato de anotações e cada termo assume uma propriedade ou uma classe específica dentro do esquema de conceitos. Nesse estágio, os editores de termos revisam as anotações e geram uma primeira versão do tesauro em SKOS. Os registros podem conter recursos que também precisam ser prezados, como: data que o registro foi criado ou modificado, títulos das páginas e tipos de licenças.

Salienta-se que o vocabulário SKOS não tem propriedades para capturar essas informações e, por essa razão, outros vocabulários são empregados, como o Dublin Core (principalmente mediante classes: *dcterms:rights*; *dcterms:created*; *dcterms:modified*). Assim, é feito o mapeamento das propriedades para que na etapa seguinte esses dados sejam convertidos para arquivos do tipo RDF.

A representação de tesouros nos vocabulários semânticos, como o SKOS, por exemplo, exige cautela, dado que o processo não se reduz apenas na atribuição de propriedades aos conceitos. As etapas de mapeamento e modelagem podem tornar evidentes os problemas oriundos do processo original de construção, principalmente relacionados aos princípios da organização lógico-semântica. A título de exemplo, a designação de termos mais amplos ou mais restritos envolve a superordenação e subordinação dos tipos gênero/espécie ou todo/parte, que embora sejam diferentes são traduzidos no tesauro pelos mesmos símbolos: TG (termo geral) e TE (termo específico). Sob outra perspectiva, as relações associativas, caracterizadas

símbolo TR (termo relacionado), são usadas para indicar diferentes tipos de associação (ação/produto da ação; processo/agente; causa/efeito).

Quando os tesouros são gerados nos modelos de dados, os conceitos assumem propriedades que vão além da representação de temas mais amplos, mais restritos ou relacionados. Demandam que as conjunções ou disjunções entre as classes de conceitos sejam estabelecidas de forma a não comprometer as relações lógicas ou semânticas do vocabulário. Para os casos em que os tesouros foram originalmente construídos e mantidos em bancos de dados relacionais, deve-se proceder com a definição de padrões para os identificadores únicos (URI) e a conversão desses registros para o formato RDF.

#### 5.4.2 Padrões para os identificadores únicos (URI)

Um tesouro aberto e vinculado deve estar disponível na *web* e acessível por URI persistentes. É fundamental definir padrões para os URI, visto que estes especificam as estruturas dos documentos em RDF e suas relações com as representações em HTML. Concernente aos tesouros, padrões de URI devem ser estabelecidos para que um URI base direcione o usuário para o tesouro como um todo, e de maneira fragmentada aponte para níveis subordinados, como esquemas, conceitos e termos.

**Quadro 49** - Exemplo de padrões de URI para um tesouro como LOD

<b>Padrões</b>	<b>Descrição</b>
<b>&lt;URI&gt;</b> <http://tesouro.musica>	Informação sobre o tesouro como um todo: <b>Macroestrutura</b>
<b>&lt;URI&gt;scheme</b> http://tesouro.musica/scheme	Informação sobre um esquema de conceitos: Macroestrutura → <b>Grupos musicais</b>
<b>&lt;URI&gt;scheme/concept</b> http://tesouro.musica/scheme/concept	Lista de conceitos neste esquema de conceitos: Grupos musicais → <b>Orquestras de câmara</b> → <b>Orquestras sinfônicas</b>
<b>&lt;URI&gt;scheme/concept/A</b> http://tesouro.musica/scheme/concept/A	Informação sobre o conceito neste esquema de conceitos: → <b>Orquestras de cordas</b>

**Fonte:** Adaptado de (ZOU, 2013).

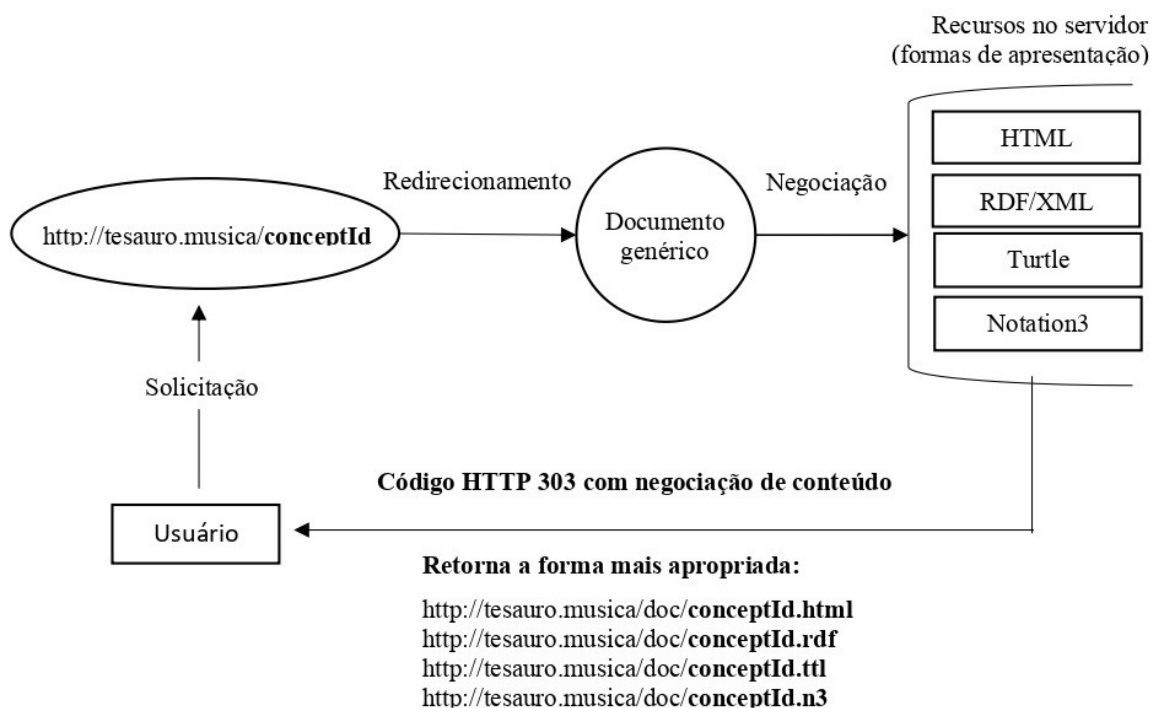
Os desenvolvedores precisam decidir como modelar os URI, de modo que os documentos na *web* sejam endereçados expressamente, permitindo o acesso e a leitura de representações HTML de um documento da *web* para navegadores tradicionais e representações

RDF para serviços habilitados da WS. Para isso, o W3C (2008c) apresenta duas estratégias: URI de *hash* e negociação de conteúdo usando o código de status HTTP 303 (consulte outro).

Na URI de *hash*, ao final do identificador um fragmento é apresentado e separado por um símbolo de *hash* (#): `http://tesouro.musica/concept#orchestra`. Isso significa que o URI não pode ser recuperado diretamente por não ser um documento na *web*. Quando ele é restaurado, o protocolo HTTP exige que a parte do fragmento seja removida e o servidor direcionado para o recurso que simboliza o objeto do mundo real. Ainda, também podem ser usadas com a negociação de conteúdo para recuperar os formatos alternativos de serialização RDF ou texto / html. Por conseguinte, a decisão é tomada com base nas preferências do cliente e na configuração do servidor.

A negociação de conteúdo usando o código 303, ilustrada na Figura 74, é aplicada para sinalizar que o recurso não é um documento regular da *web*, e, portanto, direciona uma solicitação para um documento genérico. Dessa forma, evita-se a ambiguidade entre o objeto do mundo real e o recurso que o representa. Esta configuração deve ser usada quando todas as representações alternativas transmitem o mesmo conteúdo de formas diferentes. Já os URI de *hash* são recomendados para os casos de pequenos conjuntos de recursos estáveis, ou seja, que não evoluem muito com o passar do tempo.

**Figura 74** - Uso do código 303 com negociação de conteúdo



**Fonte:** Adaptação de W3C (2008c).

Para garantir sua persistência na *web*, é recomendado acrescentar identificadores numéricos a uma URI base. Não é aconselhável utilizar termos como partes de URI, pois estes podem ser transitórios, isto é, pode haver a necessidade de modificá-los com o passar do tempo em função da atualização dos termos do tesouro. O uso de identificadores numéricos garante que os endereços permaneçam inalterados.

Os vocabulários semânticos aplicados para modelar os dados geram um grafo RDF que devem ser formalizados para obter interoperabilidade semântica com aplicativos da *web*. Assim, é preciso definir uma sintaxe de serialização específica para registrar as triplas e facilitar a leitura por humanos e máquinas. Um gráfico RDF pode ser serializado conforme diversos tipos de sintaxes. Alguns formatos, como o *N-Triples* e o *Notation 3*, são apresentados como possibilidades. A sintaxe XML facilita a integração de documentos *web*, tendo em vista que muitos são significados com a linguagem HTML/XML. Contudo, a sintaxe RDF/XML é especificada e recomendada pelo W3C para a troca de informações entre aplicativos.

O *N-triples* e o *Notation 3* fornecem uma representação mais clara de triplos RDF e podem ser facilmente compreendidos por humanos, pois consiste em uma lista explícita de sujeito, predicado e objeto. Porém, a escolha dos tipos de formatos fica a critério das equipes de construção. A Figura 75 apresenta o exemplo da serialização em RDF/XML.

É importante esclarecer que os vocabulários de propriedades, como o SKOS e SKOS-XL, operam no nível de modelagem de dados do tesouro, cujos formatos de serialização, como o XML, JSON, N-Triples e outros, lidam no nível das aplicações de informática. Logo, os dados modelados nos vocabulários de propriedades podem ser exportados em diversos formatos, sendo estes definidos conforme os padrões de saída ou de exportação dos dados.



**Figura 75** - Formato de serialização RDF/XML

```

<rdf:RDF
  xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
  xmlns:skos=" http://www.w3.org/2004/02/skos/core#"
  xmlns:skosxl=" http://www.w3.org/2008/05/skos-xl#">
  xmlns:dc=" http://purl.org/dc/elements/1.1/
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <skos:conceptSchema rdf:about="http://exemplo.com/conceptSchema">
  <skos:concept rdf:about="http://www.example.com/concepts#Orchestra">
  <skos:prefLabel> Orchestra </skos:prefLabel>
  <skosxl:literalform>Orquestra @pt-br</skosxl:prefLabel>
  <skos:closeMatch>Orchestres @fr </skos:closeMatch>
  <skos:note rdf:type= "Resource">
  <rdf:value>nota</rdf:vaule>
  <dc:creator>
  <foaf:person>
  <foaf:name>colaborador</foaf:name>
  <foaf:mbox rdf:resource= colaborador@mail.com/>
  <foaf:person>
  <dc:creator>
  <dc:date>data</dc:date>
  </skos:note>
  <skos:Concept rdf:about="http://www.example.com/concepts#Orchestra">
  <skos:prefLabel> Orchestra </skos:prefLabel>
  <skos:narrower>
  <skos:Collection>
  <rdfs:label>Tipos de orquestra</rdfs:label>
  <skos:member rdf:resource=
  "http://www.example.com/concepts#chamberorchestra"/>
  <skos:member rdf: resource=
  "http://www.example.com/concepts#symphonyorchestras"/>
  </skos:Collection>
  </skos:narrower>
  </skos:Concept>
</rdf:RDF>

```

**Fonte:** Elaborado pelo autor (2021).

Como o tesauro pode ser atualizado frequentemente, os dados não permanecem estáticos e podem ser alterados em razão das necessidades de manutenção da linguagem. É possível haver a substituição de um termo preferencial por outro dado a frequência com que é utilizado pela comunidade de usuários, ou, ainda, ocorrer a conveniência de reestruturação das hierarquias em função dos tipos de relacionamentos. Por isso, os conjuntos de dados também precisam ser melhorados periodicamente.

### 5.4.3 Conversão dos registros para o formato RDF

Os registros de tesauros podem assumir diferentes formatos de arquivos. Inclusive, podem ser extraídos na linguagem XML. Para assumir o formato RDF, os registros devem ser submetidos a uma conversão que usa técnicas específicas, tal como mecanismo de folhas de

estilo de transformação criadas com a linguagem *eXtensible Stylesheet Language for Transformation* (XSLT). A técnica usando o XSLT é indicada pelo W3C (2009a) e atestada por Harper (2006) como uma forma efetiva de conversão de cabeçalhos de assunto de registros de autoridade MARC para o SKOS/RDF, dado que os dois padrões têm uma sintaxe XML definida.

Embora muito usada, há alguns desafios relativos ao uso da linguagem XSLT, sobretudo em relação à preservação das informações armazenadas em XML. Durante a conversão, algumas informações podem ser negligenciadas devido a inexistência de propriedades correspondentes entre os padrões utilizados. Com isso, deve-se conhecer bem o conjunto de dados para que o mecanismo possa capturar um maior número de propriedades possível do registro.

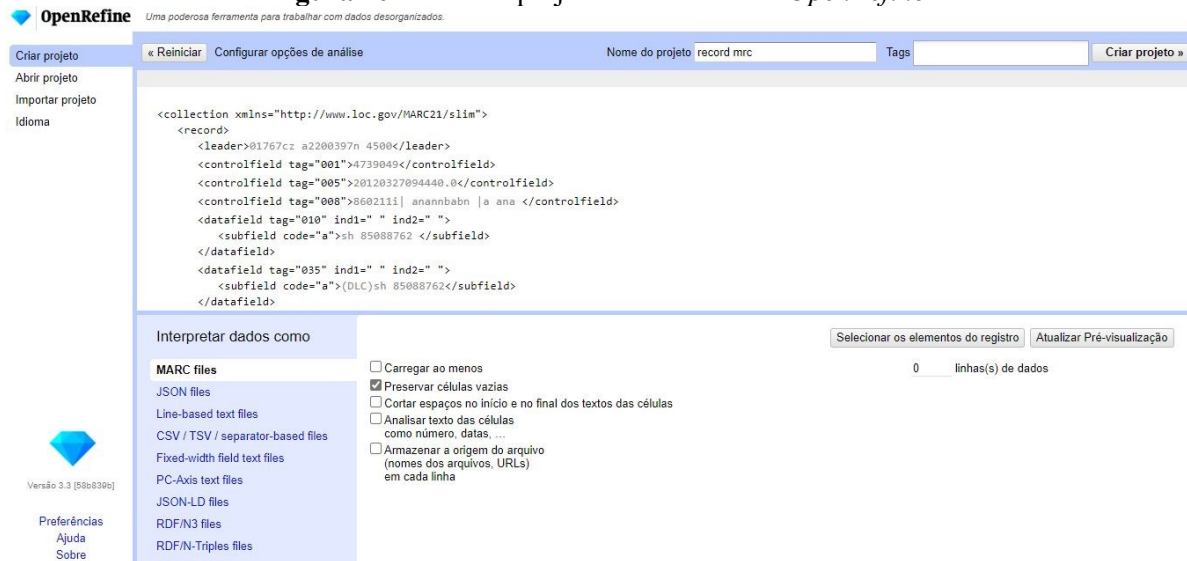
Podem haver, também, limitações tecnológicas para os casos de conversão de muitos registros, porque a transformação XSL dos arquivos é uma operação de uso intensivo quando se trata de grandes volumes de dados, provocando, ocasionalmente, lentidão ou falhas durante o processamento (TUDHOPE; BINDING, 2008). Nesse cenário, algumas ferramentas parecem mais efetivas, como o *OpenRefine* e o D2RQ Server.

O *OpenRefine*<sup>96</sup> é uma ferramenta de código aberto, de fácil uso e mais apropriada para trabalhar com uma grande quantidade de dados desestruturados. A transformação dos dados é realizada mediante linguagem *General Refine Expression Language* (GREL), cuja principal característica é a facilidade de manipular ou criar novos dados a partir de dados existentes. A ferramenta permite criar projetos por meio da importação de arquivos em diversos formatos, como: TSV, CSV, Excel (.xls e .xlsx), JSON, XML, RDF/XML e Google Data. A interpretação de arquivos do tipo MARC também é possível, como demonstra a Figura 76. Para o exemplo, um arquivo de autoridades do tipo assunto com extensão mrc foi extraído da base de autoridades da LCSH e importado para a ferramenta.

---

<sup>96</sup> <https://openrefine.org/>

**Figura 76 - Criando projetos na ferramenta *OpenRefine***

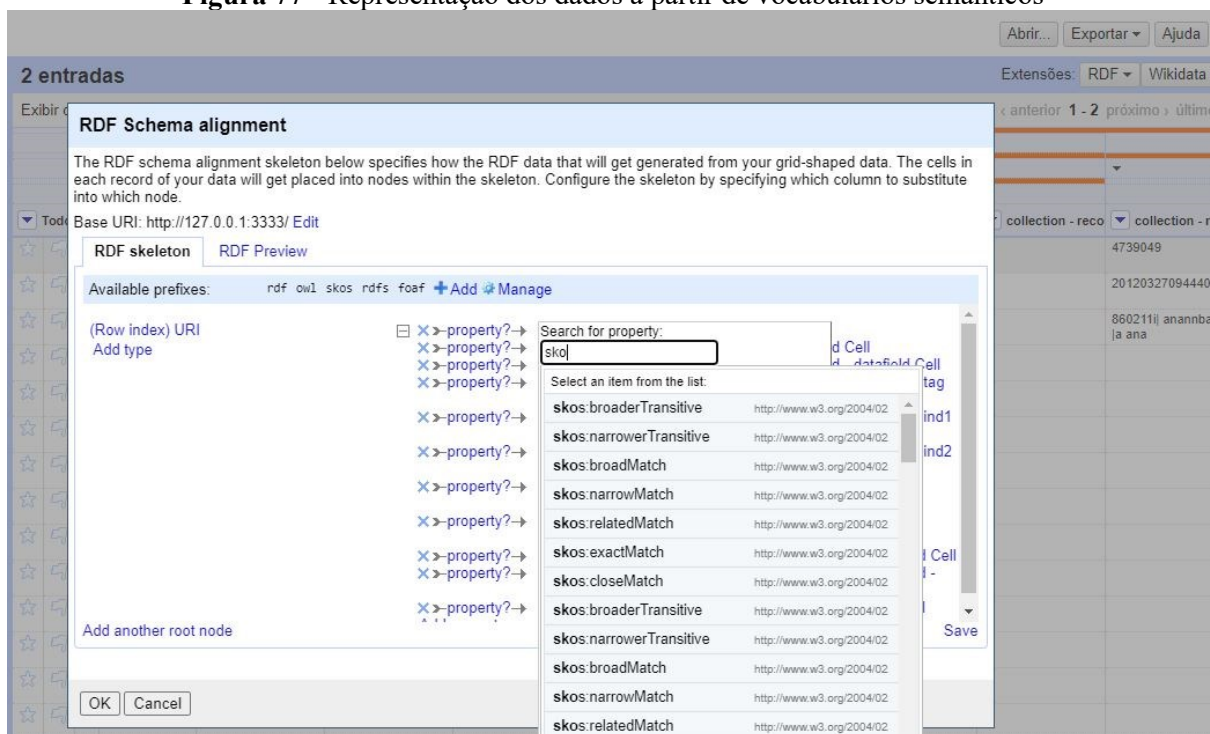


Fonte: *OpenRefine* (2020).

Os dados são importados para o aplicativo na forma de tabelas, contendo todos os elementos do registro. Os usuários podem usar facetas e filtros para selecionar subconjuntos de dados que desejam aplicar alterações. É fundamental executar limpezas nos registros, pois uma vez convertidos podem apresentar erros estruturais capazes de comprometer a integridade e dificultar sua interpretação no futuro. Nestes casos, as colunas ou linhas de dados podem ser adicionadas, movidas ou removidas para melhorar a visão dos dados.

A *OpenRefine* dispõe de uma extensão RDF, na qual os dados são representados segundo as propriedades dos vocabulários semânticos fornecidos. A Figura 77 mostra os modelos de dados disponíveis: RDF, OWL, SKOS, RDFS E FOAF. No entanto, a adição de outros modelos é permitida.

**Figura 77 - Representação dos dados a partir de vocabulários semânticos**



Fonte: *OpenRefine* (2020).

Após executar a limpeza e a representação dos dados, é possível exportá-los gerando um arquivo em RDF que é salvo no diretório de *downloads* do navegador empregado. Dessa forma, os dados podem ser lidos e interpretados por outros sistemas de gestão de vocabulários semânticos. Além disso, serem descarregados diretamente em um banco de dados de armazenamento RDF com o auxílio de ferramentas.

O D2R Server<sup>97</sup> é uma ferramenta para publicar o conteúdo de bancos de dados relacionais na *Web Semântica*. Está disponível como um servidor HTTP que permite converter dados relacionais em triplos RDF mediante linguagem de mapeamento D2RQ<sup>98</sup>. A partir desta linguagem, são criados arquivos que mapeiam cada tabela do banco de dados para as classes e propriedades do RDF. A ferramenta pode ser lidada para descrever o modelo de dados para o qual deseja converter o conteúdo – como operando as propriedades do SKOS, por exemplo –, porém, os mapeamentos entre os esquemas são realizados manualmente.

<sup>97</sup> <http://d2rq.org/d2r-server>

<sup>98</sup> <http://d2rq.org/d2rq-language>

#### 5.4.4 Armazenamento e gerenciamento dos dados RDF

Ao converter os registros do tesouro no formato RDF, surge a necessidade de aplicar sistemas de bancos de dados adequados ao seu gerenciamento. Esse formato exige o uso de ferramentas que aumentem o desempenho e a escalabilidade de soluções de armazenamento para suportar o trabalho com grandes volumes de dados e dispor de formas de consultas mais dinâmicas a esses dados.

Nos modelos de bancos de dados relacionais, os registros são apresentados como tabelas, cujos dados são organizados em linhas e colunas. Para a consulta padrão nesses bancos de dados, geralmente são dispostas linguagens declarativas, como a linguagem de consulta estruturada SQL (*Structured Query Language*). Com o advento das tecnologias da *Web Semântica* e do *Linked Data*, os registros são representados por URI persistentes e os dados integrados na *web*, o que torna essencial o uso de sistemas de gerenciamento mais eficientes a fim de garantir que aqueles possam ser consultados a partir de diferentes fontes. Dessa forma, para este cenário a linguagem de consulta mais recomendada é a SPARQL (CURÉ; BLIN, 2015).

Desde a recomendação do RDF pelo W3C, inúmeros sistemas de gerenciamento foram projetados para armazenar e recuperar dados RDF na forma de triplos que podem ser consultados com a linguagem SPARQL. Para o gerenciamento dos dados do tesouro, a literatura tem mostrado o uso das ferramentas *AllegroGraph*, *Apache Marmotta*, *Linked Media Framework*, *Mulgara Semantic Store* e *Open Link Virtuoso*. De modo geral, estas ferramentas possuem características, quais sejam:

- a) Fornecer linguagem para expressar mapeamentos do tipo *Relational Data Base* (RDB) para RDF;
- b) Além de armazenar as triplas, podem reunir dados em outras linguagens, tal como a XML. Com isso, é capaz de oferecer extensões para outros tipos de armazenamento;
- c) Viabilizar acesso persistente de grafos RDF mediante consultas federadas SPARQL, de modo a permitir que muitos bancos de dados diferentes sejam acessados e retornem resultados;
- d) Propiciar ambientes que simplifiquem a escrita de consultas SPARQL em múltiplos modelos, incluindo preenchimento automático ou auxílio aos usuários.

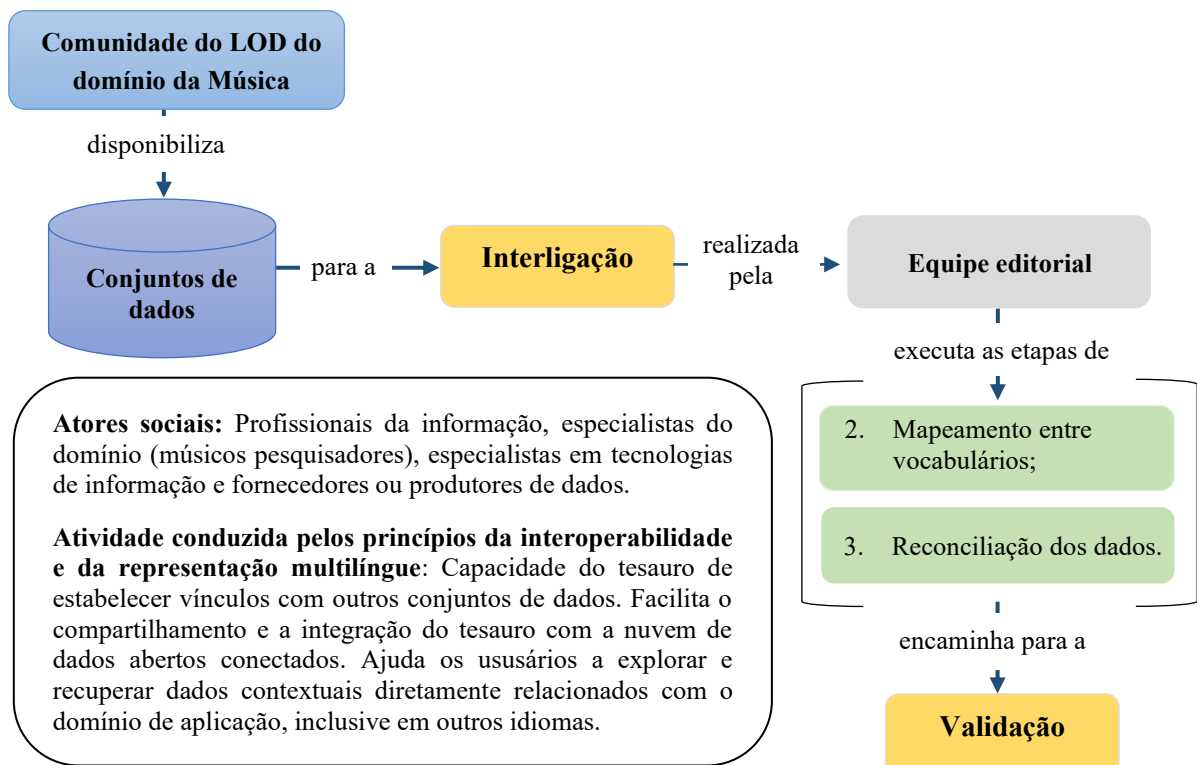
Uma vez armazenados em um banco de dados apropriados para a manipulação dos triplos RDF, devem ser estabelecidas rotinas para enriquecer e atualizar os dados. Antes de publicar

os tesouros, a equipe de construção precisa se certificar de que os dados estão conectados a outros conjuntos de dados relacionados, devendo eles serem interligados a dados de diferentes bases para ampliar os contextos de pesquisa e de consulta.

## 5.5 INTERLIGAÇÃO DOS CONJUNTOS DE DADOS

A atividade de interligação é apresentada como uma etapa de construção que coloca em evidência o princípio da interoperabilidade. Esta atividade opera procedimentos, ilustrados na Figura 78, para executar o mapeamento de termos e estabelecer os relacionamentos do tesouro com outros Sistemas de Organização do Conhecimento na *web*.

**Figura 78** - Etapas da atividade de interligação dos conjuntos de dados



**Fonte:** Elaborado pelo autor (2021).

Embora seja tratada em muitas metodologias como parte da atividade de tratamento ou preparação dos dados, optou-se por abordar tal atividade de maneira independente devido a necessidade de melhor especificação para o contexto dos tesouros.

Assim, para garantir sua ampla interligação com os demais projetos relacionados ao domínio, primeiramente é fundamental executar um mapeamento entre vocabulários. A gestão

de mapeamentos implica em identificar rótulos correspondentes em conjuntos de dados externos, sejam eles em outros tesouros, sejam eles em demais fontes de dados contextuais. As entidades semânticas relacionadas ao domínio da Música nessas fontes de dados constituem cenários para encontrar possíveis conceitos correspondentes.

### 5.5.1 Mapeamento entre vocabulários

O principal objetivo dos mapeamentos é promover a interoperabilidade semântica entre vocabulários e permitir que um conceito em um vocabulário encontre um conceito correspondente em um ou mais vocabulários de destino. A interoperabilidade é conduzida por meio de mapeamentos entre conceitos, particularmente equivalentes, exatos ou aproximados.

Os conceitos em tesouros são representados pelos termos preferidos, porém, em outros vocabulários, podem ser caracterizados como notações (em sistemas de classificação), categorias (em taxonomias), termo tópico (em esquemas de cabeçalho de assuntos), rótulos (em ontologias) ou termos (em terminologias). Como a maioria dos tesouros é representado em SKOS, os mapeamentos possivelmente são efetuados entre as propriedades *skos:prefLabel* e as propriedades do vocabulário-alvo.

Antes de executá-los é preciso decidir o tipo de modelo de gestão de mapeamentos mais adequado para o domínio de aplicação. A norma ISO:25964-2 (2013) apresenta três modelos estruturais para o mapeamento entre vocabulários:

- a) Modelo de unidade estrutural: indicado para os casos em que todos os vocabulários (de origem e de destino) compartilham a mesma estrutura de relações hierárquicas e associativas entre os conceitos. Por exemplo: mapeamento entre tesouros.
- b) Modelo vinculado diretamente: trata de vínculos entre dois ou mais vocabulários que não compartilham a mesma estrutura, os quais podem diferir em escopo, idioma e estrutura e, inclusive, englobar outros tipos de vocabulário, como lista de autoridades ou esquemas de classificação.
- c) Modelo de “Hub”: em vez de estabelecer mapeamentos diretos entre cada par de vocabulários, um vocabulário é designado como um “hub”. Em vista disso, os vocabulários de Hub retratam a estrutura central para a qual cada um dos outros vocabulários será mapeado.

Em aplicações reais, o gerenciamento pode ocorrer por meio da combinação de alguns dos modelos. No âmbito do domínio da Música, os conceitos do tesouro podem não capturar alguns elementos relacionados à percepção humana da música, classificadas como de alto nível

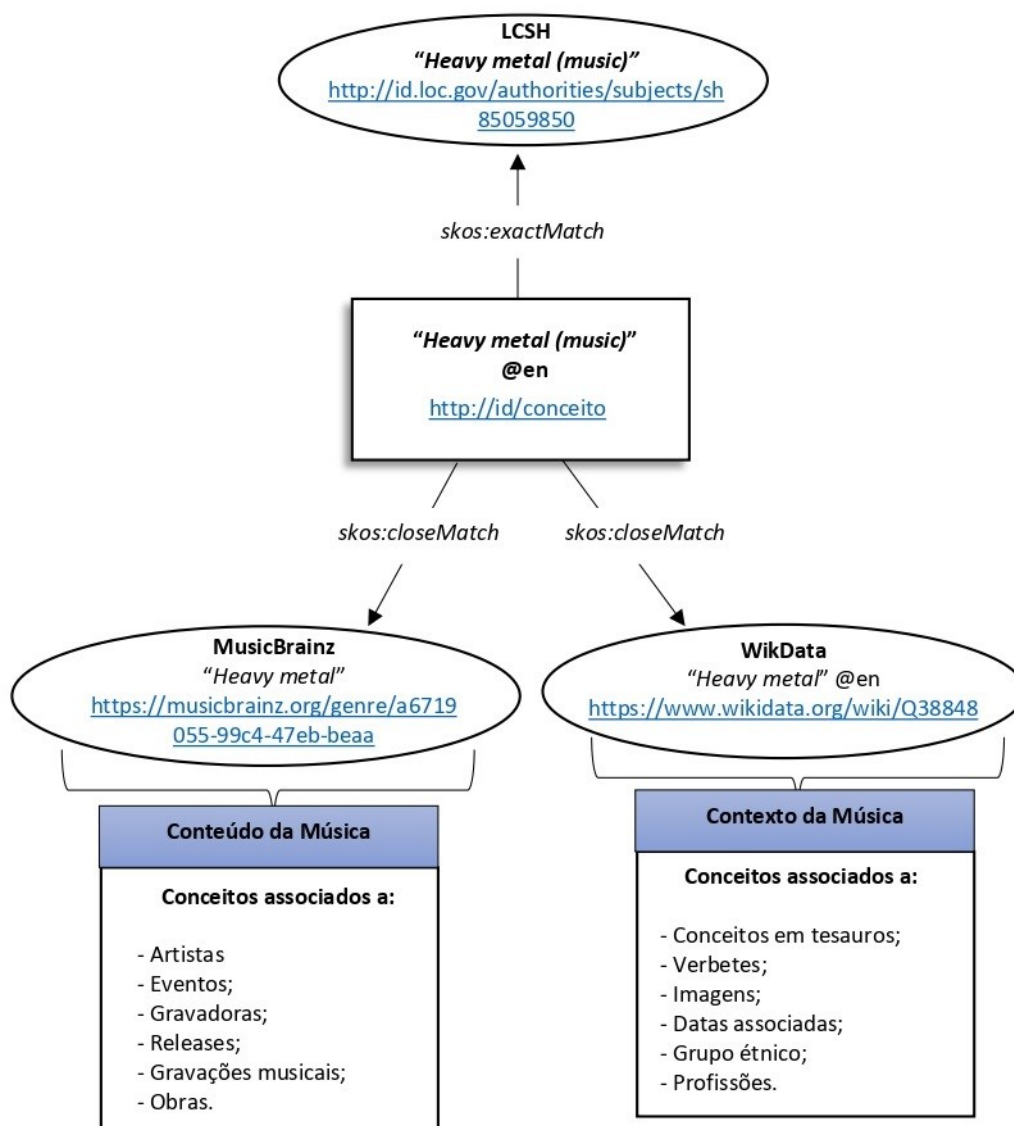
(dançabilidade, tonalidade ou timbre da música) e baixo nível (sentimento da música), uma vez que são difíceis de assumir representações conceituais em tesouros ou demais tipos de vocabulários. Os elementos do contexto do usuário e as propriedades do usuário citados por Schedl, Flexer e Urbano (2013) são naturalmente subjetivos, fruto de experiências individuais ou do nível de especialidade dos usuários.

Nesse sentido, um mapeamento de “Hub” combinado com o modelo diretamente vinculado pode ser o mais recomendado para o domínio da Música, tendo em vista que a maioria das entidades LOD relacionadas ao contexto e conteúdo da Música pode não compartilhar da mesma ordem conceitual de um tesouro. A título de exemplo, um termo preferido em um tesouro representado pela propriedade *skos:prefLabel* pode ser ligado a um conceito tópico no DBpedia (*dbo:TopicalConcept*).

A abordagem de hub permite que os mapeamentos sejam ajustados de maneira unidirecional (vinculação em única direção) com vocabulários externos, definidos como “satélites”. A vinculação pode ser determinada usando as propriedades *skos:closeMatch*, *skos:exactMatch*, *skos:narrowMatch*, *skos:broaderMatch* e *skos:relatedMatch* do SKOS. Um exemplo de mapeamento seguindo o modelo de Hub é descrito na Figura 79, que mostra como os conceitos nos tesouros podem encontrar correspondentes em rótulos semânticos de diferentes conjuntos de dados.



**Figura 79** – Modelo de Hub para o mapeamento de conceitos do tesauro com outros vocabulários



**Fonte:** Elaborado pelo autor (2021).

Nessa ilustração, o conceito "Heavy metal (music)" está associado à propriedade *skos:closeMatch*, ao conceito "Heavy metal" no WikiData e MusicBrainz. Igualmente, relacionado a conceitos aproximados ou exatos em outros tesouros na *web*, como o LCSH.

Para realizar os mapeamentos, é fundamental identificar as propriedades lidadas para representar conceitos no conjunto de dados de destino. O Quadro 50 apresenta algumas propriedades de conjuntos de dados do DPpédia, MusicBrainz e Wikidata.

**Quadro 50** - Conjuntos de dados e propriedades para a execução do mapeamento no DBPedia, Wikidata e MusicBrainz

Conjunto de dados	Propriedades ou categorias
DBPédia	dbo:TopicalConcept dbo:genre dbo:ArtisticGenre dbo:Work dbo:MusicalWork dbo:Album dbo:ArtistDiscography dbo:ClassicalMusicComposition dbo:Musical dbo:NationalAnthem dbo:Opera dbo:Single
Wikidata	category: concepts (Q151885) category: Musicology (Q164204) category: Music (Q638) category: Music genre (Q188451) category: Musical form (Q3328534) Category: Aspects of music (Q15198957)
MusicBrainz	MBID:genre

**Fonte:** Elaborado pelo autor (2021) a partir de (METABRAINZ FOUNDATION, 2019; WIKIMEDIA PROJECT, 2019; DBPEDIA ASSOCIATION, 2020).

O MusicBrainz é uma enciclopédia aberta que tem uma vasta coleção de metadados musicais na *web* representados por identificadores únicos e interligados a diversas fontes de dados relacionados ao domínio. Entre as entidades vinculadas ao gênero “*Heavy metal*” nessa enciclopédia estão, artistas, obras, lançamentos, instrumentos, eventos e *links* relacionados. Já o Wikidata é uma base de dados gratuita e colaborativa que funciona como um armazenamento central de dados estruturados de projetos parceiros de seus Wikimedia. Nele, é possível encontrar entidades vinculadas a gêneros musicais, com conceitos em tesouros, verbetes em enciclopédias eletrônicas, imagens e datas associadas, grupos étnicos e profissões associadas ao gênero musical em destaque, tal como instrumentistas.

Assim, possíveis mapeamentos recíprocos são descobertos entre os tesouros e entidades LOD externas diretamente relacionadas ao domínio. O destaque da atividade de interligação está relacionado à possibilidade do tesouro local encontrar entidades LOD externas, permitindo desempenhar o maior potencial de reconciliação com a nuvem de dados.

### 5.5.2 Reconciliação dos dados

Para estabelecer a interligação entre os conjuntos de dados não basta mapear e identificar um conceito em outro vocabulário. É preciso formalizar seu relacionamento indicando o URI HTTP, que corresponde ao conceito no vocabulário de destino. Embora seja possível estabelecer os vínculos de forma manual, a automatização desse procedimento é mais recomendada para se trabalhar com grandes volumes de dados.

A descoberta de possíveis vínculos entre vocabulários é uma etapa crítica da atividade de interligação, dado que busca gerar ligações entre itens em fontes de dados múltiplas. Em geral, esta etapa pode ser efetuada mediante ferramentas automáticas que lidam com algoritmos para calcular a similaridade de palavras e identificar correspondências candidatas, comparando valores de dados de origem com todos os rótulos conceituais do tesauro de destino. Esse tipo de solução (semi)automática é denominado de reconciliação dos dados (VAN HOOLAND et al., 2013).

Os métodos de reconciliação podem variar segundo as características das fontes de dados envolvidas. Assim, é preciso compreender como os dados estão organizados para definir os tipos de alinhamento a serem produzidos a partir deles. Concernente à reconciliação de tesouros, Isaac et al. (2008) apresentam uma abordagem que considera quatro técnicas de alinhamento:

- a) alinhamento lexical: as representações lexicais dos conceitos são comparadas. Caso alguma similaridade entre elas for identificada, o vínculo semântico é estabelecido. Pode aplicar técnicas mais complexas que consideram lematizações. Exemplo: o conceito “Balé”, no singular, concerne ao tipo de dança, e “Balés”, no plural, em alguns vocabulários pode ser empregado para aludir a obras musicais compostas para balé;
- b) alinhamento estrutural: as similaridades são identificadas com base em contextos fornecidos pela própria estrutura do vocabulário. Ferramentas podem analisar os contextos semânticos a partir dos relacionamentos hierárquicos e associativos de um conceito para identificar semelhanças. A exemplo: o conceito “Arranjo”, em um vocabulário do domínio da Música, está associado a questões de composição musical, instrumentação e orquestração. Mas, o mesmo termo pode ter sentidos diferentes em outros vocabulários, como relacionados à decoração e flores;
- c) alinhamento extensional: os contextos também são analisados, mas as ferramentas são programadas para identificar formas de termos associadas ao conceito de

origem, ou seja, o modelo estende o significado pretendido do conceito. A título de exemplo: O conceito “Ópera” pode estar associado à “Música vocal”, em um tesouro, “Coros (Música)” ou “Música dramática”, em outros tesouros;

- d) alinhamento usando bases de conhecimentos: a similaridade de um termo é identificada com base em diferentes tipos de fontes de conhecimento externas, como ontologias e redes semânticas. Esta técnica considera a relação de hipernímia (relação entre um vocabulário de sentido mais genérico e outro específico) entre dois termos correspondentes, estabelecido a partir de um vocabulário externo. Por exemplo: O conceito “Música folclórica” pode estar associado em um tesouro ao conceito “Cultura popular”, em outro tesouro, se a base de conhecimento lidada estabelecer este vínculo de relação entre os dois conceitos.

Todas as abordagens podem ser desenvolvidas, contudo a maioria das ferramentas semiautomáticas atende aos requisitos para tratar de alinhamento lexical, que é o método mais simples para o contexto dos tesouros, visto que os termos podem ser reconciliados com a lematização, computando a similaridade dos rótulos preferenciais com *strings* de outros vocabulários.

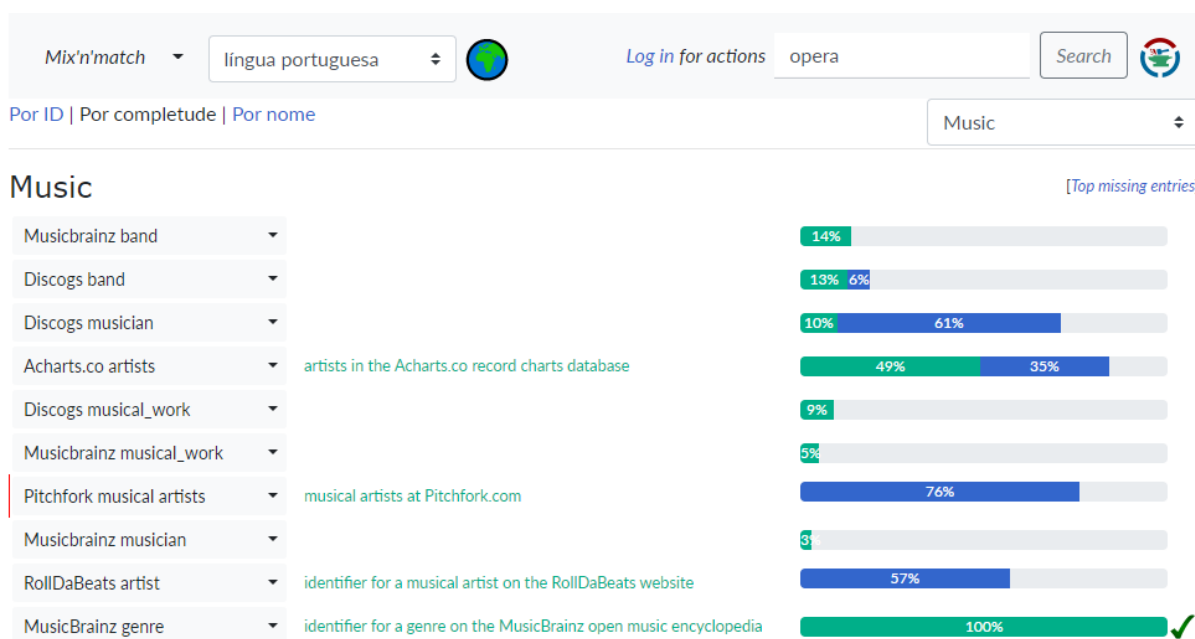
Na abordagem de alinhamento lexical e estrutural, os vínculos são determinados com as propriedades *skos:closeMatch* ou *skos:exactMatch*, cujas correspondências aproximadas ou exatas podem ser descobertas nas fontes de dados de destino. Na abordagem de alinhamento extensional ou usando a base de conhecimentos, as propriedades *skos:relatedMatch*, *skos:broaderMatch*, ou *skos:narrowMatch* podem ser mais indicadas, tendo em conta que a similaridade é calculada para identificar valores associados ao conceito na fonte de dados de origem.

Diversas ferramentas atendem aos propósitos da reconciliação dos dados, tal como o *Mix'n'match*<sup>99</sup> e o *OpenRefine*<sup>100</sup>. O *Mix'n'match* foi desenvolvido especificamente para executar a correspondência de nomes e gerar um conjunto de entidades candidatas no *Wikidata*. Através de um simples arquivo CSV com nome, descrição e URI os usuários podem importar dados, confirmar ou remover as correspondências identificadas e vincular novos dados à plataforma. No *Mix'n'match*, todos os conjuntos de dados conectados ao *Wikidata* estão disponíveis e organizados por categorias, entre elas dados em tesouros e relacionados ao domínio da Música, como ilustrado na Figura 80. A partir de sua interface *web*, os usuários podem listar e visualizar as entradas ou entidades de vocabulários que revelam vínculos com o *Wikidata*.

<sup>99</sup> <https://mix-n-match.toolforge.org/?#/>

<sup>100</sup> <https://openrefine.org/>

**Figura 80** - Conjuntos de dados do domínio da Música que possuem entidades vinculadas ao Wikidata



**Fonte:** *Mix'n'match* (2020).

O *OpenRefine* também dispõe do serviço de reconciliação, inclusive de uma extensão para vincular dados ao *Wikidata*. Por meio desta ferramenta, os usuários vinculam conjuntos de dados e removem a ambiguidade dos *links*. Ambas as ferramentas (*Mix'n'match* e *OpenRefine*) fornecem um recurso de revisão dos vínculos, contudo são recursos altamente técnicos, o que exige a atuação de profissionais bem treinados (KNOBLOCK et al., 2017).

Mesmo com as ferramentas disponíveis, a vinculação de dados ainda exige algum trabalho manual e, por isso, a total automatização não substitui completamente o conhecimento de especialistas humanos. Segundo Bryl et al. (2014), codificar todo o conhecimento e experiência de especialistas em regras para que os computadores possam entender e executar os procedimentos pode exigir muito esforço das equipes de implementação. Embora haja ferramentas apropriadas que otimizam o tempo de construção, alguma intervenção humana ainda é necessária. Logo, recomenda-se que os relacionamentos sugeridos na etapa de reconciliação sejam revisados por especialistas.

A revisão dos relacionamentos é executada sob a supervisão de especialistas tencionando verificar a qualidade dos vínculos estabelecidos entre os conceitos dos vocabulários. As ferramentas são necessárias para facilitar o trabalho intelectual destes atores e enriquecer o tesouro. Nesse estágio, sugere-se que lexicógrafos e editores de ontologias

trabalhem conjuntamente para que os relacionamentos sejam efetivos também no nível do idioma, atendendo, assim, ao princípio da representação multilíngue.

Apesar da importância do uso de ferramentas automáticas, os possíveis rótulos correspondentes, assim que identificados, precisam ser verificados pela equipe de construção. Seguindo as recomendações de Morshed et al. (2011), as etapas para avaliar uma correspondência no contexto dos tesouros são:

- (I) **verificar** se há termos não preferidos associados ao termo candidato: em alguns casos, o termo no tesouro de destino é descrito por um termo não recomendado no tesouro de origem, mas constitui uma forma variante. Por exemplo, o termo “*Heavy metal (music)*” constitui um termo preferencial no vocabulário da LCSH, porém apresenta uma correspondência na forma variada: *Heavy Rock* no *Finto Thesaurus*;
- (II) **examinar** as definições e notas de escopo relacionadas ao termo candidato: em caso de dúvidas, observar os limites semânticos de um conceito, ou seja, como ele tem sido usado no vocabulário controlado de origem. Alguns conceitos podem ter vários significados. A nota restringe o uso de um termo para apenas um significado. Exemplo, o termo “*Heavy metal (music)*”, na LCSH, é empregado para detalhar um estilo musical de artistas ou grupos de artistas quando associado a instrumentos musicais individuais. Este aparece como um qualificador: *Electric guitar music (Heavy metal)*;
- (III) **comparar** o termo correspondente com outros idiomas: Alguns termos operados no tesouro de destino em outros idiomas precisam ser traduzidos ou transliterados. É essencial encontrar palavras de sentido equivalente no idioma desejado. Por exemplo: O termo *performance*, do inglês, é apresentado como uma subdivisão de assuntos na LCSH (*Music - Performance*), porém no português o termo “Execução” é o mais recomendado (Música - Execução).

Correspondências exatas entre os vocabulários podem ser transitórias, dado que estes são construídos para atender a uma gama de aplicações diferentes. No tocante aos vocabulários multilíngues, as diferenças entre conceitos em diferentes idiomas refletem na qualidade dos mapeamentos. Portanto, deve-se estimar a necessidade de avaliar constantemente os relacionamentos com outros conjuntos de dados.

As ferramentas têm a capacidade de agilizar a atividade de interligação do tesouro à medida que rastreia diferentes tipos de fontes em vários idiomas. Devem oferecer suporte aos especialistas na definição de relacionamentos entre os conceitos mais recorrentes, reduzindo o tempo de produção do tesouro e permitindo sua produção de forma colaborativa.

A interligação com outros vocabulários marca um importante passo para consolidar a colaboração e integração com o domínio da Música, pois uma vez estabelecido os vínculos o tesouro passa a integrar uma rede semântica ampla, permitindo que os sistemas expandam sua capacidade de busca e recuperação da informação musical na *web*.

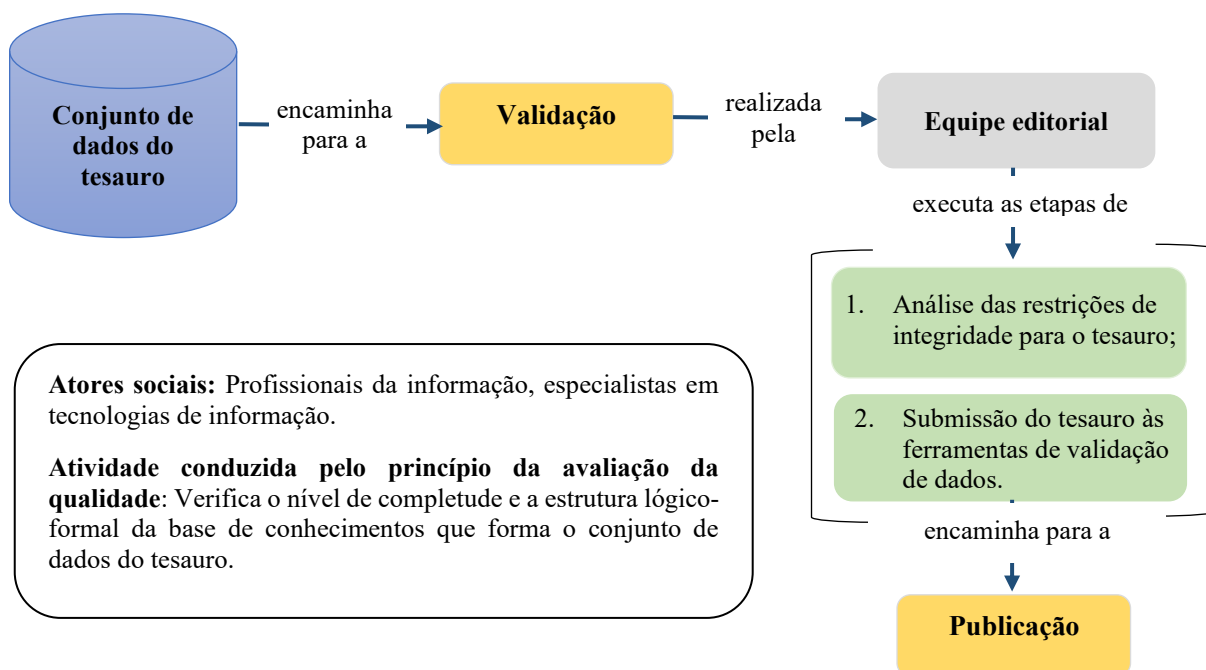
Assim, com o propósito de enriquecer as representações conceituais no tesouro e atingir o nível mínimo de representação multilíngue, a equipe de construção deve optar pelo mapeamento de conceitos correspondentes em outros tesouros ou conjuntos de dados de diversos idiomas. Ao tratar da construção de tesouros multilíngues, os profissionais precisam ter não apenas experiência com o domínio de aplicação, mas, também, conhecer bem os idiomas de destino. Conforme a ISO:2594-1 (2011), “[...] um bom entendimento das línguas naturais envolvidas ajuda na identificação de conceitos equivalentes para fins de recuperação de informações” (INTERNATIONAL STANDARD ORGANIZATION, 2011, p. 89, tradução nossa).

Um dos principais destaques para essa fase do modelo é o alinhamento de conceitos, e não apenas de termos. Isso quer dizer que as correspondências semânticas são mais relevantes que as correspondências lexicais, que figuram apenas como evidências adicionais, precisam ser identificadas e avaliadas. Ressalta-se que tal avaliação deve ser conduzida por especialistas humanos, os quais são capazes de conferir com rigor a qualidade dos relacionamentos. Após a avaliação e inclusão de todas as correspondências, ocorre a fusão de dados do tesouro com os demais conjuntos de dados relacionados.

## 5.6 VALIDAÇÃO DOS DADOS

A atividade de validação dos dados busca identificar inconsistências no conjunto de dados. As etapas descritas na Figura 81 são executadas pela equipe editorial, em especial pelos validadores humanos, que utilizam ferramentas para verificar o nível de completude e a estrutura lógico-formal da base de conhecimentos que formam o conjunto de dados do tesouro. Nessa etapa, são usadas ferramentas automáticas que estimam um conjunto de regras ou restrições de integridade para identificar inconformidades na estruturação dos dados do tesouro.

**Figura 81** - Etapas da atividade de validação dos dados



**Fonte:** Elaborado pelo autor (2021).

A manutenção dos dados pode ser viabilizada mediante um serviço de validação *web* ou pelos editores do tesouro. As ferramentas especificamente voltadas para a edição dos tesouros são capazes de produzir relatórios de erros que decorrem das atividades de construção. Quando não, os serviços de validação na *web* são projetados para realizar raciocínios automáticos sobre os dados, tendo como base a formalização das restrições de integridade para verificar se os modelos de dados usados para representar o tesouro foram aplicados corretamente.

### 5.6.1 Restrições de integridade para o tesouro

A ISO 25964 e o modelo SKOS estabelecem um conjunto de restrições de integridade para os tesouros. Enquanto a ISO especifica as condições estruturais que o tesouro deve assumir quanto à construção das hierarquias e dos relacionamentos, o modelo SKOS exhibe as regras para a estruturação dos dados em um conjunto de axiomas válidos. Esse modelo é baseado em RDF, cuja avaliação da qualidade dos dados se concentra na consistência dos esquemas RDF.

Algumas das restrições de integridade da ISO e SKOS são semelhantes, por exemplo: um conceito que não pode ter dois termos preferidos ou se apresentar como um termo mais restrito e mais amplo, ao mesmo tempo. No entanto, Martínez-González e Alvite-Díez (2019)



ressaltam haver restrições específicas aos tesouros, sobretudo concernente às relações hierárquicas e associativas que são estabelecidas de maneira equivocada.

Segundo o ISO 25964 (2011), não pode haver elementos duplicados no tesouro e conceitos órfãos. Caso a equipe opte pela exclusão de um conceito, todos os termos a ele relacionados também devem ser suprimidos. Além disso, há questões acerca dos relacionamentos associativos e hierárquicos que precisam ser observadas, como:

- a) alguns relacionamentos precisam ser estabelecidos reciprocamente. A exemplo: se “Música vocal” tem um termo específico (TE) “Ópera”, a relação inversa deve ser inferida: “Ópera” tem um termo geral (TG) “Música vocal”;
- b) se “Música vocal” é um termo geral (TG) de “Ópera”, este último não pode ser designado como termo geral (TG) de “Música vocal”, pois ciclos envolvendo relações hierárquicas são proibidos;
- c) se um relacionamento do dito TE – TG for estabelecido entre dois conceitos, não cabe um tipo de relacionamento associativo (TR) entre os mesmos.

Quanto aos problemas dos tesouros representados em SKOS, há questões de integridade e de consistência dos dados que precisam ser observadas, como: erros de rotulagem e documentação, problemas estruturais, problemas específicos de dados conectados e problemas na formalização do modelo SKOS. Para verificar se os dados estão modelados sem erros, as ferramentas de validação buscam identificar algumas das inadequações sumarizadas a seguir:

- a) **Tags de idiomas inválidos ou omitidos:** ausência de *tags* que especifiquem o idioma da linguagem ou o uso *tags* de idiomas fora dos padrões. Por exemplo:

```
<skosxl:prefLabel>Orquestra @pt-br</skosxl:prefLabel>
```

- b) **Cobertura da linguagem:** quando o tesouro tem poucos rótulos para alguns idiomas. O ideal é que para cada conceito haja um rótulo nos idiomas significados no tesouro. Exemplo: se um tesouro multilíngue cobre três idiomas (português, inglês e francês), todos os conceitos devem ser representados nesses idiomas.

```
<skos:prefLabel>Orquestra @pt-br</skos:prefLabel>
<skosxl:literalForm>Orchestras @en </skosxl:literalForm>
<skosxl:literalForm>Orchestres @fr </skosxl:literalForm>
```

- g) **Conceitos que não têm propriedades de documentação:** todos os conceitos devem dispor de propriedades úteis para documentar os significados dos conceitos; apropriado para a leitura por humanos. A exemplo:

```
<skos:prefLabel> Orquestra @pt-br </skos:prefLabel>
<skos:definition> agrupamento instrumental utilizado geralmente para a execução de música de concerto @pt-br </skos:definition>
```

- h) **Sobreposição de rótulos:** quando dois conceitos não têm o mesmo rótulo preferido, pertencem ao mesmo esquema conceitual.

Por exemplo: No esquema abaixo, o conceito “Coros (Música)” é definido como termo preferencial, enquanto é retratado como um termo alternativo para o conceito “Canto coral”.

```
<skos:prefLabel> Coros (Música) @pt-br </skos:prefLabel>

<skos:prefLabel> Canto coral @pt-br </skos:prefLabel >
<skos:altLabel> Coros (Música) @pt-br </skos:altLabel>
```

- i) **Ausência de propriedades ou propriedades sem informações:** todas as instâncias de classes do SKOS devem ser especificadas (`prefLabel`, `altLabel`, `rdfs: label`, `dc: title`) e devidamente preenchidas. Um esquema de conceito deve ser representado mediante propriedade `skos:Concept`. Este, por sua vez, deve conter um rótulo preferido. Pode-se perceber a ausência de uma forma literal (`skosxl:literalForm`), o que não é permitido nos tesouros multilíngues que lidam com a extensão SKOS-XL.

```
<skos:concept rdf:about="http://www.example.com/concepts#Orquestra">
<skos:prefLabel> Orquestra @pt-br </skos:prefLabel>
<skosxl:literalForm> -- </skosxl:literalForm>
```

- j) **Conceitos órfãos:** conceitos que não têm relações associativas ou hierárquicas. Todos devem ter, pelo menos, algum tipo de relação, conforme exemplo:

```
<skos:prefLabel> Grupos musicais @pt-br </skos:prefLabel>
<skos:narrower> Orquestras @pt-br </skos:narrower>
```

- k) **Cluster de conceitos desconectados:** todos os gráficos devem estar conectados, indicando os recursos RDF dos esquemas de conceitos mais amplos ou mais restritos nos quais os conceitos estão subordinados, conforme amostra abaixo:

```

<rdf:RDF>
  <skos:Concept rdf:about="http://exemplo.com/tesauro/concept002">
    <skos:prefLabel xml:lang="pt-br">Publicação musical </skos:prefLabel>
    <skosxl:altLabel xml:lang="fr">Édition musicale</skosxl:altLabel>
    <skos:prefLabel xml:lang="en">Music publishing </skosxl:altLabel >
    <skos:inScheme rdf:resource="http://exemplo.com/tesauro/concept000"/>
    <skos:broader rdf:resource="http://exemplo.com/tesauro/concept003"/>
    <skos:related rdf:resource="http://exemplo.com/tesauro/concept003"/>
  </skos:Concept>
</rdf:RDF>

```

- 1) **Links ausentes ou quebrados:** os recursos nos vocabulários devem apresentar uma URI desreferenciável que corresponda ao endereço único e persistente do conceito na *web*. Também devem lançar *links* para outros vocabulários, demonstrado no exemplo abaixo:

```

<skos:Concept rdf:about="http://exemplo.com/tesauro/concept001">
<skos:prefLabel xml:lang="pt-br"> Música </skos:prefLabel>
  <skos:closeMatch>
    <rdf:Description rdf:about="http://dbpedia.org/resource/Music">
      <skos:closeMatch rdf:resource=" http://exemplo.com/tesauro/concept001">
    </rdf:Description>
  </skos:closeMatch>

```

Os rótulos inconstantes ou desconexos são causados pelo uso indevido de propriedades, como o *skos:prefLabel*, *skos:altLabel* e *skos:hiddenLabel*, que são propriedades disjuntas em pares, e o *skos:conceptSchema*, que é disjunto com *skos:concept*. Esta, talvez, é uma das principais causas das inconsistências dos dados em tesauros, visto que as propriedades manipuladas incorretamente podem corromper a lógica formal do modelo SKOS. Para tentar saná-las, recomenda-se o uso de serviços de validação de dados.

No entanto, antes de reaproveitar o conjunto de dados, os projetos de implementação devem prezar por sua qualidade. Um componente de validação de dados auxilia as instituições a avaliar a qualidade da marcação empregada e analisar o nível de consistência estrutural desses recursos na *web*, que funciona como um teste de conformidade para alguns padrões. A má qualidade dos dados está, geralmente, associada à falta de disciplina ao lidar com os padrões indicados pela comunidade do LOD. Os dados conectados não devem ser publicados sem que haja qualquer validação, porque a validação é um fator crucial de usabilidade e reuso.

## 5.6.2 Serviços de validação de dados

O conjunto de regras de integridade podem ser verificados mediante serviços de validação de dados. Há serviços de validação que verificam se os dados estão em acordo com o modelo de dados RDF, e há aqueles voltados para os modelos de dados específicos, como o SKOS.

O W3C oferece um serviço de validação gratuito que pode ser aplicado para uma série de verificações de compatibilidade e integridade do tesauro no modelo RDF (Figura 82). A validação tende encontrar algumas anomalias menores, como a representação inadequada dos *triplos* ou erros de sintaxe. Essas informações devem ser repassadas ao desenvolvedor do tesauro para alimentar a manutenção de rotina.

**Figura 82** - Serviço de validação sintática para documentos RDF do W3C<sup>101</sup>

### Validation Service

[Skip Navigation](#) [Home](#)  
[Documentation](#)  
[Feedback](#)

#### Check and Visualize your RDF documents

[oldie servlet](#)

Enter a URI or paste an RDF/XML document into the text field above. A 3-tuple (triple) representation of the corresponding data model as well as an optional graphical visualization of the data model will be displayed.

Check by Direct Input

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.w3.org/">
    <dc:title>World Wide Web Consortium</dc:title>
  </rdf:Description>
</rdf:RDF>

```

Parse RDF   Restore the original example   Clear the textarea

**Display Result Options:**  
 Triples and/or Graph:   
 Graph format:

**Fonte:** (W3C, 2009b).

Na atualidade, os sistemas de construção de tesouros, livres ou comerciais, já têm componentes de validação gratuitos na *web* para verificar a consistência do tesauro no modelo de dados SKOS, como o Serviço de validação do *Poolparty*<sup>102</sup>. O W3C também dispõe um serviço de validação para o SKOS<sup>103</sup> (Figura 83). O raciocínio automático produzido por essas ferramentas materializa informações redundantes no tesauro, tal como relações mais amplas e

<sup>101</sup> <https://www.w3.org/RDF/Validator/>

<sup>102</sup> <https://qskos.poolparty.biz/login>

<sup>103</sup> <https://www.w3.org/2004/02/skos/webstage/validation>

estreitas definidas de inadequada ou quando dois conceitos distintos são representados pela mesma propriedade.

**Figura 83** - Serviço de validação SKOS do W3C

W3C Home | SWDVG Home

W3C Semantic Web Activity

SKOS: Simple knowledge organisation for the Web

SKOS Home

Introduction to SKOS

Frequently Asked Questions

Use Cases & Requirements

Specifications & Documentation

RDF Vocabularies

Find & Share Data

Software Tools & Applications

Development & Participation

Proposals & Issues

Tutorials, Presentations & Papers

>> Validation Service (alpha)

Translations

Version History

### Validation Service (alpha)

N.B. THIS WEB PAGE IS HIGHLY EXPERIMENTAL! See the [notes on using](#) below.

**Test Data**

URL:

RDF Syntax:

**Test Case**

Basic integrity test case

Thesaurus compatibility test case (includes basic integrity tests)

Try e.g.:

- <http://isegserv.itd.rl.ac.uk/cvs-public/-checkout-/skos/schemarama/data/baddatabasic.turtle> [Turtle]
- <http://isegserv.itd.rl.ac.uk/cvs-public/-checkout-/skos/schemarama/data/baddatathesaurus.turtle> [Turtle]

Fonte: (W3C, 2009b).

As ferramentas de gestão de tesouros que aplicam as regras durante a construção podem gerar relatórios de qualidade e encaminhar para os gestores do vocabulário. Nesse sentido, é importante garantir que as etapas editoriais e de validação, bem como as versões sucessivas do tesouro, sejam registradas no sistema. Esse é um pré-requisito para garantir a qualidade, pois assegura que o usuário está acessando uma versão estável e funcional do tesouro. Os mecanismos de controle e supervisão podem sinalizar quando os especialistas executam procedimentos indevidos. Igualmente, precisa compartilhar com a equipe de construção os resultados da validação a fim de minimizar erros comuns provocados pelo uso inapropriado das ferramentas.

Além de garantir a consistência dos dados durante o gerenciamento, Stellato et al. (2015) ressaltam que os sistemas de gerenciamento também devem analisar e sugerir como reparar o conteúdo automaticamente. Alguns sistemas, como o *VocBench* e o *PoolParty*, já dispõem de mecanismos de descoberta de “conceitos pendentes” que são usados para validar os conceitos representados em SKOS a partir de regras de restrição de integridade. Mediante aplicativo, os problemas são detectados e os profissionais aconselhados a proceder com os reparos.

A equipe editorial necessita conhecer bem os tipos de restrição de integridade a fim de evitá-los ao máximo. Os erros produzidos por todos os membros da equipe de construção podem ser compartilhados e discutidos entre os especialistas e, assim, impedir que situações

semelhantes possam se repetir no futuro. Este tipo de supervisão pode auxiliar na manutenção da consistência lógico-estrutural do tesouro, dado que muitos atores sociais podem estar envolvidos com a atualização, naturalmente contínua.

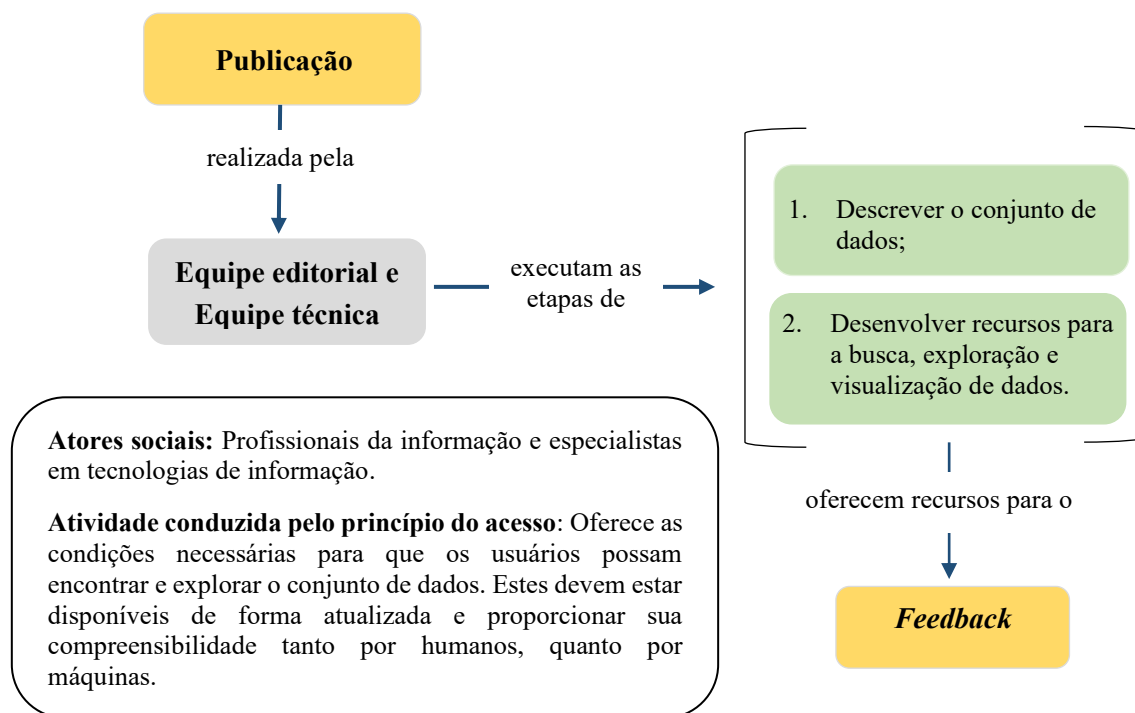
Somente posterior à reparação, os conjuntos de dados estarão prontos para serem publicados e aptos para o reuso. Não obstante, o W3C sugere que as aplicações ofereçam informações quanto à qualidade dos dados, visto que interferem na qualidade dos serviços que os utilizam. Documentar a qualidade destes, incluindo detalhes acerca de sua adequação para atender a diversas finalidades, pode facilitar a seleção dos conjuntos de dados e aumentar as chances de reaproveitamento por outras aplicações na *web*.

Indica-se que as informações referentes à qualidade estejam disponíveis para a leitura por humanos e por máquinas. Para a legibilidade desses seres, podem ser fornecidas páginas na *web* onde são descritas informações sobre sua completude (descreve a quantidade das informações disponíveis) e disponibilidade (detalha se parte ou todos os dados estão disponíveis para uso). Para a legibilidade por máquinas, podem ser operados vocabulários semânticos com propriedades para descrever informações em relação aos dados.

## 5.7 PUBLICAÇÃO

A atividade de publicação visa, através padrões e protocolos de acesso, assegurar que os dados sejam encontrados e operados pela comunidade do domínio. Para dá-los publicidade, a equipe editorial deve se concentrar nos procedimentos ilustrados na Figura 84, que consistem em: (i) fornecer informações sobre os conjuntos de dados para que humanos e máquinas possam acessá-los, compreendê-los e reutilizá-los; (ii) utilizar recursos para que os usuários possam buscar, explorar e visualizar os dados do tesouro.

**Figura 84** - Etapas da atividade de publicação



**Fonte:** Elaborado pelo autor (2021).

Na sequência, serão discutidas as etapas da atividade de publicação dos dados do tesauro que foram construídas segundo o guia de melhores práticas do W3C (2017) e tendo como referência os recursos de acesso observados na literatura analisada.

### 5.7.1 Informações sobre o conjunto de dados do tesauro

Os dados dos tesouros são aplicados em contextos distintos, como no processamento inteligente de informação, ou, de modo mais tradicional, na indexação de documentos e suporte à recuperação da informação na *web*. Dentre essas possibilidades, diversos tipos de consumidores acessam e reutilizam os dados e, por isso, suas características principais necessitam ser documentadas e disponibilizadas em endereços eletrônicos. Essas informações são explanadas por intermédio de metadados, que oferecem informações adicionais para que os usuários entendam o significado dos dados.

Nesse sentido, informações sobre licenças, versionamentos, proveniência ou origem e a qualidade dos dados podem aumentar o nível de confiança quanto à integridade e credibilidade dos dados que estão sendo compartilhados. O W3C (2017) ressalta que os metadados podem

ser dos tipos descritivos, estruturais e administrativos, devendo ser detalhados para que humanos e máquinas possam acessá-los e compreendê-los.

Os metadados descritivos expressam as características gerais dos conjuntos de dados, como título das páginas, domínio dos dados, data de publicação e publicador. Já os metadados estruturais alinham informações sobre a estrutura interna dos dados, isto é, como são formados os esquemas sob os quais os dados são organizados. E os metadados administrativos dão informações referentes às versões, qualidade e atualizações dos conjuntos de dados.

Para dar legibilidade aos dados para humanos, eles podem ser descritos como uma página da *web* ou um arquivo de texto contendo informações sobre o tesouro. Nestes recursos integram informações detalhadas sobre sua macroestrutura, lista de conceitos na forma hierárquica ou alfabética, breve histórico a respeito de como o tesouro surgiu e foi desenvolvido, entre outras informações quanto ao seu domínio ou escopo. Desse modo, permite que usuários não familiarizados com as tecnologias da *Web Semântica* possam compreender as principais características do tesouro e os limites de aplicação.

A exemplo, mostra-se a página principal do tesouro AGROVOC, ilustrado na Figura 85. Nela, pode-se observar informações sobre o tipo de dados, datas de criação e modificação e uma breve descrição a respeito do domínio de cobertura. É possível verificar, também, informações no que toca ao número de conceitos, termos e idiomas em que os conceitos estão disponíveis.

**Figura 85** - Informação sobre os dados do tesouro AGROVOC

**AGROVOC**  
ABOUT | SEARCH | RELEASES | DATA ACCESS | EDITORS | FAQ

**AGROVOC** is a controlled vocabulary covering all areas of interest of the Food and Agriculture Organization (FAO) of the United Nations, including food, nutrition, agriculture, fisheries, forestry, environment etc. It is published by FAO and edited by a community of experts. It is widely used in specialized libraries as well as digital libraries and repositories to index content and for the purpose of text mining. It is also used as a specialized tagging resource for content organization by FAO and third-party stakeholders.

**AGROVOC** consists of +37,000 concepts and +750,000 terms in up to 37 languages. Currently, AGROVOC is an SKOS-XL concept scheme, and a Linked Open Data (LOD) set edited by VocBench. **AGROVOC** is aligned with 18 other multilingual knowledge organization systems.

The latest version of AGROVOC

**How to Access AGROVOC**  
AGROVOC data can be accessed in different ways depending on the user needs:

- **ONLINE** : SKOSMOS : Search & browse
- **File for DOWNLOAD**. Two distributions are provided: (a) **CORE** , which only includes AGROVOC concepts (RDF and NT format), and (b) **Linked Open Data (LOD)** which includes AGROVOC concepts + links to **outside resources** (NT and NQ format).
- **Web Services**: SOAP Web services, accessing the RDF version of AGROVOC through the SPARQL endpoint. For further technical

Register and subscribe to receive e-mail updates

<b>TYPE</b>	Thesaurus
<b>CONTACT EMAIL</b>	agrovoc@fao.org
<b>CREATION DATE</b>	1 Jan 1980
<b>MODIFICATION DATE</b>	3 Oct 2019
<b>SUPPORTED LANGUAGE(S)</b>	Multiple Languages
<b>URL(S)</b>	<a href="http://aims.fao.org/standards/agrovoc">http://aims.fao.org/standards/agrovoc</a>
<b>DOMAIN(S)</b>	animal science and animal products, breeding and genetic improvement, farms and farming systems, fisheries and aquaculture, food

Fonte: Disponível em <http://aims.fao.org/standards/agrovoc>.



Para a legibilidade de máquinas, é recomendado fornecer acesso aos dados mediante URI dereferenciáveis, disponibilizar metadados como um arquivo RDF em um formato de serialização ou incorporá-los em uma página HTML. Além disso, devem ser usados vocabulários com propriedades para descrever o conjunto de dados, permitindo que *softwares* reconheçam facilmente os principais elementos e sua estrutura. Os metadados podem ser caracterizados por termos padrões e vocabulários de uso comum, como:

- a) *Data Catalog Vocabulary*<sup>104</sup> (VOCAB-DCAT): provê propriedades para explicar o conjunto de dados e facilitar a interoperabilidade entre catálogos de dados na *web*;
- b) *Provenance Ontology*<sup>105</sup> (PROV-O): oferece elementos para detalhar informações sobre a origem ou qualquer alteração realizada no conjunto de dados;
- c) *Dublin Core Metadata (DCMI) terms*<sup>106</sup>: fornece termos dos metadados *Dublin Core Metadata Initiative* para especificar fontes e recursos da *web*;
- d) *Vocabulary of Interlinked Datasets*<sup>107</sup> (VoID): mune termos e padrões para pormenorizar a estrutura do conjunto de dados em RDF;
- e) O *Data Quality Vocabulary*<sup>108</sup> (DQV): concede propriedades e classes para retratar a qualidade do conjunto de dados.

Nos endereços eletrônicos dos tesouros na *web* também devem constar informações sobre as licenças. Estas podem ser descritas a partir das propriedades de alguns vocabulários, como: termos do *Dublin Core* (DC): dct:license; *Creative Commons Rights Expression Language* (CCREL)<sup>109</sup>: cc:license; schema.org (SCHEMA-ORG)<sup>110</sup>: schema:license; XHTML (XHTML-VOCAB)<sup>111</sup>: xhtml:license. O uso das propriedades *Dublin Core* e *Creative Commons* para descrever as licenças em tesouros são observadas com frequência, possivelmente em razão de sua ampla aceitação em projetos do LOD, ou porque são facilmente compreendidas por humanos e igualmente legíveis por máquinas.

Para ilustrar o uso desses vocabulários no detalhamento dos conjuntos de dados, o exemplo a seguir mostra o emprego do vocabulário VoID e das propriedades do *Dublin Core* para caracterizar a estrutura do conjunto de dados do tesouro AGROVOC. Observa-se, na

<sup>104</sup> <https://www.w3.org/TR/vocab-dcat-2/>

<sup>105</sup> <https://www.w3.org/TR/prov-o/>

<sup>106</sup> <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/2006-12-18/>

<sup>107</sup> <https://www.w3.org/TR/void/>

<sup>108</sup> <https://www.w3.org/TR/vocab-dqv/>

<sup>109</sup> [https://wiki.creativecommons.org/wiki/CC\\_REL](https://wiki.creativecommons.org/wiki/CC_REL)

<sup>110</sup> <https://schema.org/>

<sup>111</sup> <https://www.w3.org/1999/xhtml/vocab>

Figura 86, a execução das propriedades do Dublin Core para a representação de metadados descritivos (*dct:created*, *dct:description*, *dct:modified*), incluindo a especificação do tipo de licença atribuída aos dados (*dct:license*). Além das informações estarem disponíveis na forma textual no endereço eletrônico do tesouro, também é incorporado a uma página em HTML no formato de serialização *Turtle*.

**Figura 86** - Fragmento da descrição legível por máquina dos dados do AGROVOC

```

@prefix :      <http://aims.fao.org/aos/agrovoc/void.ttl#> .
@prefix dct:   <http://purl.org/dc/terms/> .
@prefix dc:    <http://purl.org/dc/elements/1.1/> .
@prefix foaf:  <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos:  <http://www.w3.org/2004/02/skos/core#> .
@prefix skosxl: <http://www.w3.org/2008/05/skos-xl#> .
@prefix void:  <http://rdfs.org/ns/void#> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix lime:  <http://www.w3.org/ns/lemon/lime#> .

:Agrovoc
  rdf:type void:Dataset ;
  rdfs:label "Agrovoc Dataset"^^xsd:string ;
  foaf:homepage <http://aims.fao.org/vest-registry/vocabularies/agrovoc> ;
  dct:created "1980-01-01"^^xsd:date ;
  dct:modified "2020-07-08T09:42:42Z"^^xsd:dateTime ;
  dct:creator <http://aims.fao.org> ;
  dct:description "The AGROVOC thesaurus contains more than 37 000 concepts in 39 languages covering
  topics related to food, nutrition, agriculture, fisheries, forestry, environment and other related
  domains"^^xsd:string ;
  dct:license <https://creativecommons.org/licenses/by/3.0/igo/> ;
  dct:publisher <http://www.fao.org> ;
  dct:source <http://aims.fao.org/agrovoc> ;
  dct:subject <http://dbpedia.org/resource/Fishery> , <http://dbpedia.org/resource/Nutrition> ,
  <http://dbpedia.org/resource/Agriculture> , <http://dbpedia.org/resource/Environment> ,
  <http://dbpedia.org/resource/Food> , <http://dbpedia.org/resource/Forestry>;
  dct:title "Agrovoc"^^xsd:string ;
  void:dataDump <http://agrovoc.uniroma2.it/agrovocReleases/agrovoc_2020-07-08_core.nt.zip>;
  void:feature <http://www.w3.org/ns/formats/N-Triples> , <http://www.w3.org/ns/formats/N-Quads> ;
  void:exampleResource
    <http://aims.fao.org/aos/agrovoc/c_12332> , <http://aims.fao.org/aos/agrovoc/c_203> ,
    <http://aims.fao.org/aos/agrovoc/c_3055> , <http://aims.fao.org/aos/agrovoc/c_13551> ,
    <http://aims.fao.org/aos/agrovoc/c_6599> ;
  void:rootResource :c_330995 , :c_331000 , :c_50227 , :c_331093 , :c_7644 , :c_7778 , :c_330985 ,
  :c_330892 , :c_330991 , :c_13586 , :c_49904 , :c_330493 , :c_9001017 , :c_330705 , :c_330704 , :c_49874 ,
  :c_330834 , :c_330829 , :c_6211 , :c_330979 , :c_4788 , :c_331061 , :c_330988 , :c_330919 , :c_330998 ;
  void:sparqlEndpoint <http://agrovoc.uniroma2.it/sparql> ;

```

**Fonte:** Disponível em <http://aims.fao.org/aos/agrovoc/void.ttl>.

As especificações expostas acima constituem requisitos fundamentais para que os diferentes consumidores de dados possam compreender seu significado e estrutura. Todavia,

publicar tesouros como LOD implica também em desenvolver ferramentas e interfaces para que os dados possam ser facilmente manipulados. Como os dados dos tesouros são utilizados por usuários que, muitas vezes, não têm experiência com os padrões da *Web Semântica*, sugere-se aplicar recursos dinâmicos e interativos que ofereçam condições de busca, navegação e exploração dos dados.

### 5.7.1 Busca, exploração e visualização dos dados

Tradicionalmente, os tesouros eram projetados para que profissionais da informação treinados pudessem efetuar atividades de indexação e pesquisa. Porém, hoje, há uma demanda por vocabulários intuitivos, permitindo que usuários não treinados também possam explorá-los e utilizá-los em suas atividades de pesquisa. A intenção é incentivar toda a comunidade do domínio a aplicar conceitos do tesouro, seja para indexar seus próprios materiais de pesquisa, seja para reaproveitar os dados em outros aplicativos, ou, ainda, para encontrar dados relacionados a um tema específico, visto que os tesouros na *web* estão mapeados para outros vocabulários e fontes de dados.

Em geral, os tesouros são usados para exercer funções clássicas de recuperação mediante rede semântica de conceitos, formando uma estrutura de categorias com termos mais amplos e mais restritos. Para auxiliar os usuários na especificação de consultas e permitir que toda a comunidade encontre facilmente o que deseja, é preciso aprimorar os recursos de busca, exploração e visualização.

Concernente à busca de conceitos, os tesouros na *web* podem disponibilizar serviços de suporte à busca e recuperação de conceitos em suas plataformas de acesso. Entre as opções, estão os recursos de sugestão automática e a navegação facetada.

A sugestão automática de busca ou “autocompletar”, oferece uma lista de opções de termos assim que o usuário começa a digitar a expressão de busca. Segundo Neubert (2012), esse recurso pode ajudá-lo a recuperar rótulos preferidos de um recurso no idioma desejado. Igualmente, demonstra ser um importante meio para identificar rótulos alternativos ou ocultos ligados a um termo preferencial, pois os tesouros já têm em sua estrutura um conjunto de termos não preferenciais associados aos termos autorizados.

Na prática, se um usuário utiliza a expressão “toca-discos” e este não é considerado um conceito válido no tesouro, o sistema de busca poderá sugerir o rótulo preferencial “Fonógrafo” associado ao termo não preferencial. Assim, integrar serviços de sugestão

automática às plataformas de tesouros ajudam a identificar termos autorizados a partir de rótulos equivalentes não controlados que não são cobertos pelo tesouro, mas usados pelos usuários.

Outra maneira de facilitar as buscas é operar recursos de navegação facetada, os quais concedem que os usuários visualizem e compreendam melhor como os conceitos estão estruturados nas categoriais de assunto. Segundo Laporte et al. (2014), a navegação facetada aprimora a pesquisa semântica, pois ensina ao usuário escolher seus próprios filtros, acessando e explorando as seções ou classes de conceitos conforme interesse de busca. Usando URI HTTP, os usuários podem acessar os dados em páginas HTML, navegar e explorar os conceitos dos tesouros por interfaces convencionais de visualização de dados, como na forma de listas alfabéticas ou de estruturas hierárquica de conceitos. A Figura 87 expõe como exemplo a página da versão SKOS do UNESCO Thesaurus, que disponibiliza as formas de navegação e visualização mencionadas.

**Figura 87** - Navegação e visualização da versão SKOS do UNESCO Thesaurus

The image shows the UNESCO Thesaurus SKOS interface. At the top, there is a blue header with the 'SKOS' logo and the text 'UNESCO Thesaurus'. Below the header, there is a navigation menu with several options: 'Overview', 'Browse thesaurus', 'Alphabetical browsing', 'Hierarchical browsing', 'SPARQL Endpoint', 'Download', 'Statistics', 'Credits and legal notice', and 'UNESKOS Vocabulary'. The 'Alphabetical browsing' and 'Hierarchical browsing' options are highlighted with a red box. To the right of the navigation menu, there are language selection buttons for 'Español', 'English', 'Français', and 'Русский', and a search bar labeled 'Text to search'. Below the navigation menu, there is a list of categories with plus signs: '+ 1 Education', '+ 2 Science', '+ 3 Culture', '+ 4 Social and human sciences', '+ 5 Information and communication', '+ 6 Politics, law and economics', and '+ 7 Countries and country groupings'. At the bottom of the page, there is a contact information line: 'Contact: Juan Antonio Pastor Sánchez - pastor@um.es' and a Creative Commons license logo.

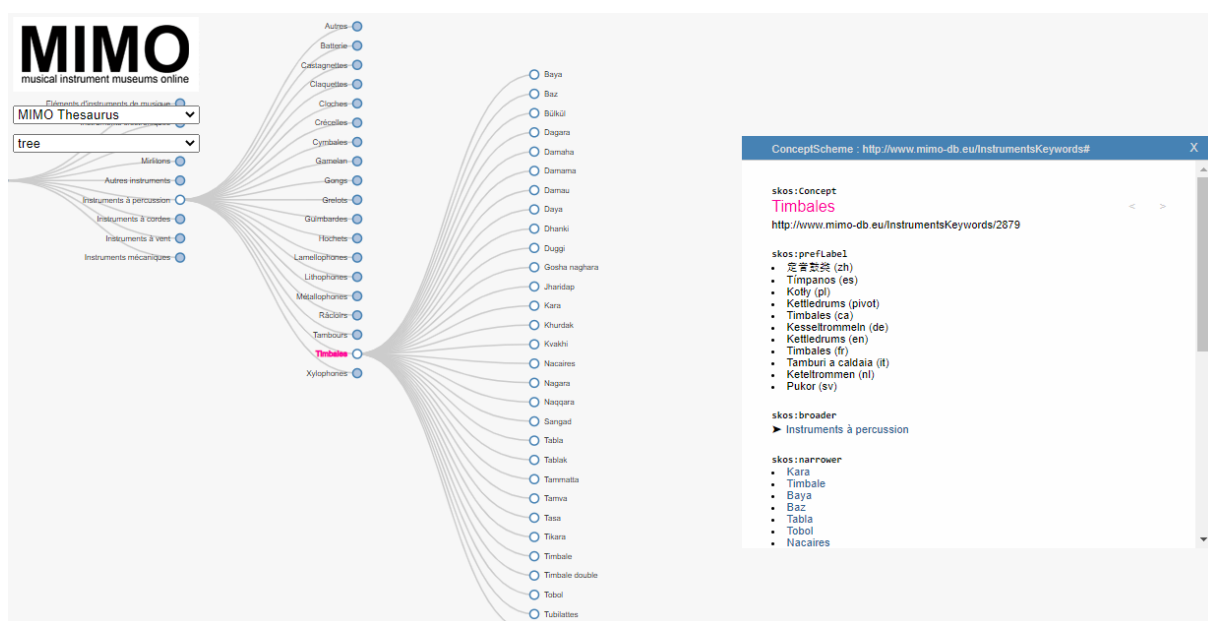
**Fonte:** Disponível em <https://skos.um.es/unescothes/view.php?fmt=1>.

Recursos interativos de visualização do tesouro ajudam os usuários a explorar e a compreender o sistema de conceitos na forma de representações gráficas. Conforme Aguilar (2013), a visualização da informação oferece uma nova abordagem para os tesouros, posto que os resultados são apresentados como uma rede formada de conceitos conectados. À medida que os usuários navegam pelas representações gráficas dessas redes de conceitos, podem visualizar

o número de conexões entre os termos, encontrar os termos mais conectados a um conceito ou localizar as conexões mais significativas entre dois nós de conceitos.

Além de melhorar a visualização dos dados, a navegação se torna mais dinâmica. No entanto, tesouros com elevado número de conceitos podem prejudicar a compreensão dos usuários em relação à organização de sua estrutura, que podem “se perder” em meio a tantas conexões. Em vista disso, os recursos de visualização devem se preocupar com a disposição gráfica dos conceitos, colocando em evidência, por exemplo, nós de superagrupamentos. A Figura 88 demonstra a diagramação do MIMO Thesaurus na forma de árvore de conceitos, destacando o termo a ser explorado.

**Figura 88** - Diagramação do MIMO Thesaurus na forma de árvore de conceitos



**Fonte:** Disponível em: <http://www.mimo-db.eu/InstrumentsKeywords/2879>.

O uso de HTML representa a forma tradicional de acesso aos dados do tesouro na *web*, em que tanto a macroestrutura como seus conceitos são caracterizados por endereços únicos (URI HTTP). Porém, o guia de recomendação das melhores práticas do W3C sugere que os dados na *web* podem ser acessados e explorados via *Application Programming Interface* (APIs) e serviços de *SPARQL Endpoint*.

Ao se tratar de acesso a grandes quantidades de dados por aplicativos na *web*, as APIs são mais recomendadas. Elas têm a função de promover o acesso aos dados abertos para que os desenvolvedores de *softwares* os reutilizem em suas aplicações. Ainda, configuram uma economia de tempo porque proporcionam acesso em tempo real a partes específicas dos dados,

geralmente distribuídos em diversos arquivos. Assim, os aplicativos dão a possibilidade de descarregar subconjuntos de dados em função do interesse dos desenvolvedores (W3C, 2017).

Em termos práticos, as APIs simplificam o desenvolvimento de aplicações, não havendo, portanto, a necessidade dos desenvolvedores conhecerem em detalhes todas as partes que compõem a estrutura de dados. No contexto dos tesouros, as APIs podem ser usadas para acessar dados dos vocabulários, tal como pesquisar conceitos que têm um rótulo em dado idioma, rótulos que contêm uma palavra específica ou dados associados aos conceitos.

De outra forma, um serviço de SPARQL *Endpoint* permite que consultas complexas sejam formuladas e os usuários recuperem o maior número possível de recursos vinculados (W3C, 2008). As consultas são executadas mediante terminal SPARQL, uma interface em que os usuários ou clientes podem consultar um conjunto de dados RDF.

As consultas são formuladas a partir de comandos ou cláusulas descritas com padrões de triplos similares aos triplos RDF, que, por sua vez, recuperam um conjunto estruturado de dados (RAUTENBERG et al., 2018). Os resultados também podem ser obtidos com a consulta simultânea a diversos servidores. Isto posto, a estratégia melhora as condições de busca, bem como a manipulação e operação sobre os dados à medida que retorna os resultados depois de processar as consultas em fontes de dados distintas.

A Figura 89 exemplifica uma consulta SPARQL no vocabulário Getty, cuja solicitação busca retornar todos os conceitos que têm um tipo particular de relacionamento associativo. Os resultados são expressos na forma de tabelas (Figura 90), onde cada linha consta os pares de termos, seus identificadores únicos e notas de definição.

**Figura 89** - Exemplo de consulta SPARQL no vocabulário Getty

The screenshot shows the 'Getty Vocabularies: LOD SPARQL Queries' interface. At the top, there are navigation tabs for 'Any', 'Search...', 'Search', and 'Brief'. Below the tabs, the 'Query:' section contains a SPARQL query:

```

1 select * {
2   ?c1 gvp:aat2100_distinguished_from ?c2. filter (str(?c1) < str(?c2))
3   ?c1 gvp:prefLabelGVP [xl:literalForm ?l1];
4     skos:scopeNote [rdf:value ?n1; dct:language gvp_lang:en].
5   ?c2 gvp:prefLabelGVP [xl:literalForm ?l2];
6     skos:scopeNote [rdf:value ?n2; dct:language gvp_lang:en]}

```

Below the query, there are checkboxes for 'Include inferred' (checked) and 'Expand results over equivalent URIs'. A 'Submit' button is located to the right.

The results section is titled '3.9 Concepts Related by Particular Associative Relation'. It includes a text explanation: 'While investigating the precise meaning of the [Associative Relationship](#) 2100 "distinguished from", we tried this query. It retrieves related concepts, their labels and scope notes (in English).'

- We use blank nodes "[.]" instead of property paths "/" because of a SPARQL parser bug (SES-2024)
- We filter by the string representation of the two concept URIs because this associative relation is symmetric, and we don't want to get it both forward and inverse in the result set.

The results table shows the following query:

```

select * {
?c1 gvp:aat2100_distinguished_from ?c2. filter (str(?c1) < str(?c2))
?c1 gvp:prefLabelGVP [xl:literalForm ?l1];
  skos:scopeNote [rdf:value ?n1; dct:language gvp_lang:en].
?c2 gvp:prefLabelGVP [xl:literalForm ?l2];
  skos:scopeNote [rdf:value ?n2; dct:language gvp_lang:en]}

```

A note at the bottom states: 'We don't need to filter by skos:inScheme because gvp:aat2100\_distinguished\_from can link only AAT concepts, nothing else.'

Fonte: Disponível em:

[http://vocab.getty.edu/queries#Concepts\\_Related\\_by\\_Particular\\_Associative\\_Relation](http://vocab.getty.edu/queries#Concepts_Related_by_Particular_Associative_Relation).

**Figura 90 - Resultados da consulta SPARQL no vocabulário Getty**

c1	c2	l1	n1	l2	n2
aat:300195328	aat:300301373	bread boxes (food containers)@en	Airtight or nearly airtight, box-like containers, of metal, wood, or plastic; used for storing bread and other baked goods to keep them fresh.@en	bread chests (working tables)@en	Chests or tables of various designs used in kitchens or bakeries, having storage areas and a working surface. Used for the storage of flour, making of bread on the table top, and often for the storage of bread. For smaller, counter-top containers used to store bread, use "bread boxes."@en
aat:300412076	aat:300419420	Mbps (megabits per second)@en	Measurement for internet connection speeds; preceded by a numerical value, denotes the download and upload speeds of digital data passing through a data transmission system, one megabit being equal to one million bits.@en	MBps (megabytes per second)@en	Measurement for how much of a file is downloaded/uploaded per second. Distinguished from megabits per second, which refers to Internet connection speeds.@en
aat:300004909	aat:300432119	sheep barns@en	Barns constructed or used for the housing of sheep, usually having a pen or pens large enough to contain several sheep or a flock, smaller pens for lambing (giving birth), and areas for shearing, feeding, and watering.@en	sheep sheds@en	Sheds to shelter sheep.@en
aat:300311318	aat:300311320	carpintería de lo prieto@en	A type of carpentry in old Spain. In contrast to 'carpentería de lo blanco,' carpentry classified as 'lo prieto' is typically not decorative. It may also constitute pieces that are heavier, rough-hewn, and irregularly shaped.@en	carpintería de lo blanco@en	Lacework carpentry associated with Mudéjar Spain and Portugal, and as such often called Mudéjar carpentry. Geometric and decorative, this type of structural woodwork is often seen in framing, paneling, and roofing.@en
aat:300412131	aat:300419421	Plumbicons (TM)@en	Proprietary name for a type of television camera tube similar to a vidicon, but in which the photoconductive layer of the signal plate is of lead monoxide; noted for	vidicons@en	Small television camera tubes of a particular kind, in which the image is formed on a transparent electrode coated with photoconductive material. the

**Fonte:** Disponível em:

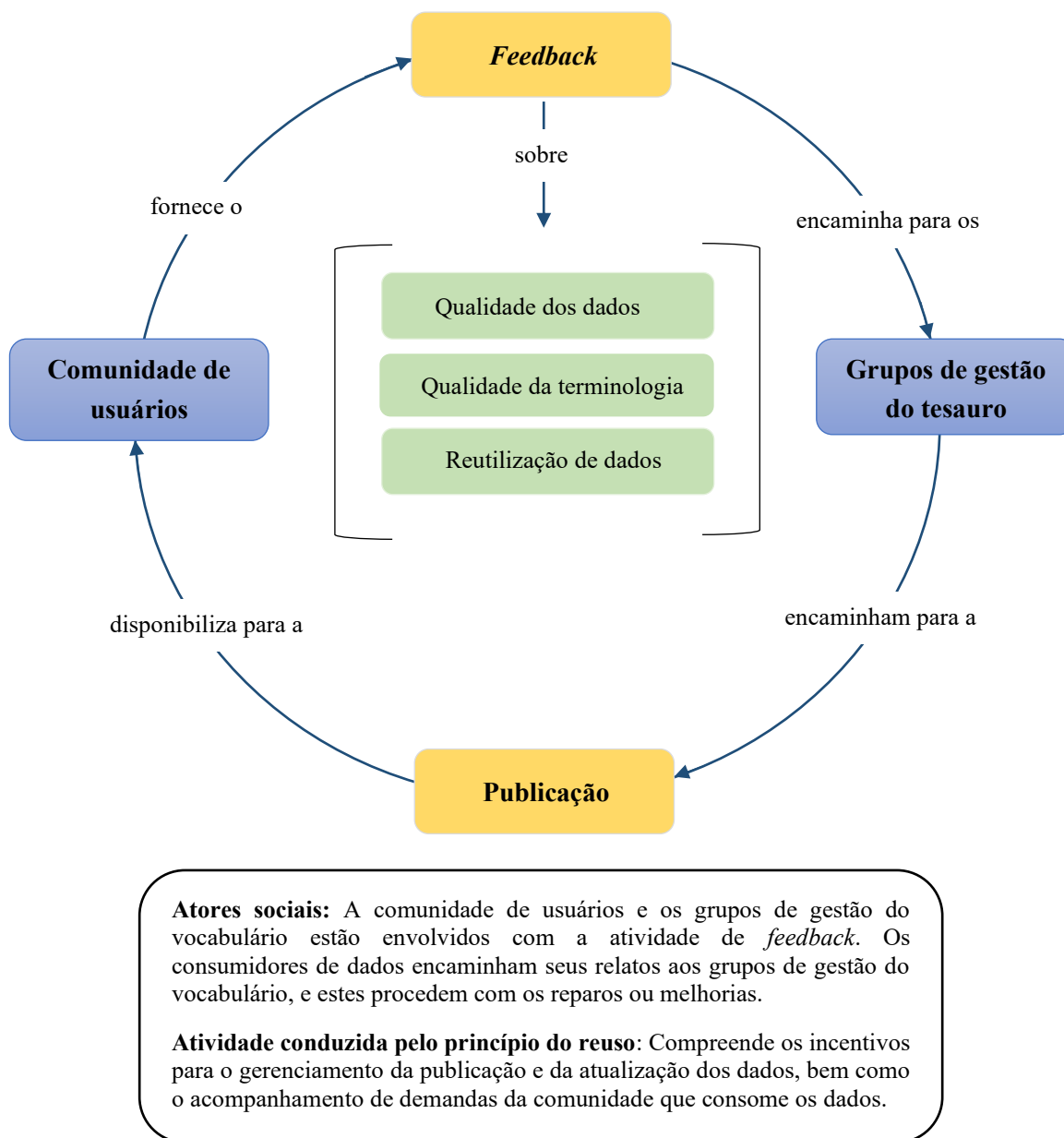
[http://vocab.getty.edu/queries#Concepts\\_Related\\_by\\_Particular\\_Associative\\_Relation](http://vocab.getty.edu/queries#Concepts_Related_by_Particular_Associative_Relation).

O número crescente de serviços de consulta SPARQL oferece aos consumidores de dados uma oportunidade de mesclá-los distribuídos pela *web*. Há uma variedade de sistemas atualmente que oferecem suporte para o desenvolvimento de serviços de consultas SPARQL, assim como o SPARQLer, D2RQ e Open Link Virtuoso (W3C, 2019).

As interfaces para explorar o conjunto de dados devem ser dinâmicas e permitir não apenas que os usuários busquem ou visualizem os conceitos do tesouro de formas alternativas, mas também disponham de canais de comunicação para que a equipe de construção receba o *feedback* dos consumidores dos dados. Esses espaços precisam dar suporte à colaboração social, de modo a operar ferramentas que ajudem a conduzir a avaliação da qualidade dos dados, assim como da terminologia.

## 5.8 FEEDBACK

Ao publicar dados no formato aberto, os tesouros poderão ser reutilizados por diferentes tipos de usuários. Todavia, para melhor compreender as necessidades dos consumidores, os editores de dados precisam disponibilizar canais de comunicação para obter *feedbacks*. A Figura 91 ilustra como ocorre a atividade.

Figura 91 - Atividade de *Feedback*

**Fonte:** Elaborado pelo autor (2021).

O *feedback* traz benefícios tanto para os usuários do tesauro como para seus editores. À medida em que aqueles descrevem suas experiências com os dados, estes podem identificar deficiências na construção e publicação dos dados e trabalhar para implementarem melhorias.

Alguns mecanismos de *feedback* são recomendados pelo guia de melhores práticas, como a disponibilização de formulários ou caixas de comentário para que os utilizadores detalhem suas experiências ou encaminhem sugestões de melhoria. Ou, ainda, sistemas de classificação em que sinalizem a qualidade dos dados a partir de pontuações. Estes são recursos



essencialmente colaborativos, pois as melhorias no conjunto de dados são produzidas com base nas contribuições da comunidade.

No contexto dos tesouros, além de auxiliar a melhorar a integridade dos dados ao comunicar suas experiências, os usuários podem ajudar a aprimorar a qualidade da terminologia. Portanto, os editores também devem dispor de mecanismos para que estes indivíduos sinalizem se o tesouro está atendendo às necessidades de busca e recuperação da informação na *web*. Dessas sugestões, podem surgir conceitos muito utilizados pela comunidade, porém pouco observados na literatura. Assim, os *feedbacks* são meios que podem auxiliar as equipes de construção de tesouros a executar a garantia do utilizador.

Desse modo, receber *feedbacks* sobre a pertinência de alguns conceitos do tesouro exige uma etapa posterior de análise das sugestões feitas pelo usuário, visto que pode ensejar a consulta de especialistas do domínio ou até promover discussões com os membros da equipe editorial. Nesse aspecto, Mcculloch (2005) ressalta que a depender da natureza do *feedback*, pode ser difícil realizar uma reunião presencial, e, portanto, alguns recursos colaborativos mais eficientes podem ser empregados, como o *e-mail* ou um fórum de discussão.

Após recebê-los, é recomendado que os editores compartilhem as avaliações com a comunidade de usuários. De acordo com o W3C (2017), essa é uma forma de demonstrar que as sugestões de melhoria estão sendo acolhidas ou de evitar que observações semelhantes possam ser enviadas novamente à equipe de construção. Uma vez que os *feedbacks* estão documentados e compartilhados, os consumidores conseguem identificar previamente quaisquer problemas que podem afetar a capacidade de reaproveitamento dos dados.

Para dar legibilidade por humanos, os *feedbacks* podem ser apresentados como parte de uma página na *web*. A ISO 25864-1 (2011) indica documentar as atualizações efetuadas no tesouro na forma de listas com os termos em ordem alfabética ou de maneira categorizada, mostrando os termos preferidos que foram adicionados ou substituídos. Para facilitar a compreensão dos usuários, podem ser fornecidos detalhes de cada termo e relacionamento, novo ou alterado. Assim, no sentido de dar destaque as alterações, os endereços eletrônicos na *web* são encorajados a incorporar dispositivos como rastreamento ou realce de novos termos, chamando a atenção dos usuários para as mudanças no tesouro.

Para dar legibilidade por máquinas, os editores de dados podem fazer uso de vocabulários padrões, tal como o *Dataset Usage Vocabulary*<sup>112</sup> (VOCAB-DUV). Este último dispõe de propriedades e classes para documentar formalmente as motivações do usuário, os

---

<sup>112</sup> <https://www.w3.org/TR/vocab-duv/>

critérios que o levou a ter uma opinião sobre os dados e os comentários associados ao *feedback* desse ator social. O (W3C, 2016b) ressalta que usar metadados para coletar os *feedbacks* pode ajudar no desenvolvimento de APIs, que fornecem acesso às opiniões e às experiências do consumidor dos dados de forma estruturada, permitindo que os resultados sejam facilmente compartilhados entre os aplicativos da *web*.

É possível observar o caráter cíclico, colaborativo e contínuo da construção dos tesouros como LOD. Sendo a última atividade do modelo, o *feedback* é considerado o componente de retroalimentação do sistema. A partir dele, os editores conseguem mensurar o alcance dos dados junto ao público e a efetividade prática do tesouro no que toca às tarefas de indexação ou recuperação da informação. Todas as interações são acompanhadas pela equipe editorial na dimensão dos grupos de gestão, que recebem a avaliação da comunidade de usuários. Uma vez identificadas as necessidades de melhorias, essa equipe deve se atentar para a possibilidade de rever as atividades de implementação, como as etapas de aquisição, os critérios de seleção de termos e a modelagem de dados do tesouro.

## 6 AVALIAÇÃO DO MODELO COM ESPECIALISTAS

Com o objetivo de atestar a confiabilidade e a consistência da proposta submeteu-se o modelo a avaliação de especialistas da área. A intenção foi verificar se o modelo traduz a realidade do fenômeno pesquisado, tendo como pressuposto de que “a avaliação da qualidade de um estudo científico é uma estratégia que promove a credibilidade de seus resultados, sendo basicamente composta de validade e confiabilidade” (SOUZA; SILVA, 2011).

Isto posto, os especialistas foram convidados a realizar a leitura do modelo, incluindo um breve relato sobre os achados da pesquisa bibliográfica e de campo. Assim, alguns questionamentos foram elaborados e encaminhados a aqueles tentando auxiliar na elaboração de um parecer (Cf. Apêndice E). Ao final deste, os colaboradores deveriam informar a aprovação do modelo, pontuar debilidades e sugerir melhorias.

O primeiro ponto questionado foi a clareza das relações estabelecidas entre as dimensões de colaboração, os princípios norteadores e as atividades de construção (Quadro 51). Esse ponto buscou verificar se a proposta contempla todos os aspectos da construção de tesouros, considerando os fluxos de tarefas e o papel de cada ator social envolvido. Quanto a isso, a maioria dos especialistas avaliou positivamente:

**Quadro 51** – Clareza na descrição e relação entre as dimensões, princípios e atividades de construção

<b>Especialistas</b>	<b>O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?</b>
Especialista 1	<i>“[...] como não consegui distinguir os papéis de cada equipe de gestão (ou editorial), em alguns momentos não ficou claro para mim as relações entre as três dimensões (colaboração, os princípios norteadores e as atividades de construção).”</i>
Especialista 2	<i>“Sim, o modelo proposto é bem completo neste aspecto.”</i>
Especialista 3	<i>“O modelo descreve, de maneira clara e didática, as relações entre as dimensões de colaboração, os princípios norteadores e o fluxo de atividades de construção e validação do tesouro. Tal proposta encontra-se ancorada nas diretrizes que norteiam o processo de construção e publicação de tesouros no contexto do LOD.”</i>
Especialista 4	<i>“O modelo proposto descreve com clareza as relações entre as dimensões de colaboração, os princípios que norteiam a publicação de tesouros em Linked Open Data e as etapas necessárias à construção desses tesouros. Os atores sociais e os recursos envolvidos nas práticas de colaboração estão suficientemente indicados na proposta. Os princípios norteadores estão bem sintetizados e encontram respaldo na literatura internacional. No decorrer da apresentação da proposta, a menção às práticas e aos princípios é recorrente.”</i>
Especialista 5	<i>“As relações são bem descritas e representadas. Há no trabalho apresentado uma riqueza de detalhes com a apresentação de figuras, fluxogramas e suficiência descritiva em cada etapa do processo.”</i>

Especialista 6	<i>“O modelo proposto descreve com clareza e leva a compressão das relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção, evidenciando como é fundamental a existência e funcionamento de cada uma das partes envolvidas.”</i>
----------------	--

**Fonte:** Dados da pesquisa (2020).

Em seguida, indagou-se os avaliadores a respeito da estruturação das etapas consolidadas no modelo e se elas atendiam às necessidades de construção de tesouros. Buscou-se com essa ação identificar se todos os procedimentos e aspectos metodológicos elencados em cada atividade estão organizados de maneira lógica. Ainda, sob o ponto de vista operacional, se as atividades e tecnologias apresentadas no modelo traduzem a realidade para a publicação de tesouros no contexto do *Linked Open Data*. As respostas seguem apontadas no quadro a seguir.

**Quadro 52** – Estruturação das etapas e conformidade às práticas do LOD

Especialistas	<b>As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do <i>Linked Open Data</i>?</b>
Especialista 1	<i>“[...] considero que o modelo não atende a todas as etapas para a construção de tesouros, ou não de uma maneira lógica que, na minha concepção, deveria ser estabelecido. Penso que o melhor para que as etapas ficassem alinhadas ao contexto do <i>Linked Open Data</i> seria o atendimento à lógica de etapas de construção recomendada pela norma ISO 25964, pois, sendo uma norma internacional, se todos os tesouros adotassem a mesma lógica de construção, teoricamente, seria mais promissor a conseguir o propósito de interoperabilidade semântica.”</i>
Especialista 2	<i>“Sim. O modelo é bastante completo e abrangente, considerando diversos aspectos da construção de um tesouro no contexto <i>Linked Open Data</i>.”</i>
Especialista 3	<i>“Considera-se que as etapas contempladas no modelo foram estruturadas de maneira lógica, atendendo as necessidades de construção de tesouros nesse contexto. Essas etapas estão vinculadas a um conjunto de atividades colaborativas que objetivam a cooperação de sujeitos, denominados atores sociais, no processo de representação da informação musical.”</i>
Especialista 4	<i>“As etapas para a construção e publicação do tesouro estão dispostas de maneira lógica, sendo os resultados de cada etapa frequentemente evidenciados como insumo para a etapa seguinte. Embora se destinem a uma abordagem mais teórica e generalista dos processos, provendo conhecimentos amplos acerca dos agentes, ferramentas e dados envolvidos – uma vez que se trata de um modelo, e não de um manual – essas etapas apresentam os elementos necessários até mesmo para subsidiar projetos práticos, no nível operacional, para a criação e publicação de tesouros como <i>Linked Open Data</i>.”</i>
Especialista 5	<i>“O conjunto de atividades indicadas “Planejamento”, “Aquisição”, “Seleção”, “Preparação dos dados”, “Interligação”, “Validação”, “Publicação”, “Feedback”, são bastante pertinentes e estão organizados de maneira lógica. As</i>

	<i>etapas de “Preparação dos dados”, “Interligação dos Dados” e “Validação dos Dados” contextualizam o modelo no Linked Open Data.”</i>
Especialista 6	<i>“As etapas propostas foram estruturadas de maneira lógica, pois seguem uma ordem de desenvolvimento do fluxo de trabalho que atendem as demandas exigidas para a criação de tesouro de maneira geral. No contexto do Linked Open Data elas atendem as necessidades de construção de tesouros, visto que estão sua ordenação e descrição estão voltadas para facilitar que a organização da informação funcione em um ambiente web.”</i>

**Fonte:** Dados da pesquisa (2020).

No que se refere às observações do Especialista 1, ressalta-se haver na literatura diversas abordagens teóricas e metodológicas voltadas à construção de tesouros, cujo modelo foi elaborado com o propósito de orientar os grupos de gestão, inclusive em relação à seleção e aplicação dos diversos métodos. Os métodos podem variar conforme a aplicação do vocabulário, o público a que se destina, os recursos e fontes disponíveis. Diante disso, procurou-se confluir algumas abordagens de construção e empregar ao contexto dos dados conectados na *web*, aspecto este pouco evidenciado nas instruções normativas, exceto pelas discussões em torno da necessidade de tornar os vocabulários interoperáveis.

Acerca das nuances da terminologia musical e a necessidade de atender ao público heterogêneo, os especialistas foram solicitados a discorrer com relação à possibilidade de o modelo atender às necessidades do domínio de aplicação. De outra forma, sob o ponto de vista da generalização, os avaliadores se posicionaram quanto à capacidade de aplicação e adaptação do modelo em outros contextos ou domínios. Os resultados seguem expressos no Quadro 53.

**Quadro 53** – Aplicação no domínio da música, em outros contextos ou domínios

<b>Especialistas</b>	<b>De modo geral, o modelo atende as necessidades do domínio de aplicação? O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?</b>
Especialista 1	<i>“[...] posso comentar que no elemento tradução, uma vez que a proposta é ser uma representação multilíngue, com interoperabilidade, percebe-se uma limitação, em especial por denominá-la de “interoperabilidade semântica”, sendo que parece que irá levar em consideração apenas o aspecto lexical. E mesmo sendo assim, deve-se pensar em distintas questões que envolvem a equivalência entre diferentes idiomas. No âmbito da música, a representação de “escala”, por exemplo, varia conforme o contexto cultural no qual está inserido. Outro exemplo seria a denominação de instrumentos de percussão latino-americanos, que têm variação de designações, dependendo do lugar. Como o seu modelo envolve a construção do tesouro, e não somente o ambiente colaborativo, a questão da tradução, neste exemplo, teria que ser melhor definida.”</i>
Especialista 2	<i>“Sim, em parte [...] Sua complexidade obrigada que sejam feitas adaptações para que ele possa ser adaptado em projetos e contexto específicos.”</i>
Especialista 3	<i>“Em linhas gerais, depreende-se que o modelo proposto atende satisfatoriamente as necessidades do domínio de aplicação, podendo servir como referência para</i>

	<i>ser adaptado/aperfeiçoado para fins de implementação em outros contextos ou domínios.”</i>
Especialista 4	<i>“O modelo foi proposto para um domínio específico, o da Música. Os conceitos e os termos deste domínio são a matéria-prima para o tesouro e os atores sociais do domínio fazem parte de diversas etapas do modelo, contribuindo nos processos de sua competência. No decorrer da apresentação das etapas, nota-se uma preocupação em trazer elementos do domínio da Música para a discussão e/ou exemplificação dos processos, atores, instrumentos, etc., com isso, a proposta busca demonstrar que atende as necessidades do domínio. [...] o modelo pode ser aplicado em outros contextos ou domínios, seja em sua totalidade, nos casos em que ainda não há um tesouro, ou parcialmente, nos casos em que a pretensão é publicar como Linked Open Data um tesouro já existente.”</i>
Especialista 5	<i>“Em termos teóricos o modelo apresentado é consistente e pode atender o domínio da Música, desde que a etapa de planejamento, com a definição do público alvo, nível de especialidade, fontes de referência e escolha dos conjuntos de interligação, seja bem delimitado [...] O modelo tem enorme potencial para aplicação em outros contextos e domínios.”</i>
Especialista 6	<i>“Acreditamos que o instrumento proposto atende as necessidades do domínio de aplicação, visto que as fontes consultadas para a coleta de termos são especializadas no domínio e também o pesquisador buscou apoio das instituições e profissionais que trabalham no domínio. Esses passos favoreceram o enriquecimento de termos e atividades para a elaboração da proposta. O instrumento pode ser adaptado ao outro domínio, desde que ao se fazer as adaptações os profissionais sigam os mesmos passos para a elaboração de sua proposta, como a consulta as fontes e instituições ligadas ao domínio no qual queira fazer a adaptação.”</i>

**Fonte:** Dados da pesquisa (2020).

Em relação à validade da proposta, o Quadro 54 mostra que cinco dos especialistas aprovaram o modelo e apenas um se absteve. Entre os que aprovaram a proposta foram consideradas, principalmente, sua completude ou capacidade de implementação. Quanto ao que se absteve, apesar de afirmar que o modelo pode atender ao que se propõe considerou que precisa ser testado.

**Quadro 54** – Resultado da avaliação dos especialistas

<b>Especialistas</b>	<b>Resultado da avaliação</b>
Especialista 1	<i>“Penso que não é uma questão de aprovar ou desaprovar o modelo, que, acredito, <b>pode atender ao que se propõe</b>, mas que seja necessário ajustar alguns elementos para que ele possa ser testado e, então, aprovado ou não, ou ser refinado.”</i>
Especialista 2	<i>“<b>Aprovamos o modelo</b>. A validade do mesmo reside na sua completeza, ressalvados os aspectos mencionados acima. Seria algo como um modelo de referência.”</i>
Especialista 3	<i>“Em linhas gerais, reitera-se que a pesquisa doutoral apresenta uma proposta inovadora e construtivista para a Organização e Representação do Conhecimento, bem como para unidades e sistemas de recuperação da</i>

	<i>informação no domínio da música. Diante do exposto, considero o modelo proposto “Aprovado”.</i>
Especialista 4	<i>“considerando os pontos destacados, aprovo o modelo [...]”</i>
Especialista 5	<i>“O modelo apresentado é uma enorme contribuição para a área, e o considero Aprovado.”</i>
Especialista 6	<i>“Por fim, aprovo o modelo proposto, visto que fica claro a forma como foi construído e enquanto produto final para o domínio da música dentro do Linked Data, pode suprir as necessidades existentes de organização e recuperação da informação.”</i>

**Fonte:** Dados da pesquisa (2020).

Contudo, alguns especialistas aduziram primordialidade de ajustes, como:

- a) Ressaltar a importância do planejamento como etapa de construção do tesouro;
- b) Considerar o elemento tradução para que o tesouro possa atender de maneira efetiva ao princípio da representação multilíngue e da interoperabilidade;
- c) Abordar a essencialidade de desenvolver políticas de gestão, atualização e revisão dos tesouros ao longo do tempo, bem como aspectos da governança dos projetos.
- d) Versar acerca da perspectiva de reuso de outros vocabulários já existentes;
- e) Dar ênfase à caracterização do público-alvo, visto haver outros tipos de usuários no domínio da Música como graduandos em instrumentos, estudantes de conservatórios e de escolas de música com ênfase prática ou teórica;

Todas as sugestões foram consideradas e os ajustes realizados no modelo a fim de refiná-lo. Portanto, considera-se a proposta válida à medida que: retrata as principais etapas de construção de tesouros no âmbito do LOD; demonstra como os atores sociais estão articulados em uma rede de colaboração; e ressalta as nuances do domínio de aplicação, em especial, a participação dos usuários e a comunidade do LOD do domínio da Música, sendo esta última representada pelos projetos que mantêm os conjuntos de dados relacionados ao conteúdo e contexto da Música.

Destaca-se, também, que o modelo tem potencial de ser aplicado ou adaptado a outros contextos. Para isso, a etapa de planejamento deve analisar as características do público-alvo, as fontes de referência ou vocabulários já existentes e os aspectos sociolinguísticos do domínio que se pretende representar.

## 7 CONSIDERAÇÕES FINAIS

Esta pesquisa demonstrou a capacidade de integração de diversos atores sociais envolvidos com a construção e a publicação de tesouros. Foram apresentados modelos de dados, padrões e tecnologias aplicados ao contexto do LOD. Assim, admite-se que ao explicitar as relações existentes entre esses atores, o uso de tecnologias e de como estes estão articulados em uma rede de colaboração, infere-se que o estudo atendeu ao objetivo geral proposto, visto colocar em evidência o caráter cíclico e cooperativo de elaboração de tesouros, revelando sua importância para a organização e recuperação da informação musical na *web* de dados conectados.

Tendo em conta o domínio da Música como área de aplicação percebeu-se o quão complexo é construir vocabulários controlados que atendam a demandas de informação de públicos diversificados. Os termos selecionados para representar o sistema de conceitos no tesouro podem apresentar variações tanto no campo lexical como no campo semântico. Sua constituição está condicionada às características do público-alvo e aos aspectos socioculturais e linguísticos de cada país ou região. Em vista disso, faz-se necessário dar importância às múltiplas aplicações da música, com realce para a pesquisa acadêmica, o processo de ensino-aprendizagem, a prática instrumental e a vida cotidiana.

Por conseguinte, ao avaliar esses elementos, o tesouro poderá contribuir de maneira efetiva com a Organização da Informação Musical, fazendo com que os sistemas de busca e recuperação atendam às necessidades de informação de usuários especialistas e não especialistas simultaneamente.

Mediante a proposta da tese, acredita-se que os objetivos específicos foram alcançados à medida que foi possível identificar práticas colaborativas na construção de vocabulários controlados do domínio da Música, estruturar os princípios que norteiam a implantação e publicação de tesouros como LOD, construir o modelo e avaliá-lo com os especialistas.

Com base nos relatos das instituições que participaram da pesquisa, especificamente bibliotecas públicas e universitárias especializadas em Música no Brasil, identificou-se os principais atores sociais, os recursos empregados durante todas as etapas de construção e as organizações que contribuem, mesmo que de maneira indireta, com a elaboração dos vocabulários controlados. Nesses relatos percebeu-se como esses elementos estão articulados dentro de uma rede de colaboração, da qual participam bibliotecários, especialistas do domínio, profissionais de TI e a comunidade de usuários. Constatou-se, ainda, que a seleção e



estruturação dos termos estão fundamentados conforme cinco garantias: a garantia literária, a garantia terminológica, a garantia estrutural, a garantia cultural e a garantia de uso.

A partir da amostra teórica (análise da literatura) estruturou-se um conjunto de princípios que norteiam a construção e publicação de tesouros no contexto do LOD e funcionam como o núcleo do modelo colaborativo proposto. Ao sobrepor os princípios junto às dimensões de colaboração e seus respectivos papéis durante a composição pode-se verificar que cada dimensão é regida por aquele conjunto, em que cada princípio abrange um agrupamento de atividades essenciais. Essa evidência permitiu a construção do diagrama cujos principais conceitos do modelo foram descritos.

Em relação aos princípios, acredita-se que o consenso, a representação multilíngue, o controle editorial e a avaliação da qualidade estão associadas às competências dos grupos de gestão e, portanto, diretamente relacionadas às atividades de construção, enquanto que os princípios de acesso, reuso e interoperabilidade apontam para o desenvolvimento de aplicações e estratégias relativas às atividades de publicação do tesouro.

A representação multilíngue se destacou entre os princípios por cumprir importante papel em mais de uma dimensão, sendo elas a dimensão dos grupos de gestão e a dimensão das comunidades do LOD do domínio da Música. No que concerne aos grupos de gestão, a representação multilíngue está associada à representação dos conceitos em dois ou mais idiomas. Quanto às comunidades do LOD do domínio da Música são consideradas fundamentais na etapa de reconciliação dos dados. Em ambos os casos, enfatiza-se a pertinência de estimar o nível de semelhança do termo do idioma de origem em comparação com os termos em outros idiomas, ou seja, do idioma de destino. Deve-se observar se eles apresentam algum nível de correspondência, tanto no campo lexical como no campo semântico para então admiti-lo no tesouro.

Concernente à representação multilíngue e à interoperabilidade na dimensão das comunidades do LOD do domínio da Música, discorre-se que todos os vocabulários na *web* compartilham uma estrutura de dados principais, que é um conceito representado por termos e com identificadores exclusivos e persistentes. Ligados ao conceito estão termos, ou nomes, notas, datas, bibliografias. Assim, vinculá-los aos tesouros e aos diferentes conjuntos de dados que cobrem o domínio da Música implica em identificar e selecionar rótulos que expressem as diferentes características da percepção da música, como o contexto e o conteúdo da música e as propriedades e o contexto do usuário.

Desse modo, os tesouros podem, através da atividade de interligação, estar vinculados às entidades LOD em conjuntos de dados externos que representem elementos de alto e baixo

nível da música, os quais funcionam como hub de dados, como é o caso do Wikidata. Pode haver desses hubs não disporem de rótulos correspondentes exatos, porém o mapeamento de correspondências aproximadas ou relacionadas pode ser executado para garantir a vinculação com os dados contextuais relacionados ao domínio.

Nesse cenário, algumas limitações são evidentes e exigem dos grupos de gestão do tesouro uma avaliação constante, visto que as representações do conhecimento de dado domínio podem variar conforme as características desses conjuntos de dados externos. Logo, podem ocorrer dificuldades relacionadas, principalmente, às barreiras linguísticas e culturais.

A respeito da interoperabilidade, uma das limitações deste estudo foi indicar conjuntos de dados externos que possam ser mapeados com o tesouro, especialmente os que representem elementos de alto e baixo nível da música, como é o caso da Plataforma AcustcBrainz. A Plataforma tem rótulos para representar, por exemplo, dados sobre a dançabilidade, a tonalidade, o timbre, o humor ou o sentimento associado a um registro sonoro. Todavia, eles estão profundamente enraizados ao conteúdo da mídia e não são representados por um termo associado a um vocabulário controlado. Ademais, os termos representam conceitos, portanto, como podem representar um sentimento associado à Música?

Possivelmente, a solução para esse problema seja integrar o conjunto de dados do tesouro a outros tipos de registros como, por exemplo, coleções em acervos bibliográficos ou repositórios de música digital, uma vez que vinculados a outros acervos as dificuldades impostas pelo alto nível de heterogeneidade dos dados do domínio da Música podem ser minimizadas. Assim, a integração com outros agrupamentos de dados permite que o tesouro não se torne um “jardim fechado” e a comunidade de usuários possa recuperar não apenas as informações que necessitam, bem como informações adicionais relativas à solicitação de busca.

Considera-se ainda que um mapeamento do tesouro para o Wikidata é desejável, visto ser capaz de fornecer acesso a informações contextuais na *web* em diversos idiomas, assim como explorar os relacionamentos ou *links* para outros SOC já mapeados para a Wikidata.

Apesar de algumas questões a serem enfrentadas, como a dificuldade de mapear conceitos em diversos idiomas ou correspondências aproximadas, não há dúvidas da importância de utilizar a Wikidata para enriquecer semanticamente os tesouros, tendo em vista esta ser uma importante rede de conceitos sobre diversos domínios, acessível, mantida por diversos usuários e caracterizada como uma base de conhecimentos aberta e colaborativa.

O modelo proposto foi avaliado e aprovado pelos especialistas da área. Entre as considerações tiveram destaque o desenvolvimento de uma abordagem construtivista para a organização do conhecimento, a clareza com que os conceitos e procedimentos foram

apresentados e sua completude quanto às necessidades de construção do tesouro no âmbito do LOD. No entanto, algumas alterações foram recomendadas, principalmente pelo Especialista 1, que reforçou a relevância de analisar as etapas elencadas pela norma ISO 25964, enfatizar a atividade de planejamento como uma das etapas de construção do tesouro e a essencialidade de avaliar o elemento tradução para atender com maior rigor aos princípios da representação multilíngue e interoperabilidade.

Ressalta-se, contudo, que mesmo diante das observações do Especialista aludido, ele afirma que o modelo pode atender ao que se propõe e sua aprovação está condicionada à realização de testes, algo que vai além do escopo desta tese.

Com a pesquisa, foi possível verificar que as instituições ainda mantêm tesouros na atualidade, apesar de haverem outras abordagens menos complexas voltadas à busca e recuperação de informação em bancos de dados ou catálogos de bibliotecas. A possível explicação para esse fenômeno: os tesouros continuam sendo instrumentos eficazes para a indexação de assuntos. Com a evolução da *web* de dados conectados, eles se destacam entre os demais SOC devido a forte adaptabilidade em ambientes de informação que mudam rapidamente, o que permite afirmar que as coleções indexadas com um tesouro representado na *web* possam estar integradas a outros conjuntos de dados.

Assim, constatou-se que os tesouros ocupam um espaço de destaque por serem considerados instrumentos de representação da informação confiáveis, seja devido ao rigor inerente ao processo de construção, seja pela efetiva contribuição para a recuperação da informação.

Como sugestão para trabalhos futuros, são destacados alguns pontos que merecem aprofundamento e não puderam ser discutidos com exatidão nesta tese, a saber:

- a) Aprofundar as discussões em torno do princípio da interoperabilidade, considerando especificamente a interoperabilidade semântica, tendo em vista a dificuldade na etapa de reconciliação dos dados do tesouro com outros conjuntos de dados externos. Muitas vezes, o mapeamento com outros agrupamentos de dados considera apenas o nível lexical, no entanto para alcançar a vinculação entre conceitos equivalentes é preciso considerar, também, o nível semântico. Algo complexo, especialmente, quando se trata de tesouros multilíngues;
- b) Adentrar aos debates em torno das políticas de governança do tesouro. Com a pesquisa, foi possível constatar a participação de muitos atores sociais no decorrer da construção, todavia é fundamental concentrar esforços para que o tesouro permaneça “vivo” e útil. Para isso, deve-se levar em conta a atuação dos diferentes

- públicos, estabelecer parcerias formais com outras organizações, formalizar o papel de cada ator social na manutenção dos projetos e elaborar diretrizes para que esse instrumento continue sendo fruto de uma ação conjunta e dialética entre a comunidade de uso, os profissionais e as instituições de especialidade no domínio;
- c) Desenvolver recursos para melhor integrar a comunidade de usuários. O modelo demonstra as possibilidades de participação desses sujeitos na construção do tesauro, em especial mediante mecanismos de *feedback*. À vista disso, é preciso detalhar os tipos de técnicas usadas para melhor coletar e processar esses dados. Embora o modelo ressalte as perspectivas de utilizar a análise de *logs* de transação ou fóruns públicos de discussão é essencial ter em mente o fornecimento de dados brutos, cuja capacidade de tratamento dependerá dos recursos disponíveis e variará conforme os tipos de instituição (pública ou privada).

Por fim, considera-se que o modelo colaborativo desenvolvido representa uma contribuição para a área da Ciência da Informação, especificamente no tocante às disciplinas de Organização do Conhecimento, Representação da Informação e Recuperação da Informação. A proposta se destaca pela reunião das abordagens mais atuais da construção de tesouros atreladas às demandas de publicação como LOD, sendo o principal contributo o desenvolvimento de um arcabouço teórico representado pela estruturação dos princípios norteadores.

Os princípios servem como o núcleo ou a base teórico-metodológica para o desdobramento de um ambiente colaborativo, algo que pode ser destacado como elemento do ineditismo desta tese. O modelo aplicado ao domínio da Música traz contribuições importantes para a área de Organização do Conhecimento de domínios complexos, em particular por retratar suas implicações para a Organização da Informação Musical. Apesar de contemplar um domínio em específico, a proposta se mostra, inclusive, de acordo com avaliação dos especialistas como relevante, podendo ser conceituado como um modelo de referência que pode ser adaptado a outros contextos ou domínios.

## REFERÊNCIAS

- ABDALLAH, S. et al. The Digital Music Lab: a Big Data infrastructure for digital musicology. **Journal on Computing and Cultural Heritage**, v. 10, n. 1, p. 1–21, 2017. Disponível em: <http://dl.acm.org/citation.cfm?doi=3034773.2983918>. Acesso em: abr. 2020.
- ABECKER, A. et al. Latest Developments of the Linked Thesaurus Framework for the Environment (LusTRE). In: PROCEEDINGS OF THE 29TH ENVIROINFO; 3RD ICT4S CONFERENCE, Copenhagen, Denmark. **Proceedings..** Copenhagen, Denmark: 2015. Disponível em: <https://dblp.org/db/conf/enviroinfo/enviroinfo2015-2.html#AbeckerSWAMP15>. Acesso em: abr. 2020.
- AGUILAR, A. G. Creación y visualización de tesauros a partir de un vocabulario de técnicas de la información y de la comunicación. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 18, n. 37, p. 303–320, 2013. Disponível em: <https://doi.org/10.5007/1518-2924.2013v18n37p303>. Acesso em: abr. 2020.
- ALBERTONI, R. et al. EARTH : an Environmental Application Reference Thesaurus in the Linked Open Data. **Semantic Web**, v. 5, n. 2, p. 165–171, 2014. Disponível em: [http://www.semantic-web-journal.net/system/files/swj520\\_0.pdf](http://www.semantic-web-journal.net/system/files/swj520_0.pdf). Acesso em: abr. 2020.
- ALBERTONI, R.; DE MARTINO, M.; PODESTÀ, P. Quality measures for skos: ExactMatch linksets: an application to the thesaurus framework LusTRE. **Data Technologies and Applications**, v. 52, n. 3, p. 405–423, 2018. Disponível em: <https://www.emeraldinsight.com/doi/10.1108/DTA-05-2017-0037>. Acesso em: abr. 2020.
- ALEXIEV, V.; ISAAC, A.; LINDENTHAL, J. On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). **International Journal on Digital Libraries**, v. 17, n. 1, p. 39–48, 2016. Disponível em: <https://doi.org/10.1007/s00799-015-0162-2>. Acesso em: abr. 2020.
- ALLIK, A. et al. Facilitating music information research with shared open vocabularies. In: CIMIANO, P. et al. (Ed.). **The Semantic Web: ESWC 2013 Satellite Events. ESWC 2013. Lecture Notes in Computer Science**, vol 7955. Berlin, Heidelberg: Springer, 2013. Disponível em: [https://doi.org/10.1007/978-3-642-41242-4\\_20](https://doi.org/10.1007/978-3-642-41242-4_20). Acesso em: jan. 2019.
- ALLIK, A.; FAZEKAS, G.; SANDLER, M. Ontological representation of audio features. **The Semantic Web – ISWC 2016: Lecture Notes in Computer Science**, vol 9982. Cham: Springer, 2016. Disponível em: [https://doi.org/10.1007/978-3-319-46547-0\\_1](https://doi.org/10.1007/978-3-319-46547-0_1). Acesso em: jan. 2019.
- ANIBALDI, S. et al. Migrating bibliographic datasets to the Semantic Web: The AGRIS case. **Semantic Web**, v. 6, n. 2, p. 113–120, 2015. Disponível em: <http://www.semantic-web-journal.net/system/files/swj463.pdf>. Acesso em: abr. 2020.
- ARAÚJO, C. A. Á. Correntes teóricas da ciência da informação. **Ciência da Informação**, v. 38, n. 3, p. 192–204, set. /dez., 2009. Disponível em: <https://doi.org/10.1590/S0100-19652009000300013>. Acesso em: jan. 2019.
- ASSUMPCÃO, F. S. **Modelo para a publicação de dados de autoridade como Linked Data**. 2018. Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Informação,

Universidade Estadual Paulista (Unesp), 2018. Disponível em: <http://hdl.handle.net/11449/152759>. Acesso em: jan. 2019.

ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. da C. Metadata Authority Description Schema (MADS): uma alternativa à utilização do formato MARC 21 para dados de autoridade. **Informação & Informação**, v. 18, n. 1, p. 106–126, 2013. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2013v18n1p106>. Acesso em: jan. 2019.

BACA, M.; GILL, M. Encoding multilingual knowledge systems in the digital age: the Getty Vocabularies. **Knowledge organization**, v. 42, n. 4, p. 232–243, 2014. Disponível em: <http://dx.doi.org/10.7152/nasko.v5il.15179>. Acesso em: abr. 2020.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação da informação: conceitos e tecnologia das máquinas de busca**. 2. ed. Porto Alegre: Bookman, 2013.

BARITÉ, M. La Garantía Literaria : Vigencia Y Proyección Teórico-metodológica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, VIII, Salvador, Bahia. **Anais...** Salvador, Bahia: ANCIB, 2007. Disponível em: <http://www.enancib.ppgci.ufba.br/artigos/GT2--068.pdf>. Acesso em: ago. 2020.

BARITÉ, M. et al. Garantia literária: elementos para uma revisão crítica após um século. **Transinformação**, v. 22, n. 2, p. 123–138, 2010. Disponível em: <http://dx.doi.org/10.1590/S0103-37862010000200003>. Acesso em: ago. 2020.

BARITÉ, M. G. Los conceptos y su representación: Una perspectiva terminológica para el tratamiento temático de la información. **Scire: representación y organización del conocimiento**, v. 6, n. 1, p. 31–53, 2000. Disponível em: <https://www.iberid.eu/ojs/index.php/scire/article/view/1123>. Acesso em: ago. 2020.

BAUER, F.; KALTENBÖCK, M. **Linked Open Data: the essentials: a quick start guide for decision makers**. Vienna, Austria: Edition Mono, 2011. Disponível em: <https://www.reeep.org/LOD-the-Essentials.pdf>. Acesso em: jan. 2019.

BEGHTOL, C. Semantic Validity : Concepts of Warrant in Bibliographic classification systems. **Library Resources & Technical Services**, v. 30, n. 2, p. 109–138, 1986. Disponível em: <http://downloads.alcts.ala.org/lrts/lrtsv30no2.pdf>. Acesso em: ago. 2020.

BEGHTOL, C. Domain analysis, literary warrant, and consensus: The case of fiction studies. **Journal of the American Society for Information Science**, v. 46, n. 1, p. 30–44, 1995. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199501\)46:1%3C30::AID-ASI4%3E3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-4571(199501)46:1%3C30::AID-ASI4%3E3.0.CO;2-F). Acesso em: ago. 2020.

BELLINI, P.; CAPPUCIO, A.; NESI, P. Collaborative and assisted SKOS generation and management. In: INTERNATIONAL CONFERENCE ON DISTRIBUTED MULTIMEDIA SYSTEMS (DMS 2011), 17., Florence, Italy, 2011. **Proceedings...** Florence, Italy: University of Florence, 2011. p. 28–33. Disponível em: <http://www.disit.org/disitmn/DMS2011-nesi-paper16-OpenSpaceInnoMind-SKOS-v1-2-6pages.pdf>. Acesso em: abr. 2020.

BENEVENTANO, D. *et al.* Semantic annotation of the CEREALAB database by the AGROVOC linked dataset. **Ecological Informatics**, v. 26, part 2, p. 119–126, 2015. Disponível em: <http://dx.doi.org/10.1016/j.ecoinf.2014.07.002>. Acesso em: abr. 2020.

BERGAMIN, G.; BERGAMIN, G.; LUCARELLI, A. The Nuovo soggettario as a service for the linked data world. **JLIS.it**, v. 4, n. 1, p. 213, 2013. Disponível em: <http://dx.doi.org/10.4403/jlis.it-5474>. Acesso em: abr. 2020.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, pp. 34-43, may, 2001. Disponível em: <https://www.jstor.org/stable/26059207>. Acesso em: jan. 2019.

BERTI JUNIOR, D. W. et al. Semiautomatização de relações em tesouros: uma proposta para o refinamento de relacionamentos semânticos a partir do tesouro Agrovoc. **Informação & Informação**, v. 22, n. 3, p. 377, 2017. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2017v22n3p377>. Acesso em: jan. 2020.

BINDING, C.; TUDHOPE, D. Improving interoperability using vocabulary linked data. **International Journal on Digital Libraries**, v. 17, n. 1, p. 5–21, 2016. Disponível em: <https://doi.org/10.1007/s00799-015-0166-y>. Acesso em: abr. 2020.

BOCCEGA, M. A. Apresentação. In: CINTRA, A. M. M. et al. (Ed.). **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002. p. 92.

BOUTIN-QUESNEL, R. et al. **Vocabulaire systématique de la terminologie**. Québec: Publications du Québec, 1985. Disponível em: [https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/voc\\_systematique\\_terminologie.pdf](https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/voc_systematique_terminologie.pdf). Acesso em: ago. 2020.

BRÄSCHER, M.; CARLAN, E. Sistemas de Organização do Conhecimento: antigas e novas linguagens. In: ROBREDO, J.; BRÄSCHER, M. (Ed.). **Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento**. Brasília: IBICT, 2010. p. 335. Disponível em: <http://repositorio.ibict.br/bitstream/123456789/36/1/eroic.pdf>. Acesso em: jan. 2019.

BROWN, S. et al. Bulding listening experience Linked Data through crowd-sourcing and reuse of Library Data. In: INTERNATIONAL WORKSHOP ON DIGITAL LIBRARIES FOR MUSICOLOGY, 1., London. **Proceedings...** London: 2014. Disponível em: <https://dl.acm.org/citation.cfm?id=2660168>. Acesso em: jan. 2019.

BRYL, V. et al. Interlinking and Knowledge Fusion. In: AUER, S.; BRYL, V.; TRAMP, S. (Ed.). **Linked Open Data - Creating Knowledge Out of Interlinked Data: results of the LOD2 Project**. Cham, Switzerland: Springer, 2014. p. 70–89. Disponível em: [https://doi.org/10.1007/978-3-319-09846-3\\_4](https://doi.org/10.1007/978-3-319-09846-3_4). Acesso em: jan. 2019.

BURGOYNE, J. A.; FUJINAGA, I.; DOWNIE, J. S. Music Information Retrieval. In: SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (Ed.). **A New Companion to Digital Humanities**. Hoboken, Nova Jersey, EUA: Wiley, 2015. p. 213–228. Disponível em: <https://doi.org/10.1002/9781118680605.ch15>. Acesso em: jan. 2019.

ÇAĞDAŞ, V.; STUBKJÆR, E. A SKOS vocabulary for Linked Land Administration: Cadastre and Land Administration Thesaurus. **Land Use Policy**, v. 49, p. 668–679, 2015. Disponível em: <https://doi.org/10.1016/j.landusepol.2014.12.017>. Acesso em: abr. 2020.

CAMPOS, M. L. de A. **Em busca de princípios comuns na área de representação da informação**: uma comparação entre o método de classificação facetada, o método de tesouro-baseado-em-conceito e a teoria geral da terminologia. 1994. 196 f. Dissertação (Mestrado em Ciência da Informação) - Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1994.

CAMPOS, M. L. de A. et al. Estudo comparativo de softwares de construção de tesouros. **Perspectivas em Ciência da Informação**, v. 1, n. 11, p. 68–81, jan. /abr., 2006. Disponível em: <http://dx.doi.org/10.1590/S1413-99362006000100006>. Acesso em: jan. 2019.

CARACCILO, C. *et al.* Thesaurus maintenance, alignment and publication as linked data: The AGROVOC use case. In: GARCÍA-BARRIOCANAL, E. *et al.* (Org.). **Metadata and Semantic Research, MTSR 2011: Communications in Computer and Information Science**, v. 240. Berlin, Heidelberg: Springer, 2011. p. 489–499. Disponível em: [https://doi.org/10.1007/978-3-642-24731-6\\_48](https://doi.org/10.1007/978-3-642-24731-6_48). Acesso em: abr. 2020.

CARACCILO, C. *et al.* The AGROVOC Linked Dataset. **Semantic Web**, v. 4, n. 3, p. 341–348, 2013. Disponível em: <https://dl.acm.org/doi/10.5555/2786071.2786087>. Acesso em: abr. 2020.

CASTELLS, M. **Sociedade em rede**. 17. ed. São Paulo: Paz & Terra, 2016.

CATARINO, M. E. Simple Knowledge Organization System : construindo sistemas de organização do conhecimento no contexto da Web Semântica. **Informação & Tecnologia (ITEC)**, v. 1, n. 1, p. 17–28, 2014. Disponível em: <http://periodicos.ufpb.br/index.php/itec/article/download/19307/10980>. Acesso em: jan. 2019.

CENTRE FOR DIGITAL MUSIC. **DBTune.org**. Disponível em: <http://dbtune.org/>. Acesso em: 27 mar. 2019.

CERVANTES, B. M. N. **A Construção De Tesouros Com a Integração De Procedimentos Terminográficos**. 2009. Tese (Doutorado em Ciência da Informação) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília, 2009. Disponível em: <http://hdl.handle.net/11449/103382>. Acesso em: jan. 2019.

CHARLES, V.; FREIRE, N.; ISAAC, A. Links, languages and semantics: linked data approaches in The European Library and Europeana. In: IFLA WORLD LIBRARY AND INFORMATION CONGRESS, 80., Lyon, august, 2014. **Proceedings...** Lyon: IFLA, 2014. Disponível em: [http://ifla2014-satdata.bnf.fr/pdf/iflalld2014\\_submission\\_Charles\\_Freire\\_Isaac.pdf](http://ifla2014-satdata.bnf.fr/pdf/iflalld2014_submission_Charles_Freire_Isaac.pdf). Acesso em: jan. 2019.

CHARMAZ, K. **A construção da teoria fundamentada**: guia prático para análise qualitativa. Porto Alegre: Artmed, 2009.

CHAU, R.; YEH, C. H. Construction of a fuzzy multilingual thesaurus and its application to cross-lingual text retrieval. In: ZHONG, N. et al. (Eds.). **Web Intelligence: Research and Development. WI 2001**. Berlin, Heidelberg: Springer, 2001. p. 340-345. Lecture Notes in Computer Science, vol 2198. Disponível em: [https://doi.org/10.1007/3-540-45490-X\\_40](https://doi.org/10.1007/3-540-45490-X_40). Acesso em: jan. 2019.

CHEN, C.; YEH, J.; SIE, S. Government Ontology and Thesaurus Construction : A Taiwanese Experience. In: FOX, E.A. et al. (eds.). **Digital libraries: implementing strategies**



and sharing experiences. Berlin, Heidelberg: Springer, 2005. Lecture Notes in Computer Science, vol 3815. Disponível em: [https://doi.org/10.1007/11599517\\_30](https://doi.org/10.1007/11599517_30). Acesso em: jan. 2019.

CHEN, P. P. S. The Entity-Relationship Model: toward a Unified View of Data. **ACM Transactions on Database Systems (TODS)**, v. 1, n. 1, p. 9–36, 1976. Disponível em: <https://dl.acm.org/doi/10.1145/320434.320440>. Acesso em: jan. 2019.

CHEUNG, C. F. et al. A multi-faceted and automatic knowledge elicitation system (MAKES) for managing unstructured information. **Expert Systems with Applications**, v. 38, n. 5, p. 5245–5258, 2011. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2010.10.033>. Acesso em: jan. 2019.

CHURCH, K. W.; HILL, M.; HANKS, P. Word Association Norms, Mutual Information, and Lexicography. In: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 27<sup>th</sup>. Vancouver, British Columbia, Canada. **Proceedings...** Vancouver, British Columbia, Canada: Association for Computational Linguistics, 1990. Disponível em: <https://www.aclweb.org/anthology/P89-1010>. Acesso em: jan. 2019.

CINTRA, A. M. M. et al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002.

COBB, J. The Journey to Linked Open Data: The GettyVocabularies. **Journal of Library Metadata**, v. 15, n. 3–4, p. 142–156, 2016. Disponível em: <https://doi.org/10.1080/19386389.2015.1103081>. Acesso em: abr. 2020.

CORRÊA, G. N.; MARCACINI, R. M.; REZENDE, S. O. Uso da mineração de textos na análise exploratória de artigos científicos. **Relatórios técnicos do Instituto de Ciências Matemáticas e Computação**, v. 30, São Carlos, SP: Instituto de Ciências Matemáticas e Computação, 2012. Disponível em: <http://repositorio.icmc.usp.br/handle/RIICMC/6631>. Acesso em: jun. 2020.

COYLE, K. Vocabularies: Term Lists and Thesauri. **Library Technology Reports**, v. 48, n. 4, p. 27–36, 2012. Disponível em: <https://journals.ala.org/index.php/ltr/article/view/4670/5540>. Acesso em: abr. 2020.

COYLE, K. et al. Linked Data Infrastructure Models: areas of focus for Program for Cooperative Cataloging Strategies. Washington, DC: The Library of Congress, 2017. Disponível em: <https://www.loc.gov/aba/pcc/documents/LinkedDataInfrastructureModels.pdf>. Acesso em: jan. 2020.

CRAWFORD, T. et al. Explorations in linked data practice for early music corpora. In: ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES, London, UK, 2014. **Proceedings...** London: IEEE, 2014. p. 309–312. Disponível em: <https://doi.org/10.1109/JCDL.2014.6970184>. Acesso em: jan. 2019.

CREATIVE COMMONS. **Sobre as Licenças**. [2018]. Disponível em: <https://br.creativecommons.org/licencas/>. Acesso em: 17 maio. 2018.

CROUCH, C. J.; YANG, B. Experiments in automatic statistical thesaurus construction. In: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND

DEVELOPMENT IN INFORMATION RETRIEVAL, 15., Copenhagen, Denmark, 1992. **Proceedings...** New York: Association for Computing Machinery, 1992. p. 77–88. Disponível em: <https://doi.org/10.1145/133160.133180>. Acesso em: jan. 2020.

CURÉ, O.; BLIN, G. Introduction. In: CURÉ, O.; BLIN, G. (Ed.). **RDF Database Systems: Triples Storage and SPARQL Query Processing**. Curie, France: Morgan Kaufmann, 2015. Disponível em: <https://doi.org/10.1016/c2013-0-14009-3>. Acesso em: jan. 2020.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, v. 7, n. 2, p. 101–107, 1978. Disponível em: <http://revista.ibict.br/ciinf/article/view/115/115>. Acesso em: ago. 2020.

DANNENBERG, R. et al. Panel: New Directions in music information retrieval. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, Habana, Cuba, September 2001., Habana, Cuba. **Proceedings...** Habana, Cuba: 2001. Disponível em: <https://quod.lib.umich.edu/i/icmc/bbp2372.2001.037/1>. Acesso em: jun. 2020.

DAVISON, S. et al. Enhancing an OAI-PMH Service Using Linked Data: A Report from the Sheet Music Consortium. **Journal of Library Metadata**, v. 13, n. 2–3, p. 141–162, 2013. Disponível em: <https://doi.org/10.1080/19386389.2013.826067>. Acesso em: jan. 2019.

DBPEDIA ASSOCIATION. **DBpedia**. [2020]. Disponível em: <https://wiki.dbpedia.org/>. Acesso em: ago. 2020.

DE ROURE, D. et al. An e-Research approach to Web-scale music analysis. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 369, n. 1949, p. 3300–3317, 2011. Disponível em: <https://doi.org/10.1098/rsta.2011.0171>. Acesso em: jan. 2019.

DEBEVERE, P. *et al.* Linking thesauri to the linked open data cloud for improved media retrieval. In: INTERNATIONAL WORKSHOP ON IMAGE ANALYSIS FOR MULTIMEDIA INTERACTIVE SERVICES (WIAMIS 2011), 12., apr. 2011. **Proceedings...** The Netherlands: TUDelft, 2011. p. 28–37. Disponível em: <https://biblio.ugent.be/publication/1210156>. Acesso em: abr. 2020.

DEERWESTER, S. et al. Indexing by Latent Semantic Analysis Scott. **Journal of the American Society for Information Science**, v. 41, n. 6, p. 391–407, 1990. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6%3C391::AID-ASII%3E3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASII%3E3.0.CO;2-9). Acesso em: jan. 2019.

DÍAZ-CORONA, D. *et al.* Profiling of knowledge organisation systems for the annotation of Linked Data cultural resources. **Information Systems**, v. 84, p. 17–28, 2019. Disponível em: <https://doi.org/10.1016/j.is.2019.04.008>. Acesso em: abr. 2020.

DIMITROVA, V. *et al.* Interoperability of language-related information: Mapping the BLL thesaurus to LEXVO and glottolog. In: International Conference on Language Resources and Evaluation, LREC 2018, 11., 2018. **Proceedings...** Paris: European Language Resources Association, 2018. p. 4555–4561. Disponível em: <http://www.lrec-conf.org/proceedings/lrec2018/pdf/259.pdf>. Acesso em: abr. 2020.

DIX, A. et al. Authority and Judgement in the digital archive. In: INTERNATIONAL WORKSHOP ON DIGITAL LIBRARIES FOR MUSICOLOGY, 1., London, sept. 2014. **Proceedings...** New York: Association for Computing Machinery, 2014. Disponível em:

<https://doi.org/10.1145/2660168.2660171>. Acesso em: abr. 2020.

DODEBEI, V. L. D. **Tesouro: linguagem de representação da memória documentária**. Rio de Janeiro: Intertexto; Interciência, 2002.

DOLAN, E. I. Review: MIMO: Musical Instrument Museums Online. **Journal of the American Musicological Society**, v. 70, n. 2, p. 555–565, 2017. Disponível em: <https://doi.org/10.1525/jams.2017.70.2>. Acesso em: jan. 2019.

DOREMUS. **DOREMUS: Doing Reusable Musical Data**. [2019]. Disponível em: <http://www.doremus.org/>. Acesso em: 7 fev. 2019.

DOWNIE, J. S. The Scientific Evaluation of Music Information Retrieval Systems: Foundations and Future. **Computer Music Journal**, v. 28, n. 2, p. 12–23, 2004. Disponível em: <http://www.mitpressjournals.org/doi/10.1162/014892604323112211>. Acesso em: fev. 2019.

DUERST, M.; SUIGNARD, M. **Internationalized Resource Identifiers (IRIs)**. The Internet Society, 2005. Disponível em: <https://www.w3.org/International/iri-edit/draft-duerst-iri.html>. Acesso em: 24 jan. 2019.

ETCHEVERRY, L.; VAISMAN, A. A. **Views over RDF Datasets: a state-of-the-art and Open Challenges**. Montevideo, Uruguai: Facultad de Ingeniería de la Universidad de la República, 2012. Disponível em: <http://arxiv.org/abs/1211.0224>. Acesso em: jan. 2019.

EUROPEANA FOUNDATION. **Product Specification for Europeana Thematic Collections**. Haia, Netherlands: Europeana, 2015. Disponível em: [https://pro.europeana.eu/files/Europeana\\_Professional/Europeana\\_Network/europeana-thematic-collections-specification.pdf](https://pro.europeana.eu/files/Europeana_Professional/Europeana_Network/europeana-thematic-collections-specification.pdf). Acesso em: jan. 2019.

EVERITT, B. S. et al. Miscellaneous Clustering Methods. In: **Cluster Analysis**. London: John Wiley & Sons, 2011. p. 215–255. Disponível em: <https://doi.org/10.1002/9780470977811.ch8>. Acesso em: jan. 2019.

FABEIRO, R.; MASAGUÉ, C.; ROVIRA JARQUE, A. Evolució de les dades obertes enllaçades (LOD) en el context dels vocabularis controlats. **Item: Revista de Biblioteconomia i Documentació**, v. 2, n. 65-66, p. 1–24, 2019. Disponível em: <https://www.raco.cat/index.php/Item/article/view/353616>. Acesso em: abr. 2020.

FERREIRA, J. A.; SANTOS, P. L. V. A. da C. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Informação & Sociedade: estudos**, v. 23, n. 2, p. 13–23, 2013. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/15436>. Acesso em: jan. 2019.

FILLON, T. et al. Telemeta: An open-source web framework for ethnomusicological audio archives management and automatic analysis. In: INTERNATIONAL WORKSHOP ON DIGITAL LIBRARIES FOR MUSICOLOGY, 1., London, sept. 2014. **Proceedings...** New York: Association for Computing Machinery, 2014. Disponível em: <https://doi.org/10.1145/2660168.2660169>. Acesso em: jan. 2019.

FINK, E. E. **American Art Collaborative (AAC) Linked Open Data (LOD) Initiative: overview and recommendations for good practices**. Washington, DC: The American Art

- Collaborative (AAC), 2018. Disponível em: <https://s3.amazonaws.com/assets.saam.media/files/documents/2020-07/OverviewandRecommendationsAccessible.pdf>. Acesso em: jan. 2019.
- FRAKES, W. B.; BAEZA-YATES, R. **Information retrieval: data structures and algorithms**. New Jersey: Prentice Hall, Englewood Cliffs, 1992.
- FROSTERUS, M. et al. Linked Open Ontology Cloud KOKO: Managing a System of Cross-Domain Lightweight Ontologies. In: CIMIANO, P. et al. (Ed.). **The Semantic Web: ESWC 2013 Satellite Events**. New York: Springer, 2013. p. 279–281. Disponível em: [https://doi.org/10.1007/978-3-642-41242-4\\_49](https://doi.org/10.1007/978-3-642-41242-4_49). Acesso em: jan. 2020.
- FUGINI, M. G. Automatic thesaurus construction supporting fuzzy retrieval of reusable components. In: SAC '95: ACM SYMPOSIUM ON APPLIED COMPUTING, Nashville Tennessee, 1995. **Proceedings...** New York: Association for Computing Machinery. Disponível em: <https://doi.org/10.1145/315891.316093>. Acesso em: jan. 2019.
- FUJITA, M. S. L. A política de indexação para representação e recuperação da informação. In: LEIVA, I. G.; FUJITA, M. S. L. (Ed.). **Política de indexação**. Marília: Oficina Universitária, 2012.
- FUTRELLE, J.; DOWNIE, J. S. Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000-2002. **Journal of New Music Research**, v. 32, n. 2, p. 121–131, 2003. Disponível em: <http://www.tandfonline.com/doi/abs/10.1076/jnmr.32.2.121.16740>. Acesso em: jan. 2019.
- GABRIEL JÚNIOR, R. F.; LAIPELT, R. do C. F. Thesa: semantic thesaurus construction tool applied interoperable. **P2P & Inovação**, v. 3, n. 2, p. 124–145, mar. /set., 2017. Disponível em: <http://dx.doi.org/10.21721/p2p.2017v3n2.p124-145%0ATHESA>. Acesso em: jan. 2019.
- GANGEMI, A.; PRESUTTI, V. The bourne identity of a web resource. In: IDENTITY REFERENCE AND THE WEB WORKSHOP (IRW) AT THE WWW CONFERENCE, May 22–26, 2006, Edinburgh, UK. **Proceedings...** 2006. Edinburgh, 2006. Disponível em: [www.ibiblio.org/hhalpin/irw2006/vpresutti.pdf](http://www.ibiblio.org/hhalpin/irw2006/vpresutti.pdf). Acesso em: jan. 2019.
- GARCÍA-GUTIÉRREZ, A. Una teoría de los conceptos. In: **Epistemología de la documentación**. Barcelona: Stonberg, 2011. p. 104–135.
- GARCÍA-TORRES, A.; PAREJA-LORA, A.; PRADANA-LÓPEZ, D. Reutilización de tesauros: el documentalista frente al reto de la web semántica. **El profesional de la información**, v. 17, n. 1, p. 8–21, 2008. Disponível em: <https://doi.org/10.3145/epi.2008.ene.02>. Acesso em: jan. 2019.
- GERBÉ, O.; KERHERVÉ, B. A model-driven approach to SKOS implementation. In: INTERNATIONAL CONFERENCE ON INTERNET AND WEB APPLICATIONS AND SERVICES (ICIW 2010), 5., 2010. **Proceedings...** Barcelona: IEEE, 2010. p. 484–488, 2010. Disponível em: <https://doi.org/10.1109/iciw.2010.79>. Acesso em: abr. 2020.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. ed. São Paulo: Atlas, 2008.

GODOY, A. S. Pesquisa qualitativa: tipos fundamentais. **Revista de Administração de Empresas**, v. 35, n. 3, p. 20–29, 1995. Disponível em: <http://dx.doi.org/10.1590/S0034-75901995000300004>. Acesso em: jan. 2019.

GOMES, H. E. **Manual de elaboração de tesouros monolíngues**. Brasília: Ministério da Educação, 1990.

GRACY, K. F.; ZENG, M. L.; SKIRVIN, L. Exploring Methods To Improve Access to Music Resources by Aligning Library Data With Linked Data: A Report of Methodologies and Preliminary Findings. **Journal of the American Society for Information Science and Technology**, v. 64, n. 10, p. 2078–2099, 2013. Disponível em: <https://doi.org/10.1002/asi.22914>. Acesso em: jan. 2019.

GRANADA, R. L.; VIEIRA, R.; LIMA, V. L. S. de. Evaluating co-occurrence order for automatic thesaurus construction. In: INTERNATIONAL CONFERENCE ON INFORMATION REUSE AND INTEGRATION, IRI 2012, 13<sup>th</sup>., Las Vegas, NV, 2012. **Proceedings...** Las Vegas: IEEE, 2012. p. 474–481, 2012. Disponível em: <https://doi.org/10.1109/IRI.2012.6303046>. Acesso em: jan. 2019.

GRFENSTETTE, G. **Explorations in automatic thesaurus discovery**. Boston: Kluwer Academic Publishers, 1994. Disponível em: <https://www.springer.com/gp/book/9780792394686>. Acesso em: jan. 2019. DOI: 10.1007/978-1-4615-2710-7.

GRFENSTETTE, G. SQLET: Short Query Linguistic Expansion Techniques, Palliating One-Word Queries by Providing Intermediate Structure to Text. In: INTERNATIONAL CONFERENCE COMPUTER-ASSISTED INFORMATION RETRIEVAL: Recherche d'Information et ses Applications - RIAO 1997, 5<sup>th</sup>., Montreal. **Proceedings...** Montreal: ACM, 1997. Disponível em: <https://dl.acm.org/doi/pdf/10.5555/2856695.2856739>. Acesso em: jan. 2019.

GRIMOÛARD, C. S. The Thesaurus for French Local Archives and the Semantic Web. **Procedia - Social and Behavioral Sciences**, v. 147, p. 206–212, aug., 2014. Disponível em: <https://doi.org/10.1016/j.sbspro.2014.07.153>. Acesso em: jan. 2019.

HAGIWARA, M.; OGAWA, Y.; TOYAMA, K. PLSI utilization for automatic thesaurus construction. In: DALE, R. et al. (Eds.) In: INTERNATIONAL JOINT CONFERENCE OF NATURAL LANGUAGE, 2<sup>th</sup>, Berlin. **Proceedings...** Berlin: Springer, 2005. Disponível em: [https://doi.org/10.1007/11562214\\_30](https://doi.org/10.1007/11562214_30). Acesso em: jan. 2019.

HAN, L. et al. Improving word similarity by augmenting PMI with estimates of word polysemy. *IEEE Transactions on Knowledge and Data Engineering*, v. 25, n. 6, p. 1307–1322, jun., 2013. Disponível em: <https://doi.org/10.1109/TKDE.2012.30>. Acesso em: jan. 2019.

HAN, Y. S.; HAN, Y. K.; CHOI, K. Lexical Concept Acquisition From Collocation Map. In: SIGLEX WORKSHOP ON ACQUISITION OF LEXICAL KNOWLEDGE FROM TEXT; ANNUAL MEETING OF THE ACL, 31<sup>st</sup>., Ohio State University, Columbus, Ohio. **Proceedings...** Columbus, Ohio: Ohio State University, 1992. Disponível em: <https://www.aclweb.org/anthology/W93-0103.pdf>. Acesso em: jan. 2019.

HANANDEH, E. Building an Automatic Thesaurus to Enhance Information Retrieval Information Retrieval. **International journal of computer science**, v. 10, n. 3, p. 22–29,



2013. Disponível em: <https://ijcsi.org/papers/IJCSI-10-1-3-676-686.pdf>. Acesso em: jan. 2019.

HARPER, C. A. Encoding Library of Congress Subject Headings in SKOS : Authority Control for the Semantic Web. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, Colima, Mexico, 2006. **Proceedings....**, Colima, México, 2006. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/842>. Acesso em: jan. 2019.

HARPRING, P. Introduction to Controlled Vocabularies: Terminologies for Art, Architecture, and Other Cultural Works. Los Angelis, California: Getty Research Institute, 2015. Disponível em: <https://www.getty.edu/publications/virtuallibrary/160606018X.html>. Acesso em: jan. 2019.

HARPRING, P. **Introdução aos Vocabulários Controlados**: Terminologia para arte, arquitetura e outras obras culturais: Getty Research Institute. São Paulo: ACAM Portinari, 2016. Coleção Gestão e Documentação de Acervos: Textos de Referência, 4.

HARPRING, P. Linking the Getty Vocabularies: the content perspective, Including an Update on CONA. In: PACIFIC NEIGHBORHOOD CONSORTIUM ANNUAL CONFERENCE AND JOINT MEETINGS (PNC), 2018. **Proceedings...** San Francisco, CA, USA: IEEE, 2018. p. 1–8. Disponível em: <https://ieeexplore.ieee.org/document/8579460>. Acesso em: abr. 2020.

HEARST, M. A. Automatic Acquisition of Hyponyms from Large Text Corpora. In: CONFERENCE ON COMPUTATIONAL LINGUISTICS, 14th, Stroudsburg, PA, 1992. **Proceedings...** Stroudsburg, PA: Association for Computational Linguistics, 1992. p. 23–28, 1992. Disponível em: <https://doi.org/10.3115/992133.992154>. Acesso em: jan. 2019.

HEATH, T.; BIZER, C. Linked Data: evolving the Web into a Global Data Space. Williston, VT: Morgan & Claypool, 2011. Disponível em: <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>. Acesso em: mar. 2019.

HINDLE, D. Noun classification from predicate-argument structures. In: ANNUAL MEETING ON ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 28<sup>th</sup>., **Proceedings...** Stroudsburg, PA: Association for Computational Linguistics, 1990. p. 268–275. Disponível em: <https://dl.acm.org/doi/10.3115/981823.981857>. Acesso em: jan. 2019.

HODGE, G.; BEDFORD, D. **The Transportation Research Thesaurus**: Capabilities and Enhancements. Washington, DC: The: National Academy of Sciences, 2018. Disponível em: <https://doi.org/10.17226/25087>. Acesso em: jun. 2020.

HULME, E. W. Principles of Book Classification. **Library Association Record**, v. 13, p. 389–394, 1911.

IBÁÑEZ, L. D. et al. Live linked data: Synchronising semantic stores with commutative replicated data types. **International Journal of Metadata, Semantics and Ontologies**, v. 8, n. 2, p. 119–133, 2013. Disponível em: <https://dx.doi.org/10.1504/IJMSO.2013.056605>. Acesso em: jun. 2020.

IBEKWE-SANJUAN, F. Constructing and maintaining knowledge organization tools: A symbolic approach. **Journal of Documentation**, v. 62, n. 2, p. 229–250, 2006. Disponível

em: <https://doi.org/10.1108/00220410610653316>. Acesso em: jan. 2019.

INSIGHT CENTRE FOR DATA ANALYTICS. **The Linked Open Data Cloud**. Disponível em: <https://lod-cloud.net/>. Acesso em: dez. 2010.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. Statement of International Cataloguing Principles. Disponível em: [http://www.ifla.org/files/cataloguing/icp/icp\\_2009-en.pdf](http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf). Acesso em: 30 abr. 2018.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 2709:2008 : Information and documentation — Format for information exchange**. Geneva, Switzerland: ISO, 2008. Disponível em: <https://www.iso.org/standard/41319.html>. Acesso em: jan. 2020.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 25964-1: Information and Documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval**. Geneva, Switzerland: ISO, 2011.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 25964-2: Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies**. Geneva, Switzerland: ISO, 2013.

ISAAC, A. et al. Integrated access to cultural heritage resources through representation and alignment of controlled vocabularies. *Library Review*, v. 57, n. 3, p. 187–199, 2008. Disponível em: <https://doi.org/10.1108/00242530810865475>. Acesso em: jan. 2019.

ISAAC, A.; MATTHEZING, H. Representing and Aligning Thesauri for an Integrated Access to Cultural Heritage Resources. In: INTERNATIONAL SEMINAR “INFORMATION ACCESS FOR THE GLOBAL COMMUNITY”, 29., 4-5 June 2007. **Proceedings...** The Hague: UDC Consortium, 2007. p. 301–310. Disponível em: <http://hdl.handle.net/10150/106244>. Acesso em: abr. 2020.

ISAAC, A.; TUDHOPE, D. S. **[ISO-THES]**. 2015. Disponível em: <http://pub.tenforce.com/schemas/iso25964/skos-thes>. Acesso em: out. 2019.

ISOTANI, S.; BITTENCOURT, I. **Dados abertos conectados**. São Paulo: Novatec, 2015.

JORGE, N. M. DA C. **Ensaio sobre o AAT-Art & Architecture Thesaurus : proposta terminológica de adaptação à realidade portuguesa**. 2011. 150 f. Universidade do Porto, Porto, 2011. Disponível em: <http://repositorio-aberto.up.pt/handle/10216/57042>. Acesso em: abr. 2020.

JUAN, P. de; IGLESIAS, C. Semantically enriching the search system of a Music Digital Library. In: DU, W.; ENSAN, F. (Ed.). *Canadian semantic web: technologies and applications*. New York: Springer, 2010. p. 175-194. Disponível em: [https://doi.org/10.1007/978-1-4419-7335-1\\_8](https://doi.org/10.1007/978-1-4419-7335-1_8). Acesso em: jan. 2019.

JUPP, S. et al. Knowledge representation for web navigation. In: SEMANTIC WEB APPLICATIONS AND TOOLS FOR LIFE SCIENCES, Edinburgh, United Kingdom, November 28, 2008. **Proceedings...** Edinburgh: CEUR Workshop Proceedings, 2008. Disponível em: <http://ceur-ws.org/Vol-435/paper03.pdf>. Acesso em: jan. 2019.

KAMALZADEH, M.; BAUR, D. A survey on music listening and management behaviours.

- In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE (ISMIR), 13, Canadá. **Proceedings...** Canadá: ISMIR, 2012. Disponível em: [http://ismir2012.ismir.net/event/papers/373\\_ISMIR\\_2012.pdf](http://ismir2012.ismir.net/event/papers/373_ISMIR_2012.pdf). Acesso em: jan. 2019.
- KEIL, K.; WARD, J. A. Applications of RISM data in digital libraries and digital musicology. **International Journal on Digital Libraries**, v. 20, p. 3–12, 2017. Disponível em: <https://doi.org/10.1007/s00799-016-0205-3>. Acesso em: jan. 2019.
- KELLY, E. Music indexing and retrieval : current problems. **Indexer**, v. 28, n. 4, p. 163–167, 2010. Disponível em: <https://doi.org/10.3828/indexer.2010.45>. Acesso em: jan. 2019.
- KELLY, E. Linked Data and music: current projects and opportunities. **Indexer**, v. 33, n. 1, p. 2–7, 2015. Disponível em: <https://doi.org/10.3828/indexer.2015.2>. Acesso em: jan. 2019.
- KEMPF, A. O.; NEUBERT, J. The role of thesauri in an open web: a case study of the STW Thesaurus for Economics. **Knowledge organization**, v. 43, p. 160–173, 2016. Disponível em: <https://doi.org/10.5771/0943-7444-2016-3-160>. Acesso em: abr. 2020.
- KLESS, D. et al. Thesaurus and Ontology Structure: Formal and Pragmatic Differences and Similarities. **Journal of the Association for Information Science and Technology**, v. 66, n. 7, p. 1348–1366, 2014. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23268>. Acesso em: jun. 2020.
- KLIE, S. M. **An RDF Schema for Thesauri**. Deventer: Information Services and Management, 2003. Disponível em: <https://www.agnosis.be/static/file/rdfs-thesaurus.pdf>. Acesso em: jan. 2019.
- KNOBLOCK, C. A. et al. Lessons Learned in Building Linked Data for the American Art Collaborative. In: D'AMATO, C. et al. (Ed.). **The Semantic Web: ISWC 2017**. Cham: Springer, 2017. p. 263-219. Lecture Notes in Computer Science, vol 10588. Disponível em: [https://doi.org/10.1007/978-3-319-68204-4\\_26](https://doi.org/10.1007/978-3-319-68204-4_26). Acesso em: jan. 2019.
- KOBASHI, N. Y. Análise documentária e representação da informação. **Revista Informare: Cadernos de pós-graduação em Ciência da Informação**, Rio de Janeiro, v. 2, n. 2, p. 5–27, 1996. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/40976>. Acesso em: jun. 2020.
- KORN, N.; OPPENHEIM, C. **Licensing Open Data: a practical guide**. [Kingdom: Discovery Programme], 2011. Disponível em: [http://discovery.ac.uk/files/pdf/Licensing\\_Open\\_Data\\_A\\_Practical\\_Guide.pdf](http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf). Acesso em: jan. 2019.
- KOUTSOMITROPOULOS, D. A.; SOLOMOU, G. D. A learning object ontology repository to support annotation and discovery of educational resources using semantic thesauri. **IFLA Journal**, v. 44, n. 1, p. 4–22, 2018. Disponível em: <https://doi.org/10.1177/0340035217737559>. Acesso em: jan. 2019.
- L'ABATE, G. *et al.* Exposing vocabularies for soil as Linked Open Data. **Information Processing in Agriculture**, v. 2, n. 3–4, p. 208–216, 2015. Disponível em: <http://dx.doi.org/10.1016/j.inpa.2015.10.002>. Acesso em: abr. 2020.
- LAIPELT, R. do C. F. A análise de logs como estratégia para a realização da garantia do usuário. **Em Questão**, v. 21, n. n. 3, p. 150–170, 2015. Disponível em:



<https://doi.org/10.19132/1808-5245213.150-170>. Acesso em: jan. 2019.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. ed. Brasília: Brinquet de Lemos, 2004.

LAPORTE, M.-A. *et al.* A semantic web faceted search system for facilitating building of biodiversity and ecosystems services. In: GALHARDAS, H.; RAHM, E. (Org.). **Data Integration in the Life Sciences, DILS 2014**: Lecture Notes in Computer Science, vol 8574. Cham: Springer, 2014. p. 50–57. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-08590-6\\_5](https://link.springer.com/chapter/10.1007/978-3-319-08590-6_5). Acesso em: abr. 2020.

LAPORTE, M. A.; MOUGENOT, I.; GARNIER, E. ThesauForm-Traits: A web based collaborative tool to develop a thesaurus for plant functional diversity research. **Ecological Informatics**, v. 11, p. 34–44, 2012. Disponível em: <http://dx.doi.org/10.1016/j.ecoinf.2012.04.004>. Acesso em: abr. 2020.

LARA, M. L. G. Documentary languages and knowledge organization systems in the context of the semantic web. **Transinformacao**, v. 25, n. 2, p. 145–150, maio/ago. 2013. Disponível em: <https://doi.org/10.1590/S0103-37862013000200005>. Acesso em: abr. 2020.

LAUSER, B. **From thesauri to Ontologies**: a short case study in the food safety area in how ontologies are more powerful than thesauri: From thesauri to RDFS to OWL. Food and Agriculture Organization (FAO), 2004. [relatório]. Disponível em: <http://hdl.handle.net/10760/15686>. Acesso em: nov. 2019.

LAWRENCE, D. W. *et al.* Developing the injury prevention and safety promotion thesaurus, international English edition: An interdisciplinary tool for indexing and searching for research literature. Progress report 1. **Safety Science**, v. 44, n. 4, p. 279–296, 2006. Disponível em: <https://doi.org/10.1016/j.ssci.2005.10.006>. Acesso em: jan. 2019.

LEE, H. M.; HUANG, C. C.; CHAO, C. Y. Association thesaurus construction for interactive query expansion based on association rule mining. **Journal of Information Science and Engineering**, v. 23, n. 2, p. 617–627, 2007. Disponível em: [https://jise.iis.sinica.edu.tw/JISESearch/pages/View/PaperView.jsf?keyId=47\\_810](https://jise.iis.sinica.edu.tw/JISESearch/pages/View/PaperView.jsf?keyId=47_810). Acesso em: jan. 2019.

LEE, M.; BAILLIE, S.; ORO, J. D. TML : A Thesaural Markup Language. In: AUSTRALASIAN DOCUMENT COMPUTING SYMPOSIUM, 4., Coffs Harbour, Australia, 1999. **Proceedings...** Coffs Harbour, Australia, 1999. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.2101>. Acesso em: nov. 2019.

LI, C. H.; SONG, W.; PARK, S. C. An automatically constructed thesaurus for neural network based document categorization. **Expert Systems with Applications**, v. 36, n. 8, p. 10969–10975, 2009. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2009.02.006>. Acesso em: jan. 2019.

LIBRARY OF CONGRESS. **MARC 21 Format for Authority Data**. 1999. Disponível em: <https://www.loc.gov/marc/authority/>. Acesso em: maio 2018.

LIBRARY OF CONGRESS. **MARC in XML**. 2008. Disponível em: <https://www.loc.gov/marc/marcxml.html>. Acesso em: maio 2018.

LIBRARY OF CONGRESS. **Overview of the BIBFRAME 2.0 Model**. 2016. Disponível em: <https://www.loc.gov/bibframe/docs/bibframe2-model.html>. Acesso em: maio 2018.

LIBRARY OF CONGRESS. **Metadata Authority Description Schema**: official web site. 2018. Disponível em: <http://www.loc.gov/standards/mads/>. Acesso em: 3 maio. 2018.

LIEBESKIND, C.; DAGAN, I.; SCHLER, J. Semiautomatic Construction of Cross-Period Thesaurus. **Journal on Computing and Cultural Heritage**, v. 9, n. 4, dez., 2016. Disponível em: <https://doi.org/10.1145/2994151>. Acesso em: jan. 2019.

LIEBESKIND, C.; DAGAN, I.; SCHLER, J. Automatic Thesaurus Construction for Modern Hebrew. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 11., Miyazaki, Japan. **Proceedings...** Miyazaki, Japan: European Languages Resources Association (ELRA), 2018. Disponível em: <https://www.aclweb.org/anthology/L18-1229/>. Acesso em: jan. 2019.

LIMA, J. C. De; CARVALHO, C. L. de. **Uma Visão da Web Semântica**. 2004. Instituto de Informática, Universidade Federal de Goiás. [relatório]. Disponível em: [http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-04](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-04). Acesso em: jan. 2019.

LIMA, V. M. A. de et al. Estudos para implantação de ferramenta de apoio à gestão de linguagens documentárias: vocabulário controlado da USP. **Transinformação**, v. 18, n. 1, p. 17–25, jan. ,abr., 2006. Disponível em: <http://dx.doi.org/10.1590/S0103-37862006000100002>. Acesso em: jan. 2019.

LOSEE, R. M. Decisions in thesaurus construction and use. **Information Processing and Management**, v. 43, n. 4, p. 958–968, 2007. Disponível em: <https://doi.org/10.1016/j.landusepol.2014.12.017>. Acesso em: jan. 2019.

MACULAN, B. C. M. dos S. **Estudo e aplicação de metodologia para reengenharia de tesauro: remodelagem do THESAGRO**. 2015. Tese (Doutorado em Ciência da Informação) Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Minas Gerais, 2015. Disponível em: <https://repositorio.ufmg.br/handle/1843/BUBD-9ZKMUV>. Acesso em: jun. 2020.

MACULAN, B. C. M. dos S.; LIMA, G. A. B. de O. Buscando uma definição para o conceito de “conceito”. **Perspectivas em Ciência da Informação**, v. 22, n. 2, p. 54–87, 2017. Disponível em: <https://doi.org/10.1590/1981-5344/2963>. Acesso em: ago. 2020.

MACULAN, B. C. M. S. et al. Refinamento de relações em tesauros: reengenharia do THESAGRO. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., João Pessoa, PB. **Anais...** João Pessoa, PB: UFPB, 2015. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2810>. Acesso em: nov. 2019.

MARCONI, M.; LAKATOS, E. **Fundamentos de metodologia científica**. 5. ed. ed. São Paulo: Atlas, 2003. 310p.

MARI VÁLLEZ et al. Updating controlled vocabularies by analysing query logs. **Online Information Review**, v. 39, n. 7, p. 870–884, 2015. Disponível em: <https://doi.org/10.1108/OIR-06-2015-0180>. Acesso em: jan. 2019.

- MARTÍNEZ-GONZÁLEZ, M. M.; ALVITE-DÍEZ, M. The support of constructs in thesaurus tools from a Semantic Web perspective: framework to assess standard conformance. **Computer Standards & Interfaces**, v. 65, p. 79–91, July, 2019a. Disponível em: <https://doi.org/10.1016/j.csi.2019.02.003>. Acesso em: nov. 2019.
- MARTÍNEZ GONZÁLEZ, M. M.; ALVITE DÍEZ, M. L. Propuesta metodológica de evaluación de gestores de tesauros compatibles con la web semántica. **Anales de Documentación**, v. 17, n. 1, p. 1–18, 2014. Disponível em: <https://doi.org/10.6018/analesdoc.17.1.186271>. Acesso em: abr. 2020.
- MARTÍNEZ-GONZÁLEZ, M. M.; ALVITE-DÍEZ, M. L. Thesauri and Semantic Web: discussion of the evolution of thesauri toward their integration with the semantic web. **IEEE Access**, v. 7, p. 153151–153170, 2019b. Disponível em: <https://doi.org/10.1109/ACCESS.2019.2948028>. Acesso em: abr. 2020.
- MATTHEWS, B.; MILES, A.; WILSON, M. **Modelling thesauri for the Semantic Web**. Semantic Web Advanced Development for Europe (SWAD-Europe), 2009. Disponível em: <http://www.webcitation.org/5m2lmCyQY>. Acesso em: 1 maio. 2018.
- MAZIARZ, M.; PIASECKI, M. Towards mapping thesauri onto plWordNet. In: GLOBAL WORDNET CONFERENCE (GWC 2018), 9<sup>th</sup>., Singapore. **Proceedings...** Singapore: Nanyang Technological University (NTU), 2018. Disponível em: [http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018\\_paper\\_63.pdf](http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018_paper_63.pdf). Acesso em: abr. 2020.
- MCCULLOCH, E. Thesauri: Practical guidance for construction. **Library Review**, v. 54, n. 7, p. 403–409, 2005. Disponível: <https://doi.org/10.1108/00242530510611893>. Acesso em: abr. 2020.
- MEROÑO-PENUELA, A. et al. The MIDI linked data cloud. In: D'AMATO C. et al. (eds). **The Semantic Web: ISWC 2017**. Cham: Springer, 2017. p. 156-164. Lecture Notes in Computer Science, vol 10588. [https://doi.org/10.1007/978-3-319-68204-4\\_16](https://doi.org/10.1007/978-3-319-68204-4_16). Acesso em: jan. 2019.
- MEROÑO-PENUELA, A.; HOEKSTRA, R. The song remains the same: Lossless conversion and streaming of MIDI to RDF and back. In: SACK, H. et al. (Eds.). **The Semantic Web: ISWC 2016**. Cham: Springer, 2016. p. 194–199. Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, v. 9989. Disponível em: [https://doi.org/10.1007/978-3-319-47602-5\\_38](https://doi.org/10.1007/978-3-319-47602-5_38). Acesso em: jan. 2019.
- METABRAINZ FOUNDATION. **MusicBrainz**. 2019. Disponível em: <https://musicbrainz.org/>. Acesso em: 7 fev. 2019.
- MIMO. Musical Instrument Museums Online. **About MIMO**. 2019. Disponível em: <http://www.mimo-international.com/MIMO/about-mimo.aspx>. Acesso em: 4 fev. 2019.
- MOCHÓN-BEZARES, G.; MÉNDEZ-RODRÍGUEZ, E. M.; LA FUENTE, G. B. de. Conversión normalizada (SKOS) de sistemas de organización del conocimiento interoperables en la web. **Investigación Bibliotecológica: archivonomía, bibliotecología e información**, v. 34, n. 82, p. 53–86, 2020. Disponível em: <http://dx.doi.org/10.22201/iibi.24488321xe.2020.82.58071>. Acesso em: abr. 2020.

MOCHÓN, G.; MÉNDEZ, E. M.; BUENO DE LA FUENTE, G. 27 pawns ready for action: A multi-indicator methodology and evaluation of thesaurus management tools from a LOD perspective. **Library Hi Tech**, v. 35, n. 1, p. 99–119, 2017. Disponível em: <https://doi.org/10.1108/LHT-11-2016-0123>. Acesso em: abr. 2020.

MORAIS, M. W. de; RAMALHO, R. A. S. Representação em SKOS de um microtesauro de conhecimentos estratégicos nas organizações. **Perspectivas em Ciência da Informação**, v. 24, n. 4, p. 187–198, 2019. disponível em: <https://doi.org/10.1590/1981-5344/4193>. Acesso em: abr. 2020.

MOREIRA, M. P.; MOURA, M. A. Construindo tesouros a partir de tesouros existentes: a experiência do TCI - Tesouro em Ciência da Informação. **DataGramZero**, v. 7, n. 4, 2006. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/6670>. Acesso em: jan. 2019.

MORIN, E.; JACQUEMIN, C. Automatic acquisition and expansion of hypernym links. **Computers and the Humanities**, v. 38, n. 4, p. 363–369, 2004. Disponível em: <https://doi.org/10.1007/s10579-004-1926-2>. Acesso em: jan. 2019.

MORSHED, A. *et al.* Bridging end users' terms and AGROVOC concept server vocabularies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2010. **Proceedings...** Pittsburgh, Pennsylvania: Dublin Core Metadata Initiative, 2010. p. 186–189. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/1015>. Acesso em: abr. 2020.

MORSHED, A. *et al.* Thesaurus alignment for linked data publishing. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2011. **Proceedings...** The Hague: Dublin Core Metadata Initiative, 2011. p. 37–46, 2011. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3623>. Acesso em: abr. 2020.

MOURA, M. A. Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias. **Informação & Sociedade: Estudos**, v. 19, n. 1, p. 59–73, 2009. Disponível em: <http://periodicos.ufpb.br/ojs2/index.php/ies/article/view/2396>. Acesso em: jan. 2019.

MUSIC TECHNOLOGY GROUP. **AcousticBrainz**. 2019. Disponível em: <https://acousticbrainz.org/>. Acesso em: 7 fev. 2019.

NARUKAWA, Cristina Miyuki. **Estudo de vocabulário controlado na indexação automática**: aplicação no processo de indexação do Sistema de Indização Semiautomática (SISA). 2011. 222 f. Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2011. Disponível em: <http://hdl.handle.net/11449/93677>. Acesso em: jan. 2019.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.50-2003 (S2014) Information Retrieval: Application Service Definition & Protocol Specification**. Maryland, USA: NISO, 2014.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005 (R2010): Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies**. Baltimore, USA: NISO, 2010.

NEUBERT, J. Linked data based library web services for economics. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2012.

**Proceedings...** Kuching, Sarawak, Malaysia: Dublin Core Metadata Initiative, 2012. p. 12–22. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3659>. Acesso em: abr. 2020.

NEUBERT, J. Wikidata as a linking hub for knowledge organization systems? Integrating an authority mapping into Wikidata and learning lessons for KOS mappings. In: EUROPEAN NETWORKED KNOWLEDGE ORGANIZATION SYSTEMS (NKOS) WORKSHOP, 17., 2017. **Proceedings...** Thessaloniki, Greece: CEUR, 2017. p. 1–12. Disponível em: <http://ceur-ws.org/Vol-1937/paper2.pdf>. Acesso em: abr. 2020.

NIININEN, S.; NYKYRI, S.; SUOMINEN, O. The future of metadata: open, linked, and multilingual – the YSO case. **Journal of Documentation**, v. 73, n. 3, p. 451–465, 2017. Disponível em: <https://doi.org/10.1108/JD-06-2016-0084>. Acesso em: abr. 2020.

NOHAMA, P. *et al.* Quality issues in Thesaurus building: a case study from the medical domain. **Revista Brasileira de Engenharia Biomédica**, v. 28, n. 1, p. 11–22, 2012. Disponível em: <http://rbejournal.org/doi/10.4322/rbeb.2012.002>. Acesso em: abr. 2020.

NORUZI, A. Folks thesauri or search thesauri: Why semantic search engines need Folks Thesauri? **Webology**, v. 15, n. 2, p. 1–10, 2018. Disponível em: <http://www.webology.org/2018/v15n2/editorial26.pdf>. Acesso em: abr. 2020.

NOVIANTI; NUGROHO, I. H. T. SKOSYN: Simple Knowledge Organization System for synonym search on content management system. In: ASIA-PACIFIC CONFERENCE ON COMPUTER AIDED SYSTEM ENGINEERING, APCASE 2014. **Proceedings...** South Kuta, Indonesia: IEEE, 2014. p. 25–30, 2014. Disponível em: <https://doi.org/10.1109/APCASE.2014.6924466> Acesso em: jan. 2019.

NURMIKKO-FULLER, T. *et al.* In Collaboration with In Concert: Reflecting a Digital Library as Linked Data for Performance Ephemera. In: INTERNATIONAL WORKSHOP ON DIGITAL LIBRARIES FOR MUSICOLOGY, 3<sup>rd</sup>., aug., 2016. **Proceedings...** New York: Association for Computing Machinery, 2016. p. 17–24. Disponível em: <https://doi.org/10.1145/2970044.2970049>. Acesso em: jan. 2019.

OPEN ARCHIVES INITIATIVE. **The Open Archives Initiative Protocol for Metadata Harvesting**. 2015. Disponível em: <http://www.openarchives.org/OAI/openarchivesprotocol.html>. Acesso em: 3 jun. 2020.

OPEN KNOWLEDGE FOUNDATION. **Open Data Commons: Legal tools for Open Data**. 2008. Disponível em: <https://opendatacommons.org/index.html>. Acesso em: 23 jan. 2019.

OPEN KNOWLEDGE INTERNATIONAL. **The Open Definition: Conformant Licenses**. 2015. Disponível em: <https://opendefinition.org/licenses/>. Acesso em: 23 jan. 2019.

PAGE, K. R. *et al.* Semantics for music analysis through linked data: How country is my country? In: IEEE INTERNATIONAL CONFERENCE ON E-SCIENCE, 6<sup>th</sup>., Brisbane, QLD, Australia, 2010. **Proceedings...** Brisbane, QLD, Australia: IEEE, 2010. p. 41–48, 2010. Disponível em: <https://doi.org/10.1109/eScience.2010.49>. Acesso em: jan. 2019.



PAGE, K. R. et al. Capturing the workflows of music information retrieval for repeatability and reuse. **Journal of Intelligent Information Systems**, v. 41, n. 3, p. 435–459, 2013. Disponível em: <https://doi.org/10.1007/s10844-013-0260-9>. Acesso em: jan. 2019.

PARK, Y. C.; CHOI, K.-S. Automatic thesaurus construction using Bayesian networks. **Information Processing & Management**, v. 32, n. 5, p. 543–553, setp., 1996. Disponível em: [https://doi.org/10.1016/0306-4573\(96\)00026-X](https://doi.org/10.1016/0306-4573(96)00026-X).

PASTOR-SÁNCHEZ, J.-A. Mercado semântico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. **Scire: representación y organización del conocimiento**, v. 19, n. 2, jul. /dic., p. 55–68, 2013. Disponível em: <https://www.iberid.eu/ojs/index.php/scire/article/view/4112>. Acesso em: abr. 2020.

PASTOR-SÁNCHEZ, J.-A. Proposal to represent the UNESCO Thesaurus for the semantic web applying ISO-25964. **Brazilian Journal of Information Science**, v. 9, n. 2, p. 1–8, 2015. Disponível em: <https://doi.org/10.36311/1981-1640.2015.v9n2.01.p1>. Acesso em: jan. 2010.

PASTOR-SÁNCHEZ, J.-A.; MARTÍNEZ-MÉNDEZ, F.-J.; RODRÍGUEZ-MUÑOZ, J.-V. Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. **El profesional de la información**, v. 21, n. 3, p. 245–253, 2012. Disponível em: <http://dx.doi.org/10.3145/epi.2012.may.04>. Acesso em: abr. 2020.

PASTOR-SANCHEZ, J.-A.; MARTÍNEZ-MENDEZ, F. J.; RODRIGUEZ-MUÑOZ, J. V. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. **Information Research**, v. 14, n. 4, p. 1–15, 2009. Disponível em: <http://www.informationr.net/ir/14-4/paper422.html>. Acesso em: jan. 2019.

PASTOR-SÁNCHEZ, J. A.; LÓPEZ-CARREÑO, R.; MARTÍNEZ-MÉNDEZ, J. F. Advantages of SKOS for providing semantic stability to the information resources: an example of use in a collaborative institutional web portal. In: ISKO INTERNATIONAL CONFERENCE, 11th., Rome. **Proceedings...** Rome: ISKO, 2010. Disponível em: <http://eprints.rclis.org/24261/>. Acesso em: jan. 2019.

PASTOR SÁNCHEZ, J. A. **Diseño de un sistema colaborativo para la creación y gestión de tesauros en internet basados en SKOS**. p. 296, 2009. Tese. Universidad de Murcia. Espanha. Disponível em: <http://hdl.handle.net/10201/3946>. Acesso em: abr. 2020.

PATRÍCIO, H. S. A Europeia e a agregação de metadados na web: análise dos esquemas ESE/EDM e da aplicação de standards da web semântica a dados de bibliotecas. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, n. 11, Lisboa, Fundação Calouste Gulbenkian, 18, 19 e 20 de outubro de 2012. **Anais...** Lisboa: Fundação Calouste Gulbenkian, 2012. Disponível em: <http://bad.pt/publicacoes/index.php/congressosbad/article/view/458>. Acesso em: jan. 2019.

PATTUELLI, C. **Linked Jazz Project**. 2016. Disponível em: <https://linkedjazz.org/>. Acesso em: fev. 2019.

PATTUELLI, M. C. et al. Linked Jazz 52nd Street: a LOD crowdsourcing tool to reveal connections among jazz artists. In: DIGITAL HUMANITIES, Lincoln, 2013. **Proceedings...**

Lincoln: University of Nebraska, 2013. Disponível em: <http://dh2013.unl.edu/abstracts/ab-254.html>. Acesso em: jan. 2019.

PETRAKI, E.; KAPETIS, C.; YANNAKOUDAKIS, E. J. Automated thesaurus population and management. **Qualitative and Quantitative Methods in Libraries**, v. 4, p. 181–189, 2015. Disponível em: <http://www.qqml-journal.net/index.php/qqml/article/view/238>. Acesso em: jan. 2019.

PETRINI, M.; POZZEBON, M. Usando Grounded Theory Na Construção De Modelos Teóricos. **Revista Gestão e Planejamento**, v. 10, n. 1, p. 1–18, 2009. Disponível em: <https://revistas.unifacs.br/index.php/rgb/article/view/990>. Acesso em: jan. 2019.

PICKERING, R. **The Music Ontology**. 2012. Disponível em: <http://musicontology.com/>. Acesso em: 7 fev. 2019.

PICKLER, M. E. V. Web Semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, v. 12, n. 1, p. 65–83, jan. /abr., 2007. Disponível em: <https://doi.org/10.1590/S1413-99362007000100006>. Acesso em: jan. 2019.

PINHO, F. A. Metafiltro para controle terminológico de metáforas no domínio da homossexualidade masculina. **Ciência da Informação**, v. 41, n. 1, p. 120–133, jan. /abr., 2014. Disponível em: <http://revista.ibict.br/ciinf/article/view/1422/1600>. Acesso em: jan. 2019.

PIOTROWSKI, M.; COLAVIZZA, G.; THIERY, F. The Labeling System: A New Approach to Overcome the Vocabulary Bottleneck. In: DH-CASE '14: Collaborative Annotations in Shared Environments: metadata, tools and techniques in the Digital Humanities, Fort Collins, CA, september 2014. **Proceedings...** New York: Association for Computing Machinery, 2014. Disponível em: <https://doi.org/10.1145/2657480.2657482>. Acesso em: abr. 2020.

PIRES, M. T. **Gruia de dados abertos**. São Paulo: Centro de Estudos sobre Tecnologia Web, 2015. Disponível em: [https://nic.br/media/docs/publicacoes/13/Guia\\_Dados\\_Abertos.pdf](https://nic.br/media/docs/publicacoes/13/Guia_Dados_Abertos.pdf). Acesso em: jun. 2020.

PIZZANI, L. et al. A arte da pesquisa bibliográfica na busca do conhecimento. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 10, n. 1, p. 53, 2012. Disponível em: <https://doi.org/10.20396/rdbci.v10i1.1896>. Acesso em: jan. 2019.

PLINI, P. et al. A Joint Initiative to Support the Semantic Interoperability within the GIIDA Project. In: GI4DM 2010 CONFERENCE: GEOMATICS FOR CRISIS MANAGEMENT (Gi4DM 2010), Torino, Italy, feb., 2010. **Proceedings...** Torino: ASITA, 2010. Disponível em: <https://arxiv.org/ftp/arxiv/papers/1012/1012.0160.pdf>. Acesso em: jan. 2019.

PU, H. T.; CHIEN, L. F. Integrating log-based and text-based methods towards automatic Web thesaurus construction. In: ASIST ANNUAL MEETING, v. 41, 2004. **Proceedings...** Providence, RI: ASIST, 2004. p. 463–471. Disponível em: <https://doi.org/10.1002/meet.1450410154>. Acesso em: jan. 2019.

QI, J.; LUO, D.; ZHU, L. Research on the collaborative method for vocabulary system constructions. In: INTERNATIONAL CONFERENCE ON COMPUTER AND AUTOMATION ENGINEERING, ICCAE 2010, 2., April 2010. **Proceedings...** Singapore:

IEEE, 2010. p. 523–526. Disponível em: <https://doi.org/10.1109/ICCAE.2010.5451268>. Acesso em: abr. 2020.

RAIMOND, Y. et al. Use of Semantic Web Technologies in Natural language interface to the BBC Web Sites. In: W3C. **Semantic Web Use Cases and Case Studies**. London: BBC, 2012. Disponível em: <https://www.w3.org/2001/sw/sweo/public/UseCases/BBC/>. Acesso em: jan. 2019.

RAIMOND, Y.; ABDALLAH, S.; SANDLER, M. The Music Ontology. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 8<sup>th</sup>., Vienna, Austria. **Proceedings...** Vienna, Austria: Austrian Computer Society, 2007. Disponível em: [http://ismir2007.ismir.net/proceedings/ISMIR2007\\_p417\\_raimond.pdf](http://ismir2007.ismir.net/proceedings/ISMIR2007_p417_raimond.pdf). Acesso em: jan. 2019.

RAJBHANDARI, S.; KEIZER, J. The AGROVOC Concept Scheme - A Walkthrough. **Journal of Integrative Agriculture**, v. 11, n. 5, may, 2012. Disponível em: [http://dx.doi.org/10.1016/S2095-3119\(12\)60058-6](http://dx.doi.org/10.1016/S2095-3119(12)60058-6). Acesso em: nov. 2019.

RAMALHO, R. A. S. Análise do modelo de dados SKOS: sistema de organização do conhecimento simples para a Web. **Informação & Tecnologia**, v. 2, n. 1, p. 66–79, 2015. Disponível em: <http://periodicos.ufpb.br/ojs2/index.php/itec/article/view/25995/14480>. Acesso em: jan. 2019.

RANGANATHAN, S. R. **Prolegomena to library classification**. 3. ed. Bombay: Asia Publ. House, 1967.

RAUTENBERG, S. et al. **Guia prático para publicação de dados abertos conectados na web**. Curitiba: Appris, 2018.

RDA STEERING COMMITTEE. **Resource Description and Access**. 2016. Disponível em: <https://www.rdatoolkit.org/rdaessentials>. Acesso em: maio 2018.

REES-POTTER, L. K. Dynamic thesaural systems: A bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. **Information Processing and Management**, v. 25, n. 6, p. 677–689, 1989. Disponível em: [https://doi.org/10.1016/0306-4573\(89\)90101-5](https://doi.org/10.1016/0306-4573(89)90101-5). Acesso em: jan. 2019.

RODRÍGUEZ-DONCEL, V.; GÓMEZ-PÉREZ, A.; MIHINDUKULASOORIYA, N. Rights declaration in Linked Data. In: INTERNATIONAL CONFERENCE ON CONSUMING LINKED DATA, 4., Sydney, 2013. **Proceedings...** New York: ACM, 2013. Disponível em: [http://ceur-ws.org/Vol-1034/RodriguezDoncelEtAl\\_COLD2013.pdf](http://ceur-ws.org/Vol-1034/RodriguezDoncelEtAl_COLD2013.pdf). Acesso em: jan. 2019.

RONCATO, L. A Grounded Theory: Teoria Fundamentada em Dados. In: SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 38., Goiânia, 2006. **Anais...** Goiânia: SBPO, 2006. Disponível em: <http://www.din.uem.br/sbpo/sbpo2006/pdf/arq0133.pdf>. Acesso em: jan. 2019.

RONDEAU, G. **Introduction à la terminologie**. 2. ed. Québec, Canadá: Gaëtan Morin, 1984.

ROSATI, I. et al. A thesaurus for phytoplankton trait-based approaches: development and applicability. **Ecological Informatics**, v. 42, p. 129–138, 2017. Disponível em:



<https://doi.org/10.1016/j.ecoinf.2017.10.014>. Acesso em: abr. 2020.

RUGE, G. Experiments on linguistically-based term associations. **Information Processing and Management**, v. 28, n. 3, p. 317–332, 1992. Disponível em: [https://doi.org/10.1016/0306-4573\(92\)90078-E](https://doi.org/10.1016/0306-4573(92)90078-E). Acesso em: jan. 2019.

SÁ-SILVA, J. R.; ALMEIDA, C. D. de; GUINDANI, J. F. Pesquisa documental: pistas teóricas e metodológicas. **Revista Brasileira de História & Ciências Sociais**, v. 1, n. 1, p. 1–15, 2009. Disponível em: <https://periodicos.furg.br/rbhcs/article/view/10351>. Acesso em: jan. 2019.

SANTIS, R. de. **Sistemas de Organização do Conhecimento para Domínios Complexos: abordagem a canções populares na web semântica utilizando propriedades fuzzy**. 2016. Tese (doutorado) – Universidade Federal do Rio de Janeiro, Escola de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós Graduação em Ciência da Informação, 2016. Disponível em: <http://ridi.ibict.br/handle/123456789/879>.

SANTOS, C. A. et al. Sistema de gestão para linguagem documentária: metadados e rede colaborativa no vocabulário controlado do SISBI/USP. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 16., Rio de Janeiro, 2010. **Anais...** Rio de Janeiro: SNBU, 2010. Disponível em: <https://repositorio.usp.br/item/002133890>. Acesso em: jan. 2019.

SANTOS, C. A. C. M. dos et al. Elaboração de vocabulário controlado em formato SKOS usando Tematres: Implicações metodológicas e web-semântica. In: Complexidade e Organização do Conhecimento: Desafios do nosso século. Rio de Janeiro: Sociedade Brasileira de Organização do Conhecimento, 2013. p. 205–210.

SANTOS, J. C. F. dos; MOREIRA, W. SKOS: uma análise sobre as abordagens e suas aplicações na Ciência da Informação. **Informação & Informação**, v. 23, n. 3, p. 362–389, set. /dez., 2018. Disponível em: <http://dx.doi.org/10.5433/1981-8920.2018v23n3p362>. Acesso em: out. 2019.

SANTOS, J. L. G. dos et al. Perspectivas metodológicas para o uso da teoria fundamentada nos dados na pesquisa em enfermagem e saúde. **Escola Anna Nery - Revista de Enfermagem**, v. 20, n. 3, p. 1–8, 2016. Disponível em: <https://doi.org/10.5935/1414-8145.20160056>. Acesso em: abr. 2019.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41–62, 1996. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>. Acesso em: jan. 2019.

SCHEDL, M.; FLEXER, A.; URBANO, J. The neglected user in music information retrieval research. **Journal of Intelligent Information Systems**, v. 41, n. 3, p. 523–539, 2013. Disponível em: <https://doi.org/10.1007/s10844-013-0247-6>. Acesso em: jan. 2019.

SCHEDL, M.; GÓMEZ, E.; URBANO, J. **Music Information Retrieval: Recent Developments and Applications**. Boston: Now, 2014. Disponível em: <https://doi.org/10.1561/15000000042>. Acesso em: dez. 2018.

SCHNEIDER, J. W.; BORLUND, P. Introduction to bibliometrics for construction and maintenance of thesauri: Methodical considerations. **Journal of Documentation**, v. 60, n. 5,

p. 524–549, 2004. Disponível em: <https://doi.org/10.1108/00220410410560609>. Acesso em: jan. 2019.

SCHOLZ, A. H. **Europeana Publishing Guide v1.8**: a guide to the metadata and content requirements for data partners publishing material in Europeana Collections. The Hague, Netherlands: Europeana Publishing, 2019. Disponível em: <https://pro.europeana.eu/post/publication-policy>. Acesso em: nov. 2019.

SCOZ, D. Web-Mining Conceitos E Aplicações. **Revista Eletrônica do Alto Vale do Itajaí**, v. 3, n. 4, p. 82–86, 2014. Disponível em: <https://doi.org/10.5965/2316419003042014082>. Acesso em: jun. 2020.

SEGUNDO, J. E. S.; CONEGLIAN, C. S.; LUCAS, E. R. O. Conceitos e tecnologias da Web semântica no contexto da colaboração acadêmico-científica: um estudo da plataforma Vivo. **Transinformação**, v. 29, n. 3, p. 297–309, 2017. Disponível em: <https://doi.org/10.1590/2318-08892017000300007>. Acesso em: jan. 2019.

SEVERINO, A. J. **Metodologia do trabalho científico**. 24 ed. rev. São Paulo: Cortez, 2016.

SHOLIKAH, R. W. et al. Co-occurrence technique and dictionary based method for Indonesian thesaurus construction. In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGY, 5<sup>th</sup>., Malacca City, Malaysia. **Proceedings...** Malacca City, Malaysia: IEEE, 2017. Disponível em: <https://doi.org/10.1109/ICoICT.2017.8074649>. Acesso em: jan. 2019.

SILVA, A. G. **Informação legislativa ao alcance do cidadão** : contribuição dos Sistemas de Organização do Conhecimento. 2015. 229 p. Tese (doutorado) - Programa de Pós-Graduação em Ciência da Informação - Escola de Comunicações e Artes, Universidade de São Paulo, 2015., 2015. Disponível em: [www.teses.usp.br/teses/disponiveis/27/27151/tde-18052015-155605](http://www.teses.usp.br/teses/disponiveis/27/27151/tde-18052015-155605). Acesso em: jan. 2019.

SILVA, E. L. da; MENEZES, E. M. **Metodologia da Pesquisa e Elaboração de Dissertação**. 4. ed. rev ed. Florianópolis: UFSC, 2005.

SIMÕES, M. da G. **Da abstracção à complexidade formal**: relações conceptuais num tesouro. Coimbra: Almedina, 2008.

SIMOU, N.; CHORTARAS, A.; STAMOU, G. Enriching and Publishing Cultural Heritage as Linked Open Data. In: IOANNIDES, M.; MAGNENAT-THALMANN, N.; PAPAGIANNAKIS, G. (Ed.). **Mixed Reality and Gamification for Cultural Heritage**. ChamCham: Springer, 2017. p. 201–223. Disponível em: [https://doi.org/10.1007/978-3-319-49607-8\\_7](https://doi.org/10.1007/978-3-319-49607-8_7). Acesso em: jan. 2019.

SMADJA, F. Retrieving Collocations from Text: Xtract. **Computational linguistics**, v. 19, n. 1, p. 143, 1993. Disponível em: <https://dl.acm.org/doi/10.5555/972450.972458>. Acesso em: jan. 2019.

SOERGEL, D. Thesauri and ontologies in digital libraries:1. Structure and use in knowledge-based assistance to users. In: ACM INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES, 2., july, Portland, Oregon, 2002. **Proceedings..** New York: ACM, 2002. p. 415. Disponível em: <https://doi.org/10.1145/544220.544364>. Acesso em: jan. 2019.

SOERGEL, D. et al. Reengineering thesauri for new applications: The AGROVOC example. **Journal of Digital Information**, v. 4, n. 4, p. 1–23, 2004. Disponível em: <http://www.fao.org/3/af234e/af234e00.htm>. Acesso em: jan. 2019.

SOLOMOU, G.; PAPTAEODOROU, T. The use of SKOS vocabularies in digital repositories: The DSpace case. In: IEEE INTERNATIONAL CONFERENCE ON SEMANTIC COMPUTING, 4<sup>th</sup>., Pittsburgh, PA, USA, 2010. **Proceedings...** Pittsburgh, PA: IEEE, 2010. p. 542–547. Disponível em: <https://doi.org/10.1109/ICSC.2010.83>. Acesso em: abr. 2019.

SONG, W. et al. Intelligent information retrieval system using automatic thesaurus construction. **International Journal of General Systems**, v. 40, n. 4, p. 395–415, 2011. Disponível em: <https://doi.org/10.1080/03081079.2010.530026>. Acesso em: jan. 2019.

SOTOMAYOR, M.; VELOZ, F. Thesaurus-based named entity recognition system for detecting spatio-temporal crime events in Spanish language from Twitter. In: 2017 IEEE 2nd Ecuador Technical Chapters Meeting, 2<sup>nd</sup>., Salinas, Ecuador, 2017. **Proceedings...** New York: IEEE, 2017. p. 1–5. Disponível em: <https://doi.org/10.1109/ETCM.2017.8247537>. Acesso em: jan. 2019.

SOUZA, S. da S. de; SILVA, D. M. G. V. da. Validação de modelo teórico: conhecendo os processos interativos na rede de apoio às pessoas com tuberculose TT. **Acta paul. enferm**, v. 24, n. 6, p. 778–783, 2011. Disponível em: <https://doi.org/10.1590/S0103-21002011000600008>. Acesso em: abr. 2019.

SPAMPINATO, D.; ZANGARA, I. Classical antiquity and semantic content management on linked open data. In: INTERNATIONAL WORKSHOP ON COLLABORATIVE ANNOTATIONS IN SHARED ENVIRONMENT, 1., Florence, Italy, 2013. **Proceedings...** New York: ACM, 2013. Disponível em: <https://doi.org/10.1145/2517978.2517992>. Acesso em: abr. 2019.

SPITERI, L. F. Word Association Testing and Thesaurus Construction : A Pilot Study Word Association Testing and Thesaurus Construction : A Pilot Study. **Cataloging & Classification Quarterly**, v. 40, n. 1, p. 55–78, 2005. Disponível em: [https://doi.org/10.1300/J104v40n01\\_05](https://doi.org/10.1300/J104v40n01_05). Acesso em: abr. 2019.

STELLATO, A. *et al.* A Collaborative Framework for Managing and Publishing KOS. In: EUROPEAN NETWORKED KNOWLEDGE ORGANISATION SYSTEMS (NKOS) WORKSHOP, 10., Berlin, 28-29 September 2011. **Proceedings...** Berlin, 2011. Disponível em: <http://www.fao.org/docrep/article/am814e.pdf>. Acesso em: abr. 2020.

STELLATO, A. *et al.* VocBench: A Web Application for Collaborative Development of Multilingual Thesauri. In: GANDON, F. et al. (eds). **The Semantic Web, Latest Advances and New Domains, ESWC 2015: Lecture Notes in Computer Science**, vol 9088. Cham: Springer, 2015a. p. 38–53. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-18818-8\\_3](https://link.springer.com/chapter/10.1007/978-3-319-18818-8_3). Acesso em: abr. 2020.

STELLATO, A. Dictionary, Thesaurus or Ontology? Disentangling Our Choices in the Semantic Web Jungle. **Journal of Integrative Agriculture**, v. 11, n. 5, p. 710–719, 2012. Disponível em: [http://dx.doi.org/10.1016/S2095-3119\(12\)60060-4](http://dx.doi.org/10.1016/S2095-3119(12)60060-4). Acesso em: abr. 2020.

STRAUSS, A.; CORBIN, J. **Pesquisa qualitativa: técnicas e procedimentos para o desenvolvimento de Teoria Fundamentada**. 2. ed. ed. Porto Alegre: Artmed, 2008.

SUMMERS, E. et al. LCSH, SKOS and Linked Data. In: DCMI INTERNATIONAL CONFERENCE DUBLIN CORE AND METADATA APPLICATIONS, Berlin, setp., 2008. **Proceedings...** Berlin: DCMI, 2008. p. 25–33. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/916>. Acesso em: jan. 2019.

SVENONIUS, E. Subject languages: introduction, vocabulary selection, and Classification. In: **The Intellectual Foundation of Information Organization**. London: MIT Press, 2001. p. 127–146. Disponível em: <https://doi.org/10.7551/mitpress/3828.003.0010>. Acesso em: out. 2020.

TAYLOR, M. **Zthes: a Z39.50 Profile for Thesaurus Navigation**. 2000. Disponível em: <https://www.loc.gov/z3950/agency/profiles/zthes-04.html>. Acesso em: nov. 2019.

THE OBO FOUNDRY. **NCI Thesaurus OBO Edition**. 2019. Disponível em: <http://www.obofoundry.org/ontology/ncit.html>. Acesso em: nov. 2019.

THIOLLENT, M. **Metodologia da pesquisa-ação**. São Paulo: Cortez, 2009.

TORRE-BASTIDA, A.-I.; GONZÁLEZ-RODRÍGUEZ, M.; VILLAR-RODRÍGUEZ, E. Datos abiertos enlazados (LOD) y su implantación en bibliotecas: iniciativas y tecnologías. **El Profesional de la Información**, v. 24, n. 2, p. 113, 2015. Disponível em: <http://profesionaldelainformacion.com/contenidos/2015/mar/04.html>. Acesso em: abr. 2020.

TRISTÃO, Ana Maria Delazari; FACHIN, Gleisy Regina Bóries; ALARCON, Orestes Estevam. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, vol.33, n.2, p.161-171, 2004. <https://doi.org/10.1590/S0100-19652004000200017>.

TSENG, Y. H. Automatic thesaurus generation for Chinese documents. **Journal of the American Society for Information Science and Technology**, v. 53, n. 13, p. 1130–1138, 2002. Disponível em: <https://doi.org/10.1002/asi.10146>. Acesso em: dez. 2018.

TUDHOPE, D.; BINDING, C. Experiences with Knowledge Organization System Services from the STAR Project. **Signum**, n. 5, p. 1–9, 2008. Disponível em: <https://journal.fi/signum/article/view/3483>. Acesso em: abr. 2020.

TUDORACHE, T. et al. Ontology Development for the Masses: Creating ICD-11 in WebProtégé. In: CIMIANO, P.; PINTO, H. S. (Ed.). **Knowledge Engineering and Management by the Masses: EKAW 2010**. Cham: Springer, 2010. p. 74–89. Lecture Notes in Artificial Intelligence, v. 6317. Disponível em: [https://doi.org/10.1007/978-3-642-16438-5\\_6](https://doi.org/10.1007/978-3-642-16438-5_6). Acesso em: jan. 2019.

UCLA DIGITAL LIBRARY PROGRAM. **Sheet Music Consortium**. 2011. Disponível em: <http://digital2.library.ucla.edu/sheetmusic/index.html>. Acesso em: 13 fev. 2019.

VAN HOOLAND, S. et al. Evaluating the Success of Vocabulary Reconciliation for Cultural Heritage Collections Seth. **Journal of the American Society for Information Science and Technology**, v. 64, n. 3, p. 464–479, 2013. Disponível em: <https://doi.org/10.1002/asi.22763>. Acesso em: jan. 2019.

VAN HOOLAND, S.; VERBORGH, R. **Linked Data for Libraries, archives and museums: how to clean, link and publish your metadata**. London: Facet, 2015.

VANDERBILT, K. et al. A prototype system for multilingual data discovery of International Long-Term Ecological Research (ILTER) Network data. **Ecological Informatics**, v. 40, n. 2017, p. 93–101, 2017. Disponível em: <http://dx.doi.org/10.1016/j.ecoinf.2016.11.011>. Acesso em: jan. 2019.

VICKERY, B. C. Thesaurus: a new word in documentation. **Journal of Documentation**, v. 16, n. 4, p. 181–189, 1960. Disponível em: <https://doi.org/10.1108/eb026286>. Acesso em: jan. 2019.

VITI, E. My First Ten Years: Nuovo Soggettario growing, development and integration with other Knowledge Organization Systems. **Knowledge Organization**, v. 44, n. 8, p. 624–637, 2018. Disponível em: <https://doi.org/10.5771/0943-7444-2017-8-624>. Acesso em: abr. 2020.

VOSS, J. Collaborative thesaurus tagging the Wikipedia way. **Arxiv preprint**, v. 1, n. 1, p. 7, 2006. Disponível em: <http://arxiv.org/abs/cs/0604036>. Acesso em: jan. 2019.

W3C. **Thesaurus Links**. SWAD-Europe Thesaurus Activity, 2004. Disponível em: [https://www.w3.org/2001/sw/Europe/reports/thes/thes\\_links.html](https://www.w3.org/2001/sw/Europe/reports/thes/thes_links.html). Acesso em: out. 2019.

W3C. **SKOS Core Vocabulary**. Cambridge, 2005a. Disponível em: <https://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/>. Acesso em: maio. 2018.

W3C. **Quick guide to publishing a thesaurus on the semantic web**. Cambridge, 2005b. Disponível em: <http://www.w3.org/TR/swbp-thesaurus-pubguide>. Acesso em: jan. 2019.

W3C. **Semantic Web Deployment Working Group (SWDWG) Charter**. Cambridge, 2006. Disponível em: <https://www.w3.org/2006/07/swdwg-charter>. Acesso em: 19 nov. 2019.

W3C. **Best Practice Recipes for Publishing RDF Vocabularies**. Cambridge, 2008a. Disponível em: <https://www.w3.org/TR/swbp-vocab-pub/>. Acesso em: jan. 2019.

W3C. **SPARQL Query LangUage for RDF**. Cambridge, 2008b. Disponível em: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>. Acesso em: maio 2018.

W3C. **Cool URIs for the Semantic Web**. Cambridge, 2008c. Disponível em: <https://www.w3.org/TR/cooluris/>. Acesso em: jun. 2020.

W3C. **SKOS Simple Knowledge Organization System Reference**. Cambridge, 2009a. Disponível em: <https://www.w3.org/TR/skos-reference/>. Acesso em: maio. 2018.

W3C. **SKOS: Software Tools & Applications**. Cambridge, 2009b. Disponível em: <https://www.w3.org/2001/sw/wiki/SKOS>. Acesso em: jun. 2020.

W3C. **SPARQL 1.1 Protocol**. Cambridge, 2013a. Disponível em: <https://www.w3.org/TR/2013/REC-sparql11-protocol-20130321/>. Acesso em: set. 2020.

W3C. **SPARQL 1.1 Federated Query**. Cambridge, 2013b. Disponível em: <https://www.w3.org/TR/sparql11-federated-query/>. Acesso em: jan. 2019.



- W3C. **RDF 1.1 Primer**. Cambridge, 2014a. Disponível em: <https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>. Acesso em: maio 2018.
- W3C. **RDF 1.1 Concepts and Abstract Syntax**. Cambridge, 2014b. Disponível em: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>. Acesso em: maio 2018.
- W3C. **RDF 1.1 Semantics**. Cambridge, 2014c. Disponível em: <https://www.w3.org/TR/rdf11-mt/>. Acesso em: 4 maio. 2018c.
- W3C. **Best Practices for Publishing Linked Data**. Cambridge, 2014d. Disponível em: <https://www.w3.org/TR/ld-bp/#VOCABULARIES>. Acesso em: 26 jun. 2020.
- W3C. **Extensible Markup Language (XML)**. Cambridge, 2016a. Disponível em: <http://www.w3.org/XML>. Acesso em: maio 2018.
- W3C. **Data on the Web Best Practices: Dataset Usage Vocabulary**. Cambridge, 2016b. Disponível em: <https://www.w3.org/TR/vocab-duv/>. Acesso em: jun. 2020.
- W3C. **Data on the web best practices**. Cambridge, 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: ago. 2020.
- W3C. **About W3C**. Cambridge, 2018. Disponível em: <https://www.w3.org/>. Acesso em: nov. 2018.
- WATSON, K. **Beginning C# 2005 databases**. Indianapolis, IN: Wiley Publishing, 2006.
- WETZ, P. *et al.* Matching linked open data entities to local thesaurus concepts. In: SEMANTICS, I., Posters & Demonstrations Track. Graz, Austria, September 5-7, 2012. **Proceedings...** Europe: CEUR Workshop Proceedings, 2012. p. 6–11. Disponível em: <http://ceur-ws.org/Vol-932/paper2.pdf>. Acesso em: abr. 2020.
- WIKIMEDIA PROJECT. **Wikidata**. 2019. Disponível em: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page). Acesso em: abr. 2019.
- WOELFER, J. P.; LEE, J. H. The role of music in the lives of homeless young people: a preliminary report. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE, 13<sup>th</sup>., Mosteiro S.Bento Da Vitória, Porto, Portugal, october 8-12, 2012. **Proceedings...** Canadá: ISMIR, 2012. p. 955–958, Disponível em: <http://doi.acm.org/10.1145/2212776.2212874>.
- WOODS, W. a. **Conceptual Indexing: a better way to organize knowledge**. CA, United States: Sun Microsystems, 1997. Disponível em: <https://dl.acm.org/doi/pdf/10.5555/974965>. Acesso em: 20 nov. 2019.
- WRIGHT, D. G.; HARRISON, K. A.; WATKINS, J. Automated tagging of environmental data using a novel SKOS formatted environmental thesaurus. **Earth Sci. Inform**, v. 8, p. 103–110, 2015. Disponível em: <https://link.springer.com/article/10.1007/s12145-014-0183-1>. Acesso em: abr. 2020.
- XIAN, G.; SUN, W. Open, interlink and mashup multilingual agricultural data semantically based on KOSs. In: INTERNATIONAL CONFERENCE ON SYSTEMS AND INFORMATICS, ICSAI 2016, 3., Shanghai, China, 2017. **Proceedings...** Shanghai, China:

IEEE, 2017. p. 405–409, 2017. Disponível em: <https://doi.org/10.1109/ICSAI.2016.7810989>. Acesso em: abr. 2020.

XU, H.; YU, B. Automatic thesaurus construction for spam filtering using revised back propagation neural network. **Expert Systems with Applications**, v. 37, n. 1, p. 18–23, 2010. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2009.02.059>. Acesso em: jan. 2019.

YE, Z. et al. Exploring social annotation tags to enhance information retrieval performance. In: AN A. et al. (eds). **Active Media Technology: AMT 2010**. Lecture Notes in Computer Science, v. 6335. Berlin, Heidelberg: Springer, 2010. p. 255–266. Disponível em: [https://doi.org/10.1007/978-3-642-15470-6\\_27](https://doi.org/10.1007/978-3-642-15470-6_27). Acesso em: jan. 2019.

YIN, W.; ZHU, M.; CHEN, T. Domain Thesaurus Construction from Wikipedia. In: INTERNATIONAL CONFERENCE ON COMPUTER, NETWORKS AND COMMUNICATION ENGINEERING, Beijing, China during May 23-24, 2013. **Proceedings...** Amsterdam, The Netherlands: Atlantis Press, 2013. p. 87–92. Disponível em: <https://dx.doi.org/10.2991/iccnce.2013.22>. Acesso em: jan. 2019.

YOON, J. W. Towards a user-oriented thesaurus for non-domain-specific image collections. **Information Processing and Management**, v. 45, n. 4, p. 452–468, 2009. Disponível em: <http://dx.doi.org/10.1016/j.ipm.2009.03.004>. Acesso em: jan. 2019.

ZAPILKO, B. *et al.* TheSoz: a SKOS representation of the Thesaurus for the Social Sciences. **Semantic Web**, v. 4, n. 3, p. 257–263, 2013. Disponível em: <https://dl.acm.org/doi/abs/10.5555/2786071.2786076>. Acesso em: abr. 2020.

ZARZOUR, H.; AL-AYYOUB, M.; JARARWEH, Y. Towards social collaborative editing of distributed linked data. In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION SYSTEMS, 9., Irbid, Jordan, may, 2018. **Proceedings...** New York: IEEE, 2018. p. 90–95. Disponível em: <https://doi.org/10.1109/IACS.2018.8355447>. Acesso em: jan. 2019.

ZARZOUR, H.; SELLAMI, M. B-Set: A synchronization method for distributed semantic stores. In: INTERNATIONAL CONFERENCE ON COMPLEX SYSTEMS, Agadir, Morocco, feb., 2012. **Proceeding...** New York: IEEE, 2012. p. 1–6. Disponível em: <https://doi.org/10.1109/ICoCS.2012.6458550>. Acesso em: jan. 2019.

ZARZOUR, H.; SELLAMI, M. Achieving consistency in collaborative image annotation systems. In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION SYSTEMS, 6., Irbid, Jordan, 2014. **Proceedings...** New York: IEEE, 2014a. p. 1–7. Disponível em: <https://doi.org/10.1109/IACS.2014.6841963>. Acesso em: jan. 2019.

ZARZOUR, H.; SELLAMI, M. Using commutative replicated data type for collaborative video annotation. In: INTERNATIONAL CONFERENCE ON MULTIMEDIA COMPUTING AND SYSTEMS, Marrakech, Morocco, apr., 2014. **Proceedings...** New York: IEEE, 2014b. p. 523–529. Disponível em: <https://doi.org/10.1109/ICMCS.2014.6911203>. Acesso em: jan. 2019.

ZAVERI, Amrapali et al. Quality Assessment for Linked Data: A Survey: A Systematic Literature Review and Conceptual Framework. **Semantic Web**, v. 1, p. 1–5, 2012. Disponível em: <http://www.semantic-web-journal.net/system/files/swj773.pdf>. Acesso em: jan. 2019.

ZENG, M. L. Knowledge organization Systems. **Knowledge organization**, v. 35, n. 2–3, p. 160–182, 2008. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ko35200823h.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ko35200823h.pdf). Acesso em: nov. 2018.

ZENG, M. L.; GRACY, K. F.; SKIRVIN, L. Navigating the Intersection of Library Bibliographic Data and Linked Music Information Sources: A Study of the Identification of Useful Metadata Elements for Interlinking. **Journal of Library Metadata**, v. 13, n. 2–3, p. 254–278, 2013. Disponível em: <https://doi.org/10.1080/19386389.2013.827513>. Acesso em: jan. 2019.

ZENG, M. L.; MAYR, P. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. **International Journal on Digital Libraries**, v. 20, n. 3, p. 1–22, 2018. Disponível em: <https://doi.org/10.1007/s00799-018-0241-2>. Acesso em: jan. 2019.

ZHANG, J.; WOLFRAM, D. Visual analysis of obesity-related query terms on HealthLink. **Online Information Review**, v. 33, n. 1, p. 43–57, 2009. Disponível em: <https://doi.org/10.1108/14684520910944382>. Acesso em: nov. 2019.

ZHANG, M. L.; ZHOU, Z. H. Multilabel neural networks with applications to functional genomics and text categorization. **IEEE Transactions on Knowledge and Data Engineering**, v. 18, n. 10, p. 1338–1351, 2006. Disponível em: <https://doi.org/10.1109/TKDE.2006.162>. Acesso em: nov. 2019.

ZHANG, S.; ZHANG, C.; YANG, Q. Data preparation for data mining. **Applied Artificial Intelligence**, v. 17, n. 5–6, p. 375–381, 2003. Disponível em: <https://doi.org/10.1080/713827180>. Acesso em: jun. 2020.

ZHU, L.; QIAO, X.; SI, L. Constructing an open-sharing Chinese Scientific & Technical Vocabulary Service. In: INTERNATIONAL CONFERENCE ON NETWORKING AND DIGITAL SOCIETY, ICNDS 2009, Guiyang, Guizhou, China. **Proceedings...** Guiyang, Guizhou, China: IEEE, 2009. p. 284–287. Disponível em: <https://doi.org/10.1109/ICNDS.2009.150>. Acesso em: abr. 2020.

ZOGLAMI, K.; KERHERVÉ, B.; GERBÉ, O. Using a SKOS engine to create, share and transfer terminology data sets. In: INTERNATIONAL CONFERENCE ON SIGNAL IMAGE TECHNOLOGY AND INTERNET-BASED SYSTEMS, 7., Dijon, France, jan. 2011. **Proceedings...** New York: IEEE, 2011. Disponível em: <https://doi.org/10.1109/SITIS.2011.59>. Acesso em: jan. 2019.

ZOU, Q. Building a platform for linked open thesauri. **Library Hi Tech**, v. 31, n. 4, p. 620–637, 2013. Disponível em: <https://doi.org/10.1108/LHT-03-2013-0022>. Acesso em: jan. 2019.

ZOU, Q. Represent Changes of Knowledge Organization Systems on the Semantic Web. **International Journal of Librarianship**, v. 3, n. 1, p. 67, 2018. Disponível em: <https://doi.org/10.23974/ijol.2018.vol3.1.64>. Acesso em: abr. 2020.



## APÊNDICES

## APÊNDICE A – CRITÉRIOS DE INCLUSÃO E EXCLUSÃO PARA AS BUSCAS SISTEMÁTICAS NAS BASES DE DADOS

Apresenta os critérios de inclusão e exclusão das Buscas Sistemáticas realizadas no âmbito desta pesquisa, dispostas no Quadro 55.

Inicialmente buscou-se investigar a pertinência e adequação dos termos de busca e das bases de dados nacionais e internacionais que cobrem as áreas da Ciência da Informação e Ciência da Computação. A revisão teve como principal objetivo responder ao seguinte questionamento: quais são os padrões e as tecnologias empregadas na construção e na publicação de tesouros no contexto do LOD?

**Quadro 55** - Critérios de inclusão e exclusão

Critérios	Incluir	Excluir
<b>Idiomas</b>	Português, Inglês e Espanhol	Outros idiomas
<b>Tipo de publicação</b>	Artigos de revisão Artigos de pesquisa Capítulos de livro Estudos de caso	Resumos Resenhas de livros Editoriais Relatórios de patentes Notícias Manuais práticos
<b>Categorias/áreas</b>	<i>Information Science Library Science</i> <i>Computer Science Interdisciplinary Applications</i> <i>Computer Science</i> <i>Social Sciences</i>	Outras categorias
<b>Período cronológico</b>	2004-2020	Outros períodos

**Fonte:** Elaborado pelo autor (2021).

## APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO



**Universidade Federal de Santa Catarina**  
**Programa de Pós-Graduação em Ciência da Informação**  
Campus Professor João David Ferreira Lima – Trindade – Florianópolis – Santa Catarina –  
Brasil – CEP: 88.040-900

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado(a) como voluntário(a) a participar da pesquisa: **Proposta de Modelo Colaborativo para a Elaboração de Tesouros do Domínio da Música no Contexto do *Linked Open Data***. Esta pesquisa está associada ao projeto de doutorado do Me. Everton Rodrigues Barbosa, do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina.

Este estudo surge da carência de estrutura teórica conceitual a respeito da construção colaborativa de tesouros na perspectiva das tecnologias do *Linked Open Data* (LOD). Os tesouros são instrumentos de representação e organização do conhecimento, os quais buscam estruturar de forma sistemática um domínio do saber por meio de um esquema de conceitos (termos) e suas relações semânticas. Tradicionalmente, os tesouros são utilizados na organização de coleções de bibliotecas, arquivos e museus, e, atualmente têm ganhado destaque na representação de recursos digitais pois são capazes de facilitar a busca e a recuperação de informações na *web*. Desse modo, os tesouros tem incorporado tecnologias e padrões do LOD, sendo esse um conceito que preconiza práticas cooperantes. No entanto, as metodologias colaborativas para o contexto dos dados vinculados podem variar conforme as tecnologias empregadas e as áreas de aplicação. Nessa oportunidade, o domínio da Música foi escolhido como área de aplicação devido à natureza heterogênea dos dados musicais na *web* e a necessidade deste domínio em atender demandas informacionais de um público diversificado (usuários especialistas e não especialistas). Outrossim, os desafios em torno da representação da Informação Musical envolvem a importância de elaborar instrumentos de organização do conhecimento que atendam às necessidades de usuários treinados (músicos e pesquisadores), e ao mesmo tempo, busquem suprir demandas de busca do público em geral, ou seja, indivíduo ou grupo de pessoas que utilizam a informação musical apenas para uma atividade recreativa. Nesse ponto, acredita-se que desenvolver um modelo colaborativo para a elaboração de tesouros

fará com que as instituições dedicadas à Organização da Informação Musical no Brasil otimizem as condições de acesso e uso da informação para atender um público heterogêneo a partir de tecnologias do LOD.

A coleta de dados ocorrerá por meio de aplicação de um questionário.

Existe um desconforto e risco mínimo para a pessoa que se submete em responder o questionário, que vão desde desconforto físico e psicológico, até consequências de uma eventual quebra de sigilo, ainda que involuntária e não intencional, nas relações pessoais e profissionais dos entrevistados, uma vez que as respostas às questões podem conter informações sobre política e estratégia das instituições envolvidas. No entanto, destacam-se alguns possíveis benefícios que a pesquisa pode proporcionar ao setor de patrimônio cultural, em especial às bibliotecas, como por exemplo, identificar tecnologias, modelos de dados e processos que estimulem práticas cooperantes entre pessoas e instituições. O modelo colaborativo poderá sustentar o debate sobre a criação de políticas ou programas voltados à construção colaborativa de vocabulários do domínio da Música, especialmente para o contexto dos dados vinculados na web, trazendo maior agilidade ao processo de construção desses vocabulários e maior visibilidade às coleções das instituições cooperantes.

A você é assegurado o acompanhamento e assistência, caso tenha alguma dúvida sobre os procedimentos ou sobre o projeto. Você poderá entrar em contato com o pesquisador a qualquer momento pelo telefone ou e-mail.

Você é livre para recusar-se a participar, retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar não irá acarretar qualquer penalidade.

O pesquisador Me. Everton Rodrigues Barbosa e seu orientador Prof. Dr. Angel Freddy Godoy Viera irão tratar a sua identidade com padrões profissionais de sigilo. Tomarão todas as providências necessárias para manter o sigilo, mas sempre existe a remota possibilidade da quebra do sigilo, mesmo que involuntário e não intencional, cujas consequências serão tratadas nos termos da lei. Os resultados deste trabalho poderão ser apresentados em encontros ou revistas científicas. Apresentarão apenas os resultados obtidos como um todo, sem revelar seu nome, instituição ou qualquer informação relacionada à sua privacidade.

Duas vias deste documento serão rubricadas e assinadas por você e pelo pesquisador responsável. Recomenda-se guardar cuidadosamente a sua via, pois é um documento que traz importantes informações de contato e garante os seus direitos como participantes da pesquisa.

Você poderá recusar a participar do estudo, ou retirar seu consentimento a qualquer momento, sem precisar justificar. Caso deseje sair da pesquisa não sofrerá qualquer prejuízo.

A participação no estudo não acarretará custos para você e não será oferecida nenhuma compensação financeira adicional. No entanto, caso você tenha qualquer despesa decorrente da participação nessa pesquisa, como transporte e alimentação, eles serão assumidos pelos pesquisadores e reembolsados para você. De igual maneira, caso sofra algum dano material ou imaterial devidamente comprovado e decorrentes diretamente dessa pesquisa, os pesquisadores garantem indenizá-lo conforme determina a lei vigente.

O pesquisador responsável, que também assina esse documento, compromete-se a conduzir a pesquisa de acordo com o que preconiza a Resolução 466/12 de 12/06/2012, que trata dos preceitos éticos e da proteção aos participantes da pesquisa.

Eu, \_\_\_\_\_,

RG \_\_\_\_\_ fui informada (o) dos objetivos da pesquisa acima de maneira clara e detalhada e esclareci minhas dúvidas. Sei que em qualquer momento poderei solicitar novas informações e motivar minha decisão se assim o desejar. O pesquisador Me. Everton Rodrigues Barbosa e o professor orientador Prof. Dr. Angel Freddy Godoy Viera certificaram-me de que todos os dados desta pesquisa serão confidenciais.

Em caso de dúvidas poderei chamar o pesquisador Me. Everton Rodrigues Barbosa no telefone (47)99935-2943 ou pelo e-mail [evertonpos@gmail.com](mailto:evertonpos@gmail.com), o professor orientador Prof. Dr. Angel Freddy Godoy Viera no telefone (48)3721-4075 ou pelo e-mail [a.godoy@ufsc.br](mailto:a.godoy@ufsc.br) ou o Comitê de Ética em Pesquisa com Seres Humanos (CEPSH) da Universidade Federal de Santa Catarina (UFSC), sito à Rua Desembargador Vitor Lima, nº 222, sala 401, Trindade – Florianópolis, SC. Telefone (48)3721-6094 ou e-mail [cep.propesq@contato.ufsc.br](mailto:cep.propesq@contato.ufsc.br).

Declaro que concordo em participar deste estudo. Recebi uma cópia deste termo de consentimento livre e esclarecido e me foi dada a oportunidade de ler e esclarecer as minhas dúvidas.

---

Nome	Assinatura do Participante	Data
------	----------------------------	------

---

Nome	Assinatura do Pesquisador	Data
------	---------------------------	------

---

Nome	Assinatura do Professor Orientador	Data
------	------------------------------------	------

## APÊNDICE C – QUESTIONÁRIO



Universidade Federal de Santa Catarina

Programa de Pós-Graduação em Ciência da Informação

Campus Professor João David Ferreira Lima – Trindade – Florianópolis – Santa Catarina –

Brasil – CEP: 88.040-900

### Questionário semiestruturado

O questionário está organizado em oito seções, com questões abertas e fechadas, desenvolvidas com o objetivo de identificar procedimentos, recursos e tecnologias aplicadas à construção dos vocabulários controlados das instituições respondentes. A pesquisa apresenta o domínio da Música como área principal, portanto, algumas questões precisam ser especificadas considerando o domínio de aplicação.

<b>1</b>	<b>Caracterização do acervo e sistemas de gestão</b>
1.1	Qual o sistema utilizado pela instituição para a gestão de seu acervo bibliográfico?
1.2	A instituição disponibiliza a consulta de seu acervo por meio de catálogos <i>online</i> ? ( ) Sim ( ) Não
1.3	A instituição disponibiliza coleções digitais especializadas no domínio da Música? ( ) Sim ( ) Não Em caso afirmativo, quais tipos de documentos são disponibilizados?
1.4	A gestão do vocabulário controlado ocorre por meio de quais sistemas? ( ) O mesmo sistema utilizado para a gestão do acervo bibliográfico ( ) Um sistema específico para gestão de vocabulários controlados Especifique o nome do sistema utilizado: _____
<b>2</b>	<b>Elaboração e manutenção do vocabulário controlado</b>
2.1	O vocabulário controlado é atualizado com que frequência? ( ) Diariamente ( ) Semanalmente ( ) Mensalmente ( ) Semestralmente ( ) Anualmente Outro: _____
2.2	Quais são os profissionais envolvidos na elaboração e manutenção do vocabulário do domínio da Música? ( ) Bibliotecários

	<p>( ) Docentes  ( ) Pesquisadores  ( ) Informáticos  ( ) Músicos  ( ) Especialistas em domínios correlatos (História, educação, etc.).  Outros: _____  Obs.: É possível assinalar mais de uma opção.  Especifique quais são os domínios de especialidade dos profissionais envolvidos:</p>
2.3	<p>A comunidade pode sugerir inclusões ou modificações de termos no vocabulário controlado?  ( ) Sim ( ) Não  Em caso afirmativo, as inclusões ou modificações são sugeridas por:  ( ) Membros da equipe de controle  ( ) Usuários  ( ) Docentes  ( ) Pesquisadores  ( ) Pessoas ou grupos externos à instituição  ( ) Músicos  ( ) Especialistas em domínios correlatos (História, antropologia, etc.).  ( ) Outros: _____  Obs.: É possível assinalar mais de uma opção.</p>
2.4	<p>No que se refere as recomendações de inclusão ou modificação de termos no vocabulário realizados pela comunidade (usuários do sistema):</p> <p>a) Quais são os canais de comunicação utilizados para que a comunidade recomende as inclusões ou modificações?  Descreva tecnologias ou ferramentas utilizadas.</p> <p>b) Como são realizadas as recomendações de inclusão ou modificação dos termos pela comunidade?  Descreva brevemente as etapas de solicitação, análise e inclusão.</p>
2.5	<p>A unidade de informação faz parte de alguma rede cooperativa de catalogação ou indexação de documentos especializados no domínio da Música ou áreas correlatas?  ( ) Sim ( ) Não</p> <p>c) Em caso afirmativo, quais são as instituições que cooperam?  d) Em ambos os casos, na sua opinião, qual a importância do controle do vocabulário para as redes cooperativas de indexação ou catalogação?</p>
2.6	<p>Existe algum programa ou projeto de <i>cooperação formal</i>* para a manutenção do vocabulário controlado geral, ou específico para o domínio da Música?  ( ) Sim ( ) Não  Em caso afirmativo:</p> <p>a) Quais são as entidades ou grupos de pessoas que participam?  b) Por meio de quais canais de comunicação ocorre a cooperação formal?</p>

	<p>Especifique as tecnologias ou ferramentas utilizadas.</p> <p>*parcerias firmadas entre instituições ou grupos de pessoas com intenção de cooperar.</p>
2.7	<p>Existe alguma iniciativa de <b>cooperação informal*</b> para a manutenção do vocabulário controlado geral, ou específico para o domínio da Música?</p> <p>( ) Sim ( ) Não</p> <p>Em caso afirmativo:</p> <p>a) Quais são as entidades ou grupos de pessoas que participam?</p> <p>b) Por meio de quais canais de comunicação ocorre a cooperação informal?</p> <p>Especifique as tecnologias ou ferramentas utilizadas.</p> <p>* parcerias não firmadas entre instituições ou grupos de pessoas que não possuem a intenção direta de cooperar. Nesse caso, a cooperação pode ocorrer por meio de <i>feedbacks</i> explícitos ou implícitos do uso dos vocabulários, como por exemplo, a compilação de termos utilizados pelo usuário durante sua interação com o sistema de busca.</p>
<b>3</b>	<b>Compilação ou levantamento de termos para a construção do vocabulário</b>
3.1	Como são selecionados os termos candidatos a inclusão no vocabulário do domínio da Música?
3.2	<p>Os termos para inclusão no vocabulário controlado são selecionados a partir de terminologias ou vocabulários existentes?</p> <p>( ) Sim ( ) Não</p> <p>a) Em caso afirmativo, quais são as terminologias ou vocabulários existentes utilizados?</p> <p>( ) Vocabulário Controlado da USP</p> <p>( ) <i>Library of Congress Subject Headings</i></p> <p>( ) <i>Musical Instrument Museums Online</i></p> <p>( ) Tesouro EuroVoc</p> <p>( ) UNESCO Thesaurus</p> <p>( ) Thesaurus Brasileiro da Educação - Brased</p> <p>( ) Tesouro de Folclore e Cultura Popular Brasileira</p> <p>( ) Terminologia de Assuntos da Biblioteca Nacional</p> <p>Outro: _____</p> <p>Obs.: É possível assinalar mais de uma alternativa.</p> <p>b) Em caso afirmativo, existe uma parceria formal com as instituições que mantêm esses vocabulários?</p> <p>( ) Sim, existe uma colaboração direta firmada</p> <p>( ) Não, os termos são apenas recolhidos de seus endereços na <i>web</i></p>
3.3	<p>Os termos são incluídos no vocabulário exatamente como se apresentam nas terminologias ou vocabulários existentes, ou podem ser adaptados à realidade da instituição?</p> <p>( ) Usados exatamente como são apresentados nos vocabulários existentes</p> <p>( ) Passíveis de adaptação</p> <p>Para ambos os casos, justifique o porquê:</p>
3.4	Quais são os profissionais*, unidades especializadas ou departamentos responsáveis pela compilação dos termos do domínio da Música?



	*descrever formação e cargos que ocupam.
3.5	Quais as competências básicas atribuídas aos profissionais responsáveis pela compilação dos termos do domínio da Música?
3.6	O processo de compilação dos termos ocorre apenas na fase inicial de construção do vocabulário, ou em outras etapas do processo de construção? Especifique quais etapas podem ocorrer a compilação de termos.
3.7	A instituição utiliza alguma tecnologia ou ferramenta para a compilação automática ou semiautomática de termos? Por exemplo, tecnologias de mineração de texto e dados, pré-processamento de documentos, indexação automática, ferramentas bibliométricas, etc. ( ) Sim ( ) Não Em caso positivo, especifique o tipo de ferramenta:
3.8	Caso a instituição utilize alguma ferramenta para a compilação automática ou semiautomática de termos, é necessário que os documentos estejam em algum formato ou padrão de dados? Por exemplo, doc, PDF, txt, CSV, XML, etc. ( ) Sim ( ) Não Em caso afirmativo, quais formatos ou padrão de dados são exigidos?
3.9	O vocabulário controlado contempla termos em outros idiomas? ( ) Sim ( ) Não Em caso afirmativo, como ocorre o processo de compilação e/ou tradução?
<b>4</b>	<b>Especificação ou supressão de termos</b>
4.1	Os termos compilados são passíveis de especificação? ( ) Sim ( ) Não Em caso afirmativo: a) Quais são os casos que exigem o processo de especificação?
4.2	Os termos compilados são passíveis de supressão (exclusão)? ( ) Sim ( ) Não Em caso afirmativo: a) Quais são os casos que exigem o processo de supressão?
4.3	São consultados especialistas do domínio da Música (docentes, pesquisadores, músicos e profissionais de áreas correlatas) para a consolidação e inclusão do termo no vocabulário? ( ) Sim ( ) Não Em caso afirmativo: a) Qual a formação desses especialistas e a importância de consultá-los? b) Quais são os canais de comunicação utilizados para realizar a consulta aos especialistas?
<b>5</b>	<b>Registro e validação de termos</b>
5.1	Como são realizados os registros dos termos? ( ) Fichas terminológicas ( ) Base de autoridades ( ) Apenas no registro bibliográfico (catalogação) ( ) <i>Software</i> de gestão de vocabulários controlados

	( ) Outros: _____
5.2	Os termos registrados no sistema passam por uma espécie de validação por outros membros da equipe de controle do vocabulário? ( ) Sim ( ) Não Em caso afirmativo, quais são os profissionais* responsáveis pela validação? *descrever formação e cargos que ocupam.
5.3	Existe alguma espécie de processamento automático, semiautomático ou manual aplicado à validação dos termos? ( ) Sim ( ) Não Em caso afirmativo, especifique quais são os instrumentos ou tecnologias utilizadas? A inclusão de termos por parte de servidores externos ao Setor de Autoridades é informada e a chefia do setor valida o registro.
5.4	A comunidade geral (usuários do sistema) participa de alguma forma na etapa de validação? ( ) Sim ( ) Não Em caso afirmativo, especifique como ocorre e os instrumentos ou tecnologias utilizadas?
5.5	Como a equipe chega ao consenso sobre a inclusão de um termo no vocabulário?
<b>6</b>	<b>Testes e avaliação do vocabulário controlado</b>
6.1	Os vocabulários controlados do domínio da Música são submetidos à avaliação dos usuários? ( ) Sim ( ) Não Em caso afirmativo, a) Como ocorre a participação dos usuários na fase de testes e avaliação? b) Por meio de quais instrumentos ou recursos tecnológicos os <i>feedbacks</i> são adquiridos? Em caso negativo, c) A instituição pretende incluí-los nesse processo? De que maneira?
6.2	Conforme sua experiência, qual a importância dos usuários na etapa de testes e avaliação do vocabulário controlado?
<b>7</b>	<b>Ferramentas colaborativas e representação do vocabulário na web</b>
7.1	O controle do vocabulário é realizado de maneira distribuída e compartilhada? Ou seja, os profissionais envolvidos operam geograficamente dispersos, criando ou editando a estrutura de termos por meio de ferramentas colaborativas? ( ) Sim ( ) Não Em caso afirmativo: a) Em quais etapas de elaboração do vocabulário controlado a edição colaborativa é necessária? Por quê?
7.2	O trabalho da equipe de construção utiliza quais tipos de recursos ou ferramentas de colaboração?

	<input type="checkbox"/> Web conferências <input type="checkbox"/> <i>Chats</i> (sala de bate-papo) <input type="checkbox"/> Fórum ou lista de discussão <input type="checkbox"/> e-mail <input type="checkbox"/> Blogs ou mídias sociais <input type="checkbox"/> Reuniões presenciais <input type="checkbox"/> Processos em <i>workflow</i> (fluxo de trabalho) <input type="checkbox"/> Gerenciamento distribuído de projetos (gestão de projetos) <input type="checkbox"/> Outro: _____ Obs.: É possível assinalar mais de uma alternativa.
7.3	Os recursos e ferramentas de colaboração assinalados na questão 7.2 são utilizados, principalmente, em quais etapas de construção: <input type="checkbox"/> Compilação de termos <input type="checkbox"/> Especificação e supressão de termos <input type="checkbox"/> Registro e validação de termos <input type="checkbox"/> Teste e avaliação
7.4	Quais são as características da plataforma ou sistema utilizado para a construção e manutenção do vocabulário controlado? <input type="checkbox"/> <b>Trabalha com computação em nuvem:</b> o conjunto de dados não está numa máquina específica, todos na instituição podem acessar. <input type="checkbox"/> <b>Editável simultaneamente:</b> Todos que tiverem autorização poderão editar o conjunto de dados do vocabulário. <input type="checkbox"/> <b>Acessível de forma remota:</b> é possível visualizar as informações, mediante login e senha, de qualquer lugar que se esteja. <input type="checkbox"/> <b>Compatível com diversos dispositivos:</b> diversos aparelhos estão preparados para usar o aplicativo, seja <i>smartphone</i> ou <i>desktop</i> . <input type="checkbox"/> <b>Oferecido na forma de <i>Software as a Service (SaaS)</i>:</b> utiliza <i>software</i> comercial, acessado via Internet. O cliente paga pelo serviço e o fornecedor se responsabiliza por toda a estrutura necessária a utilização (servidores, conectividade, cuidados com segurança da informação). <input type="checkbox"/> outras características colaborativas: Obs.: É possível assinalar mais de uma alternativa.
7.5	A plataforma ou sistema utilizado para a construção e manutenção do vocabulário controlado utiliza alguma tecnologia, padrão ou modelo de dados relacionados ao conjunto de práticas do <i>Linked Open Data</i> ? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desconheço Em caso afirmativo, quais tecnologias, padrões ou modelos de dados são utilizados: <input type="checkbox"/> <b><i>Resource Description Framework (RDF)</i>:</b> linguagem formal para a descrição de informação estruturadas sobre recursos da <i>Web</i> . <input type="checkbox"/> <b><i>Uniform Resource Identifiers (URI)</i>:</b> cadeia de caracteres compacta usada para identificar ou denominar um recurso na web. <input type="checkbox"/> <b><i>eXtensible Markup Language (XML)</i>:</b> linguagem de marcação para a criação de documentos com dados organizados hierarquicamente, tais como textos, banco de dados ou desenhos vetoriais.

	<p><input type="checkbox"/> <b>Simple Knowledge Organization System (SKOS)</b>: modelo de dados utilizado para representar os Sistemas de Organização do Conhecimento (tesauros, taxonomias, ontologias e cabeçalhos de assuntos) na web, em um esquema de dados legíveis por máquina.</p> <p><input type="checkbox"/> <b>SPARQL</b>: conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web.</p> <p><input type="checkbox"/> <b>Web Ontology Language (OWL)</b>: é uma linguagem para definir e instanciar ontologias na web.</p> <p><input type="checkbox"/> Outro :</p> <p>Obs.: É possível assinalar mais de uma alternativa.</p>
7.6	<p>A instituição disponibiliza seu vocabulário controlado como <i>Linked Open Data</i>?</p> <p><input type="checkbox"/> Sim <input type="checkbox"/> Não</p> <p>Em caso afirmativo, especificar: Qual o tipo de licença atribuída ao conjunto de dados?</p>
7.7	<p>O vocabulário da instituição é vinculado a outros vocabulários controlados estruturados como <i>Linked Open Data</i>?</p> <p><input type="checkbox"/> Sim <input type="checkbox"/> Não</p> <p>a) Em caso afirmativo, quais são os vocabulários:</p> <p><input type="checkbox"/> <i>Library of Congress Subject Headings (LCSH)</i></p> <p><input type="checkbox"/> <i>UNESCO Thesaurus</i></p> <p><input type="checkbox"/> EuroVoc</p> <p><input type="checkbox"/> <i>MIMO Thesaurus</i></p> <p><input type="checkbox"/> Outros: _____</p> <p>b) Em caso afirmativo, como ocorre o processo de mapeamento e vinculação desses dados?</p>
7.8	<p>O conjunto de dados do vocabulário controlado é vinculado a outros conjuntos de dados ligados ao domínio da Música?</p> <p><input type="checkbox"/> Sim <input type="checkbox"/> Não</p> <p>a) Em caso afirmativo, quais são os conjuntos de dados?</p> <p><input type="checkbox"/> DBpedia</p> <p><input type="checkbox"/> DBTune</p> <p><input type="checkbox"/> <i>Musical Instrument Museums Online (MIMO)</i></p> <p><input type="checkbox"/> Répertoire International des Sources Musicales (RISM)</p> <p><input type="checkbox"/> MusicBrainz</p> <p><input type="checkbox"/> AcousticBrainz</p> <p><input type="checkbox"/> Doremus</p> <p>Outros: _____</p> <p>b) Em caso afirmativo, como ocorre o processo de mapeamento e vinculação desses dados?</p>
<b>8</b>	<b>Percepção da colaboração nas etapas de construção do vocabulário controlado</b>
8.1	De acordo com sua experiência, em qual etapa de construção ou manutenção do vocabulário controlado a colaboração é mais evidente? Por quê?

## APÊNDICE D – AMOSTRA TEÓRICA: ESTUDOS ANALISADOS

Número	Documentos
1	AGUILAR, A. G. Creación y visualización de tesauros a partir de un vocabulario de técnicas de la información y de la comunicación. <b>Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação</b> , v. 18, n. 37, p. 303–320, 2013. Disponível em: <a href="https://doi.org/10.5007/1518-2924.2013v18n37p303">https://doi.org/10.5007/1518-2924.2013v18n37p303</a> . Acesso em: abr. 2020.
2	ALBERTONI, R. <i>et al.</i> EARTH: an Environmental Application Reference Thesaurus in the Linked Open Data. <b>Semantic Web</b> , v. 5, n. 2, p. 165–171, 2014. Disponível em: <a href="https://dl.acm.org/doi/10.5555/2786102.2786112">https://dl.acm.org/doi/10.5555/2786102.2786112</a> . Acesso em: abr. 2020.
3	ALBERTONI, R.; DE MARTINO, M.; PODESTÀ, P. Quality measures for skos: ExactMatch linksets: an application to the thesaurus framework LusTRE. <b>Data Technologies and Applications</b> , v. 52, n. 3, p. 405–423, 2018. Disponível em: <a href="https://doi.org/10.1108/DTA-05-2017-0037">https://doi.org/10.1108/DTA-05-2017-0037</a> . Acesso em: abr. 2020.
4	ALEXIEV, V.; ISAAC, A.; LINDENTHAL, J. On the composition of ISO 25964 hierarchical relations (BTG, BTP, BTI). <b>International Journal on Digital Libraries</b> , v. 17, n. 1, p. 39–48, 2016. Disponível em: <a href="https://doi.org/10.1007/s00799-015-0162-2">https://doi.org/10.1007/s00799-015-0162-2</a> . Acesso em: abr. 2020.
5	ANIBALDI, S. <i>et al.</i> Migrating bibliographic datasets to the Semantic Web: The AGRIS case. <b>Semantic Web</b> , v. 6, n. 2, p. 113–120, 2015. Disponível em: <a href="http://www.semantic-web-journal.net/system/files/swj463.pdf">http://www.semantic-web-journal.net/system/files/swj463.pdf</a> . Acesso em: abr. 2020.
6	BACA, M.; GILL, M. Encoding multilingual knowledge systems in the digital age: the Getty Vocabularies. <b>Knowledge organization</b> , v. 42, n. 4, p. 232–243, 2014. Disponível em: <a href="https://doi.org/10.5771/0943-7444-2015-4-232">https://doi.org/10.5771/0943-7444-2015-4-232</a> . Acesso em: abr. 2020.
7	BELLINI, P.; CAPPUCCIO, A.; NESI, P. Collaborative and assisted SKOS generation and management. In: INTERNATIONAL CONFERENCE ON DISTRIBUTED MULTIMEDIA SYSTEMS (DMS 2011), 17., Florence, Italy, 2011. <b>Proceedings...</b> Florence, Italy: University of Florence, 2011. p. 28–33. Disponível em: <a href="http://www.disit.org/disitmn/DMS2011-nesi-paper16-OpenSpaceInnoMind-SKOS-v1-2-6pages.pdf">http://www.disit.org/disitmn/DMS2011-nesi-paper16-OpenSpaceInnoMind-SKOS-v1-2-6pages.pdf</a> . Acesso em: abr. 2020.
8	BENEVENTANO, D. <i>et al.</i> Semantic annotation of the CEREALAB database by the AGROVOC linked dataset. <b>Ecological Informatics</b> , v. 26, part 2, p. 119–126, 2015. Disponível em: <a href="http://dx.doi.org/10.1016/j.ecoinf.2014.07.002">http://dx.doi.org/10.1016/j.ecoinf.2014.07.002</a> . Acesso em: abr. 2020.
9	BERGAMIN, G.; BERGAMIN, G.; LUCARELLI, A. The Nuovo soggettario as a service for the linked data world. <b>JLIS.it</b> , v. 4, n. 1, p. 213, 2013. Disponível em: <a href="http://dx.doi.org/10.4403/jlis.it-5474">http://dx.doi.org/10.4403/jlis.it-5474</a> . Acesso em: abr. 2020.

10	BINDING, C.; TUDHOPE, D. Improving interoperability using vocabulary linked data. <b>International Journal on Digital Libraries</b> , v. 17, n. 1, p. 5–21, 2016. Disponível em: <a href="https://doi.org/10.1007/s00799-015-0166-y">https://doi.org/10.1007/s00799-015-0166-y</a> . Acesso em: abr. 2020.
11	ÇAĞDAŞ, V.; STUBKJÆR, E. A SKOS vocabulary for Linked Land Administration: Cadastre and Land Administration Thesaurus. <b>Land Use Policy</b> , v. 49, p. 668–679, 2015. Disponível em: <a href="https://doi.org/10.1016/j.landusepol.2014.12.017">https://doi.org/10.1016/j.landusepol.2014.12.017</a> . Acesso em: abr. 2020.
12	CARACCILO, C. <i>et al.</i> Thesaurus maintenance, alignment and publication as linked data: The AGROOVOC use case. In: GARCÍA-BARRIOCANAL, E. <i>et al.</i> (Org.). <b>Metadata and Semantic Research, MTSR 2011: Communications in Computer and Information Science</b> , v. 240. Berlin, Heidelberg: Springer, 2011. p. 489–499. Disponível em: <a href="https://doi.org/10.1007/978-3-642-24731-6_48">https://doi.org/10.1007/978-3-642-24731-6_48</a> . Acesso em: abr. 2020.
13	CARACCILO, C. <i>et al.</i> The AGROVOC Linked Dataset. <b>Semantic Web</b> , v. 4, n. 3, p. 341–348, 2013. Disponível em: <a href="https://dl.acm.org/doi/10.5555/2786071.2786087">https://dl.acm.org/doi/10.5555/2786071.2786087</a> . Acesso em: abr. 2020.
14	COBB, J. The Journey to Linked Open Data: The GettyVocabularies. <b>Journal of Library Metadata</b> , v. 15, n. 3–4, p. 142–156, 2016. Disponível em: <a href="https://doi.org/10.1080/19386389.2015.1103081">https://doi.org/10.1080/19386389.2015.1103081</a> . Acesso em: abr. 2020
15	COYLE, K. Vocabularies: Term Lists and Thesauri. <b>Library Technology Reports</b> , v. 48, n. 4, p. 27–36, 2012. Disponível em: <a href="https://journals.ala.org/index.php/ltr/article/view/4670/5540">https://journals.ala.org/index.php/ltr/article/view/4670/5540</a> . Acesso em: abr. 2020.
16	DEBEVERE, P. <i>et al.</i> Linking thesauri to the linked open data cloud for improved media retrieval. In: INTERNATIONAL WORKSHOP ON IMAGE ANALYSIS FOR MULTIMEDIA INTERACTIVE SERVICES (WIAMIS 2011), 12., apr. 2011. <b>Proceedings...</b> The Netherlands: TUDelft, 2011. p. 28–37. Disponível em: <a href="https://biblio.ugent.be/publication/1210156">https://biblio.ugent.be/publication/1210156</a> . Acesso em: abr. 2020.
17	DÍAZ-CORONA, D. <i>et al.</i> Profiling of knowledge organisation systems for the annotation of Linked Data cultural resources. <b>Information Systems</b> , v. 84, p. 17–28, 2019. Disponível em: <a href="https://doi.org/10.1016/j.is.2019.04.008">https://doi.org/10.1016/j.is.2019.04.008</a> . Acesso em: abr. 2020.
18	DIMITROVA, V. <i>et al.</i> Interoperability of language-related information: Mapping the BLL thesaurus to LEXVO and glottolog. In: International Conference on Language Resources and Evaluation, LREC 2018, 11., 2018. <b>Proceedings...</b> Paris: European Language Resources Association, 2018. p. 4555–4561. Disponível em: <a href="http://www.lrec-conf.org/proceedings/lrec2018/pdf/259.pdf">http://www.lrec-conf.org/proceedings/lrec2018/pdf/259.pdf</a> . Acesso em: abr. 2020.
19	FABEIRO, R.; MASAGUÉ, C.; ROVIRA JARQUE, A. Evolució de les dades obertes enllaçades (LOD) en el context dels vocabularis controlats. <b>Item: Revista de</b>

	<b>Biblioteconomia i Documentació</b> , v. 2, n. 65-66, p. 1–24, 2019. Disponível em: <a href="https://www.raco.cat/index.php/Item/article/view/353616">https://www.raco.cat/index.php/Item/article/view/353616</a> . Acesso em: abr. 2020.
20	GERBÉ, O.; KERHERVÉ, B. A model-driven approach to SKOS implementation. In: INTERNATIONAL CONFERENCE ON INTERNET AND WEB APPLICATIONS AND SERVICES (ICIW 2010), 5., 2010. <b>Proceedings...</b> Barcelona: IEEE, 2010. p. 484–488, 2010. Disponível em: <a href="https://doi.org/10.1109/iciw.2010.79">https://doi.org/10.1109/iciw.2010.79</a> . Acesso em: abr. 2020.
21	HARPRING, P. Linking the Getty Vocabularies: the content perspective, Including an Update on CONA. In: PACIFIC NEIGHBORHOOD CONSORTIUM ANNUAL CONFERENCE AND JOINT MEETINGS (PNC), 2018. <b>Proceedings...</b> San Francisco, CA, USA: IEEE, 2018. p. 1–8. Disponível em: <a href="https://ieeexplore.ieee.org/document/8579460">https://ieeexplore.ieee.org/document/8579460</a> . Acesso em: abr. 2020.
22	ISAAC, A.; MATTHEZING, H. Representing and Aligning Thesauri for an Integrated Access to Cultural Heritage Resources. In: INTERNATIONAL SEMINAR “INFORMATION ACCESS FOR THE GLOBAL COMMUNITY”, 29., 4-5 June 2007. <b>Proceedings...</b> The Hague: UDC Consortium, 2007. p. 301–310. Disponível em: <a href="http://hdl.handle.net/10150/106244">http://hdl.handle.net/10150/106244</a> . Acesso em: abr. 2020.
23	JORGE, N. M. DA C. <b>Ensaio sobre o AAT-Art &amp; Architecture Thesaurus : proposta terminológica de adaptação à realidade portuguesa</b> . 2011. 150 f. Universidade do Porto, Porto, 2011. Disponível em: <a href="http://repositorio-aberto.up.pt/handle/10216/57042">http://repositorio-aberto.up.pt/handle/10216/57042</a> . Acesso em: abr. 2020.
24	KEMPF, A. O.; NEUBERT, J. The role of thesauri in an open web: a case study of the STW Thesaurus for Economics. <b>Knowledge organization</b> , v. 43, p. 160–173, 2016. Disponível em: <a href="https://doi.org/10.5771/0943-7444-2016-3-160">https://doi.org/10.5771/0943-7444-2016-3-160</a> . Acesso em: abr. 2020.
25	L’ABATE, G. <i>et al.</i> Exposing vocabularies for soil as Linked Open Data. <b>Information Processing in Agriculture</b> , v. 2, n. 3–4, p. 208–216, 2015. Disponível em: <a href="http://dx.doi.org/10.1016/j.inpa.2015.10.002">http://dx.doi.org/10.1016/j.inpa.2015.10.002</a> . Acesso em: abr. 2020.
26	LAPORTE, M. A.; MOUGENOT, I.; GARNIER, E. ThesauForm-Traits: A web based collaborative tool to develop a thesaurus for plant functional diversity research. <b>Ecological Informatics</b> , v. 11, p. 34–44, 2012. Disponível em: <a href="http://dx.doi.org/10.1016/j.ecoinf.2012.04.004">http://dx.doi.org/10.1016/j.ecoinf.2012.04.004</a> . Acesso em: abr. 2020.
27	LAPORTE, M.-A. <i>et al.</i> A semantic web faceted search system for facilitating building of biodiversity and ecosystems services. In: GALHARDAS, H.; RAHM, E. (Org.). <b>Data Integration in the Life Sciences, DILS 2014</b> : Lecture Notes in Computer Science, vol 8574. Cham: Springer, 2014. p. 50–57. Disponível em: <a href="https://link.springer.com/chapter/10.1007/978-3-319-08590-6_5">https://link.springer.com/chapter/10.1007/978-3-319-08590-6_5</a> . Acesso em: abr. 2020.
28	LARA, M. L. G. Documentary languages and knowledge organization systems in the

	context of the semantic web. <b>Transinformacao</b> , v. 25, n. 2, p. 145–150, maio/ago. 2013. Disponível em: <a href="https://doi.org/10.1590/S0103-37862013000200005">https://doi.org/10.1590/S0103-37862013000200005</a> . Acesso em: abr. 2020.
29	MACULAN, B. C. M. DOS S. <b>Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do THESAGRO</b> . 2015. 343 f. Tese. Universidade Federal de Minas Gerais, 2015. Disponível em: <a href="https://repositorio.ufmg.br/handle/1843/BUBD-9ZKMUV">https://repositorio.ufmg.br/handle/1843/BUBD-9ZKMUV</a> . Acesso em: abr. 2020.
30	MARTÍNEZ GONZÁLEZ, M. M.; ALVITE DÍEZ, M. . L. Propuesta metodológica de evaluación de gestores de tesauros compatibles con la web semántica. <b>Anales de Documentación</b> , v. 17, n. 1, p. 1–18, 2014. Disponível em: <a href="https://doi.org/10.6018/analesdoc.17.1.186271">https://doi.org/10.6018/analesdoc.17.1.186271</a> . Acesso em: abr. 2020.
31	MARTÍNEZ-GONZÁLEZ, M. M.; ALVITE-DÍEZ, M. L. Thesauri and Semantic Web: discussion of the evolution of thesauri toward their integration with the semantic web. <b>IEEE Access</b> , v. 7, p. 153151–153170, 2019. Disponível em: <a href="https://doi.org/10.1109/ACCESS.2019.2948028">https://doi.org/10.1109/ACCESS.2019.2948028</a> . Acesso em: abr. 2020.
32	MAZIARZ, M.; PIASECKI, M. Towards mapping thesauri onto plWordNet. In: GLOBAL WORDNET CONFERENCE (GWC 2018), 9., 2018. <b>Proceedings...</b> Singapore: Nanyang Technological University (NTU), 2018. Disponível em: <a href="http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018_paper_63.pdf">http://compling.hss.ntu.edu.sg/events/2018-gwc/pdfs/GWC2018_paper_63.pdf</a> . Acesso em: abr. 2020.
33	MCCULLOCH, E. Thesauri: Practical guidance for construction. <b>Library Review</b> , v. 54, n. 7, p. 403–409, 2005. Disponível: <a href="https://doi.org/10.1108/00242530510611893">https://doi.org/10.1108/00242530510611893</a> . Acesso em: abr. 2020.
34	MOCHÓN, G.; MÉNDEZ, E. M.; BUENO DE LA FUENTE, G. 27 pawns ready for action: a multi-indicator methodology and evaluation of thesaurus management tools from a LOD perspective. <b>Library Hi Tech</b> , v. 35, n. 1, p. 99–119, 2017. Disponível em: <a href="https://doi.org/10.1108/LHT-11-2016-0123">https://doi.org/10.1108/LHT-11-2016-0123</a> . Acesso em: abr. 2020.
35	MOCHÓN BEZARES, G.; MÉNDEZ RODRÍGUEZ, E. M.; LA FUENTE, G. B. DE. Conversión normalizada (SKOS) de sistemas de organización del conocimiento interoperables en la web. <b>Investigación Bibliotecológica: archivonomía, bibliotecología e información</b> , v. 34, n. 82, p. 53–86, 2020. Disponível em: <a href="http://dx.doi.org/10.22201/iibi.24488321xe.2020.82.58071">http://dx.doi.org/10.22201/iibi.24488321xe.2020.82.58071</a> . Acesso em: abr. 2020.
36	MORAIS, M. W. DE; RAMALHO, R. A. S. Representação em SKOS de um microtesouro de conhecimentos estratégicos nas organizações. <b>Perspectivas em Ciência da Informação</b> , v. 24, n. 4, p. 187–198, 2019. Disponível em: <a href="https://doi.org/10.1590/1981-5344/4193">https://doi.org/10.1590/1981-5344/4193</a> . Acesso em: abr. 2020.



37	MORSHED, A. <i>et al.</i> Thesaurus alignment for linked data publishing. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2011. <b>Proceedings...</b> The Hague: Dublin Core Metadata Initiative, 2011. p. 37–46, 2011. Disponível em: <a href="https://dcpapers.dublincore.org/pubs/article/view/3623">https://dcpapers.dublincore.org/pubs/article/view/3623</a> . Acesso em: abr. 2020.
38	MORSHED, A. <i>et al.</i> Bridging end users' terms and AGROVOC concept server vocabularies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2010. <b>Proceedings...</b> Pittsburgh, Pennsylvania: Dublin Core Metadata Initiative, 2010. p. 186–189. Disponível em: <a href="https://dcpapers.dublincore.org/pubs/article/view/1015">https://dcpapers.dublincore.org/pubs/article/view/1015</a> . Acesso em: abr. 2020.
39	NEUBERT, J. Wikidata as a linking hub for knowledge organization systems? Integrating an authority mapping into Wikidata and learning lessons for KOS mappings. In: EUROPEAN NETWORKED KNOWLEDGE ORGANIZATION SYSTEMS (NKOS) WORKSHOP, 17., 2017. <b>Proceedings...</b> Thessaloniki, Greece: CEUR, 2017. p. 1–12. Disponível em: <a href="http://ceur-ws.org/Vol-1937/paper2.pdf">http://ceur-ws.org/Vol-1937/paper2.pdf</a> . Acesso em: abr. 2020.
40	NEUBERT, J. Linked data based library web services for economics. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2012. <b>Proceedings...</b> Kuching, Sarawak, Malaysia: Dublin Core Metadata Initiative, 2012. p. 12–22. Disponível em: <a href="https://dcpapers.dublincore.org/pubs/article/view/3659">https://dcpapers.dublincore.org/pubs/article/view/3659</a> . Acesso em: abr. 2020.
41	NIININEN, S.; NYKYRI, S.; SUOMINEN, O. The future of metadata: open, linked, and multilingual – the YSO case. <b>Journal of Documentation</b> , v. 73, n. 3, p. 451–465, 2017. Disponível em: <a href="https://doi.org/10.1108/JD-06-2016-0084">https://doi.org/10.1108/JD-06-2016-0084</a> . Acesso em: abr. 2020.
42	NOHAMA, P. <i>et al.</i> Quality issues in Thesaurus building: a case study from the medical domain. <b>Revista Brasileira de Engenharia Biomédica</b> , v. 28, n. 1, p. 11–22, 2012. Disponível em: <a href="http://rbejournal.org/doi/10.4322/rbeb.2012.002">http://rbejournal.org/doi/10.4322/rbeb.2012.002</a> . Acesso em: abr. 2020.
43	PASTOR SÁNCHEZ, J. A. <b>Diseño de un sistema colaborativo para la creación y gestión de tesauros en internet basados en SKOS</b> . p. 296, 2009. Tese. Universidad de Murcia. Espanha. Disponível em: <a href="http://hdl.handle.net/10201/3946">http://hdl.handle.net/10201/3946</a> . Acesso em: abr. 2020.
44	PASTOR-SÁNCHEZ, J.-A. Marcado semántico: tecnologías y aplicación para la representación de sistemas de organización del conocimiento en el contexto Linked Open Data. <b>Scire: Representación y Organización del Conocimientos</b> , v. 19, n. 2, jul. /dic., p. 47–60, 2013. Disponível em: <a href="https://www.ibersid.eu/ojs/index.php/scire/article/view/4112">https://www.ibersid.eu/ojs/index.php/scire/article/view/4112</a> . Acesso em: abr. 2020.
45	PASTOR-SÁNCHEZ, J. A.; LÓPEZ-CARREÑO, R.; MARTÍNEZ-MÉNDEZ, J. F. Advantages of SKOS for providing semantic stability to the information resources: an example of use in a collaborative institutional web portal. In: ISKO INTERNATIONAL

	CONFERENCE, 11., 2010. <b>Proceedings...</b> Rome: ISKO, 2010. Disponível em: <a href="http://eprints.rclis.org/24261/">http://eprints.rclis.org/24261/</a> . Acesso em: abr. 2020.
46	PASTOR-SÁNCHEZ, J.-A.; MARTÍNEZ-MÉNDEZ, F.-J.; RODRÍGUEZ-MUÑOZ, J.-V. Aplicación de SKOS para la interoperabilidad de vocabularios controlados en el entorno de linked open data. <b>El profesional de la información</b> , v. 21, n. 3, p. 245–253, 2012. Disponível em: <a href="http://dx.doi.org/10.3145/epi.2012.may.04">http://dx.doi.org/10.3145/epi.2012.may.04</a> . Acesso em: abr. 2020.
47	PIOTROWSKI, M.; COLAVIZZA, G.; THIERY, F. The Labeling System: A New Approach to Overcome the Vocabulary Bottleneck. In: DH-CASE '14: Collaborative Annotations in Shared Environments: metadata, tools and techniques in the Digital Humanities, Fort Collins, CA, september 2014. <b>Proceedings...</b> New York: Association for Computing Machinery, 2014. Disponível em: <a href="https://doi.org/10.1145/2657480.2657482">https://doi.org/10.1145/2657480.2657482</a> . Acesso em: abr. 2020.
48	QI, J.; LUO, D.; ZHU, L. Research on the collaborative method for vocabulary system constructions. In: INTERNATIONAL CONFERENCE ON COMPUTER AND AUTOMATION ENGINEERING, ICCAE 2010, 2., April 2010. <b>Proceedings...</b> Singapore: IEEE, 2010. p. 523–526. Disponível em: <a href="https://doi.org/10.1109/ICCAE.2010.5451268">https://doi.org/10.1109/ICCAE.2010.5451268</a> . Acesso em: abr. 2020.
49	RAJBHANDARI, S.; KEIZER, J. The AGROVOC Concept Scheme - A Walkthrough. <b>Journal of Integrative Agriculture</b> , vol. 11, n. 5, may 2012. Disponível em: <a href="http://dx.doi.org/10.1016/S2095-3119(12)60058-6">http://dx.doi.org/10.1016/S2095-3119(12)60058-6</a> . Acesso em: abr. 2020.
50	ROSATI, I. <i>et al.</i> A thesaurus for phytoplankton trait-based approaches: development and applicability. <b>Ecological Informatics</b> , v. 42, p. 129–138, 2017. Disponível em: <a href="https://doi.org/10.1016/j.ecoinf.2017.10.014">https://doi.org/10.1016/j.ecoinf.2017.10.014</a> . Acesso em: abr. 2020.
51	SANTOS, C. A. C. M. DOS <i>et al.</i> Elaboração de vocabulário controlado em formato SKOS usando Tematres: implicações metodológicas e web-semântica. In: DODEBEL, Vera; GUIMARÃES, José Augusto Chaves (Orgs.). <b>Complexidade e Organização do Conhecimento: desafios do nosso século</b> . Rio de Janeiro: Sociedade Brasileira de Organização do Conhecimento, 2013. p. 205–210. Disponível em: <a href="https://repositorio.usp.br/item/002433663">https://repositorio.usp.br/item/002433663</a> . Acesso em: abr. 2020.
52	SOLOMOU, G.; PAPTAEODOROU, T. The use of SKOS vocabularies in digital repositories: The DSpace case. In: INTERNATIONAL CONFERENCE ON SEMANTIC COMPUTING, ICSC 2010, 4. <b>Proceedings...</b> 2010. Pittsburgh, PA, USA: IEEE, 2010. p. 542–547. Disponível em: <a href="https://doi.org/10.1109/ICSC.2010.83">https://doi.org/10.1109/ICSC.2010.83</a> . Acesso em: abr. 2020.
53	SPAMPINATO, D.; ZANGARA, I. Classical antiquity and semantic content management on linked open data. In: DH-CASE'13: Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities, Florence, Italy, september 2013.

	<b>Proceedings...</b> New York: Association for Computing Machinery, 2013. Disponível em: <a href="https://doi.org/10.1145/2517978.2517992">https://doi.org/10.1145/2517978.2517992</a> . Acesso em: abr. 2020.
54	STELLATO, A. <i>et al.</i> A Collaborative Framework for Managing and Publishing KOS. In: EUROPEAN NETWORKED KNOWLEDGE ORGANISATION SYSTEMS (NKOS) WORKSHOP, 10., Berlin, 28-29 September 2011. <b>Proceedings...</b> Berlin, 2011. Disponível em: <a href="http://www.fao.org/docrep/article/am814e.pdf">http://www.fao.org/docrep/article/am814e.pdf</a> . Acesso em: abr. 2020.
55	STELLATO, A. Dictionary, Thesaurus or Ontology? Disentangling Our Choices in the Semantic Web Jungle. <b>Journal of Integrative Agriculture</b> , v. 11, n. 5, p. 710–719, 2012. Disponível em: <a href="http://dx.doi.org/10.1016/S2095-3119(12)60060-4">http://dx.doi.org/10.1016/S2095-3119(12)60060-4</a> . Acesso em: abr. 2020.
56	STELLATO, A. <i>et al.</i> VocBench: A Web Application for Collaborative Development of Multilingual Thesauri. In: GANDON, F. <i>et al.</i> (eds). <b>The Semantic Web, Latest Advances and New Domains, ESWC 2015: Lecture Notes in Computer Science</b> , vol 9088. Cham: Springer, 2015. p. 38–53. Disponível em: <a href="https://link.springer.com/chapter/10.1007/978-3-319-18818-8_3">https://link.springer.com/chapter/10.1007/978-3-319-18818-8_3</a> . Acesso em: abr. 2020.
57	TUDHOPE, D.; BINDING, C. Experiences with Knowledge Organization System Services from the STAR Project. <b>Signum</b> , n. 5, p. 1–9, 2008. Disponível em: <a href="https://journal.fi/signum/article/view/3483">https://journal.fi/signum/article/view/3483</a> . Acesso em: abr. 2020.
58	VANDERBILT, K. <i>et al.</i> A prototype system for multilingual data discovery of International Long-Term Ecological Research (ILTER) Network data. <b>Ecological Informatics</b> , v. 40, n. 2017, p. 93–101, 2017. Disponível em: <a href="http://dx.doi.org/10.1016/j.ecoinf.2016.11.011">http://dx.doi.org/10.1016/j.ecoinf.2016.11.011</a> . Acesso em: abr. 2020.
59	VITI, E. My First Ten Years: Nuovo Soggettario growing, development and integration with other Knowledge Organization Systems. <b>Knowledge Organization</b> , v. 44, n. 8, p. 624–637, 2018. Disponível em: <a href="https://doi.org/10.5771/0943-7444-2017-8-624">https://doi.org/10.5771/0943-7444-2017-8-624</a> . Acesso em: abr. 2020.
60	WANG, T. <i>et al.</i> A thesaurus and online encyclopedia merging method for large scale domain-ontology automatic construction. In: WANG, M. (ed.). <b>Knowledge Science, Engineering and Management. KSEM 2013: Lecture Notes in Computer Science</b> , vol 8041. Berlin, Heidelberg: Springer, 2013. p. 132–146. Disponível em: <a href="https://doi.org/10.1007/978-3-642-39787-5_11">https://doi.org/10.1007/978-3-642-39787-5_11</a> . Acesso em: abr. 2020.
61	WETZ, P. <i>et al.</i> Matching linked open data entities to local thesaurus concepts. In: SEMANTICS, I., Posters & Demonstrations Track. Graz, Austria, September 5-7, 2012. <b>Proceedings...</b> Europe: CEUR Workshop Proceedings, 2012. p. 6–11. Disponível em: <a href="http://ceur-ws.org/Vol-932/paper2.pdf">http://ceur-ws.org/Vol-932/paper2.pdf</a> . Acesso em: abr. 2020.

62	<p>WRIGHT, D. G.; HARRISON, K. A.; WATKINS, J. Automated tagging of environmental data using a novel SKOS formatted environmental thesaurus. <b>Earth Sci. Inform</b>, v. 8, p. 103–110, 2015. Disponível em: <a href="https://link.springer.com/article/10.1007/s12145-014-0183-1">https://link.springer.com/article/10.1007/s12145-014-0183-1</a>. Acesso em: abr. 2020.</p>
63	<p>XIAN, G.; SUN, W. Open, interlink and mashup multilingual agricultural data semantically based on KOSs. In: INTERNATIONAL CONFERENCE ON SYSTEMS AND INFORMATICS, ICSAI 2016, 3., Shanghai, China, 2017. <b>Proceedings...</b> Shanghai, China: IEEE, 2017. p. 405–409, 2017. Disponível em: <a href="https://doi.org/10.1109/ICSAI.2016.7810989">https://doi.org/10.1109/ICSAI.2016.7810989</a>. Acesso em: abr. 2020.</p>
64	<p>ZAPILKO, B. <i>et al.</i> TheSoz: a SKOS representation of the Thesaurus for the Social Sciences. <b>Semantic Web</b>, v. 4, n. 3, p. 257–263, 2013. Disponível em: <a href="https://dl.acm.org/doi/abs/10.5555/2786071.2786076">https://dl.acm.org/doi/abs/10.5555/2786071.2786076</a>. Acesso em: abr. 2020.</p>
65	<p>ZENG, M. L.; MAYR, P. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. <b>International Journal on Digital Libraries</b>, v. 20, n. 3, p. 1–22, 2018. Disponível em: <a href="https://doi.org/10.1007/s00799-018-0241-2">https://doi.org/10.1007/s00799-018-0241-2</a>. Acesso em: abr. 2020.</p>
66	<p>ZHU, L.; QIAO, X.; SI, L. Constructing an open-sharing Chinese Scientific &amp; Technical Vocabulary Service. In: INTERNATIONAL CONFERENCE ON NETWORKING AND DIGITAL SOCIETY, ICNDS 2009, Guiyang, Guizhou, China. <b>Proceedings...</b> Guiyang, Guizhou, China: IEEE, 2009. p. 284–287. Disponível em: <a href="https://doi.org/10.1109/ICNDS.2009.150">https://doi.org/10.1109/ICNDS.2009.150</a>. Acesso em: abr. 2020.</p>
67	<p>ZOU, Q. Represent Changes of Knowledge Organization Systems on the Semantic Web. <b>International Journal of Librarianship</b>, v. 3, n. 1, p. 67, 2018. Disponível em: <a href="https://doi.org/10.23974/ijol.2018.vol3.1.64">https://doi.org/10.23974/ijol.2018.vol3.1.64</a>. Acesso em: abr. 2020.</p>

## APÊNDICE E – INSTRUÇÕES PARA A ELABORAÇÃO DO PARECER



**Universidade Federal de Santa Catarina**  
**Programa de Pós-Graduação em Ciência da Informação**  
Campus Professor João David Ferreira Lima – Trindade – Florianópolis – Santa Catarina –  
Brasil – CEP: 88.040-900

### INSTRUÇÕES PARA A ELABORAÇÃO DO PARECER

**Título da pesquisa:** Proposta de modelo colaborativo para a construção e publicação de tesouros no contexto do *Linked Open Data*: aplicação no domínio da Música

Os questionamentos a seguir foram elaborados com o propósito de auxiliar na elaboração do parecer:

- a) O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?
- b) As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do *Linked Open Data*?
- c) De modo geral, o modelo atende as necessidades do domínio de aplicação?
- d) O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?

Ao final do parecer, solicito que os colaboradores informem se **aprovam** ou **desaprovam** o modelo. Para ambos os casos, os colaboradores poderão pontuar debilidades e sugerir melhorias.

## APÊNDICE F – PARECERES DOS ESPECIALISTAS

### PARECER: Especialista 1

**Título da pesquisa:** Proposta de modelo colaborativo para a construção e publicação de tesouros no contexto do *Linked Open Data*: aplicação no domínio da Música

**a) O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?**

“A proposta está fundamentada na concepção de que o processo colaborativo é resultado das atividades de interação entre os grupos de atores sociais, distribuídos em três dimensões, a saber: comunidade de usuários, grupos de gestão do vocabulário e a comunidade do LOD do domínio da Música”. Em relação à questão da letra “a”, como não consegui distinguir os papéis de cada equipe de gestão (ou editorial), em alguns momentos não ficou claro para mim as relações entre as três dimensões (colaboração, os princípios norteadores e as atividades de construção).

**b) As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do *Linked Open Data*?**

Em relação à questão da letra “b”, considero que o modelo não atende a todas as etapas para a construção de tesouros, ou não de uma maneira lógica que, na minha concepção, deveria ser estabelecido. Penso que o melhor para que as etapas ficassem alinhadas ao contexto do *Linked Open Data* seria o atendimento à lógica de etapas de construção recomendada pela norma ISO 25964, pois, sendo uma norma internacional, se todos os tesouros adotassem a mesma lógica de construção, teoricamente, seria mais promissor a conseguir o propósito de interoperabilidade semântica.

**c) De modo geral, o modelo atende as necessidades do domínio de aplicação?**

Em relação à questão da letra “c”, posso comentar que no elemento tradução, uma vez que a proposta é ser uma representação multilíngue, com interoperabilidade, percebe-se uma limitação, em especial por denominá-la de “interoperabilidade semântica”, sendo que parece que irá levar em consideração apenas o aspecto lexical. E mesmo sendo assim, deve-se pensar

em distintas questões que envolvem a equivalência entre diferentes idiomas. No âmbito da música, a representação de “escala”, por exemplo, varia conforme o contexto cultural no qual está inserido. Outro exemplo seria a denominação de instrumentos de percussão latino-americanos, que têm variação de designações, dependendo do lugar. Como o seu modelo envolve a construção do tesouro, e não somente o ambiente colaborativo, a questão da tradução, neste exemplo, teria que ser melhor definida.

**d) O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?**

Em relação à questão da letra “d”, considero que o Modelo colaborativo, como um todo, pode ser aplicado a outros domínios, para a construção de tesouros, desde que as considerações que fiz em meus comentários sejam levadas em consideração.

**[resultado da avaliação e sugestões]**

Conclusão: Penso que não é uma questão de aprovar ou desaprovar o modelo, que, acredito, pode atender ao que se propõe, mas que seja necessário ajustar alguns elementos para que ele possa ser testado e, então, aprovado ou não, ou ser refinado. Talvez, um dos problemas do seu modelo seja o de não considerar o Planejamento como parte da construção de tesouros, assim como não considerar a definição como essencial à modelagem conceitual do tesouro, quando se determinam a delimitação do conceito e as suas relações com outros conceitos na estrutura do tesouro, que, como consequência, evidenciam as relações semânticas. Também é preciso delimitar o que é modelagem conceitual daquilo que é modelagem de dados, e os princípios que são aplicáveis a cada um deles. Considero que também é necessário determinar os diferentes grupos na equipe editorial, delimitando o conceito e os seus papéis, assim como determinar quais os tipos de permissões adotados e o processo de *workflow* correspondente.

## PARECER: Especialista 2

**Título da pesquisa:** Proposta de modelo colaborativo para a construção e publicação de tesouros no contexto do *Linked Open Data*: aplicação no domínio da Música

**e) O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?**

Sim, o modelo proposto é bem completo neste aspecto.

**f) As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do *Linked Open Data*?**

Sim, em parte, ver os comentários.

**g) De modo geral, o modelo atende as necessidades do domínio de aplicação?**

Sim, em parte, ver os comentários.

**h) O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?**

Poderia, mas com vários ajustes.

### [resultado da avaliação e sugestões]

Foi difícil avaliar o modelo porque o texto enviado não trazia um sumário que nos permitisse compreender não só o modelo, mas também seu processo de concepção, nem os dados da sua fase de Pesquisa de Campo. Não tivemos acesso aos dados do questionário aplicado nem a que instituições ele foi aplicado.

O modelo é bastante completo e abrangente, considerando diversos aspectos da construção de um tesouro no contexto *Linked Open Data*. Aspectos importantes, no entanto, foram pouco desenvolvidos ou desconsiderados, como descrito a seguir.

- As metodologias conceituais/teóricas de construção de tesouros, como teoria da Classificação Facetada, de Ranganathan (1967), a teoria do Conceito, de Dahlberg (1978) e a teoria da terminologia (CABRÉ, 2005). Estas teorias fornecessem os fundamentos, implicam e têm impacto nos diferentes procedimentos metodológicos arrolados no modelo. Estas diferenças deveriam ser consideradas.

- O modelo é de difícil implementação prática, devido à sua complexidade, em especial em realidades como a brasileira (mas também no contexto internacional, os chamados vocabulários “zumbis” ou “fantasmas”, mencionados em NISO (2017). No Brasil



existe um histórico de projetos de tesouros paralisados, concluídos, descontinuados. Poder-se-ia listar o Tesouros de Objetos de Patrimônio, <http://www.tesauromuseus.com.br/download/tesauro.pdf>, o Tesouro do Folclore e Cultura Popular Brasileira, o Tesouro de Acervos Científicos, [http://site.mast.br/hotsite\\_museologia/thesaurus.html](http://site.mast.br/hotsite_museologia/thesaurus.html), a Lista de Cabeçalhos de Assunto da Rede Bibliodata (LCARB), entre outros.

- Existe também uma conhecida dificuldade de as instituições brasileiras trabalharem em projetos interinstitucionais, sua estrutura organizacional está pouco preparada para este tipo de projeto. O questionário enviado às instituições ajudaria a avaliar melhor esta questão.

- O modelo colaborativo para a construção e publicação de tesouros menciona comunidades, equipes, etapas de planejamento do projeto, mas não sua governança. A questão da governança dos projetos ao longo do tempo, para além da sua concepção, seria um requisito importante a ser contemplado no modelo.

- O modelo se foca no desenvolvimento de tesouros a partir do zero, quando significativa literatura internacional mostra o quanto é complexo o desenvolvimento de um tesouro com características Linked Open Data – os LOV – Linked Open Vocabularies (MENDEZ e GREENBERG, 2012), (ZENG, 2019).

- O modelo não aborda a perspectiva do reuso de outros vocabulários já existentes. Assim, mostra falta de sintonia com as tendências internacionais de desenvolvimento de LOV, onde o reuso parece mais realista e adaptável ao cenário nacional.

- O modelo não aborda também uma questão chave que é a manutenção ao longo do tempo de um tesouro (NISO, 2017). Reportamos o que já destacamos sobre os vocabulários “zumbis” e “fantasmas”, em especial no contexto brasileiro. A manutenção é crítica para manter um tesouro vivo, garantindo assim seu reuso. Como acentuamos anteriormente, esta questão esbarra no problema institucional, uma fragilidade em nosso país.

- **Aprovamos o modelo.** A validade do mesmo reside na sua completeza, ressalvados os aspectos mencionados acima. Seria algo como um modelo de referência. Sua complexidade obrigada que sejam feitas adaptações para que ele possa ser adaptado em projetos e contexto específicos.

### PARECER: Especialista 3

#### PARECER ACERCA DO MODELO PROPOSTO

**Título da pesquisa:** Proposta de modelo colaborativo para a construção e publicação de tesouros no contexto do *Linked Open Data*: aplicação no domínio da Música

**a) O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?**

O modelo descreve, de maneira clara e didática, as relações entre as dimensões de colaboração, os princípios norteadores e o fluxo de atividades de construção e validação do tesouro. Tal proposta encontra-se ancorada nas diretrizes que norteiam o processo de construção e publicação de tesouros no contexto do LOD.

**b) As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do *Linked Open Data*?**

Considera-se que as etapas contempladas no modelo foram estruturadas de maneira lógica, atendendo as necessidades de construção de tesouros nesse contexto. Essas etapas estão vinculadas a um conjunto de atividades colaborativas que objetivam a cooperação de sujeitos, denominados atores sociais, no processo de representação da informação musical.

**c) De modo geral, o modelo atende as necessidades do domínio de aplicação?**

**d) O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?**

Em linhas gerais, depreende-se que o modelo proposto atende satisfatoriamente as necessidades do domínio de aplicação, podendo servir como referência para ser adaptado/aperfeiçoado para fins de implementação em outros contextos ou domínios.

#### [resultado da avaliação e sugestões]

O parecer em questão avalia a proposta de modelo colaborativo para a construção e publicação de tesouros no contexto *Linked Open Data* (LOD) no domínio da música, universo informacional cujo conteúdo merece e necessita ser especificamente representado, sobretudo a partir da garantia do usuário.

Visando o aprimoramento do modelo proposto, e considerando o seu caráter colaborativo, sugere-se uma atenção especial à dimensão da comunidade de usuários cuja participação está centrada no compartilhamento de suas experiências por meio de fóruns públicos de discussão, de comentários ou de pesquisas na interface de busca realizadas por esses sujeitos informacionais.

As características de um modelo podem influenciar nos incentivos ou motivações que conduzem os sujeitos a utilizarem e participarem em um sistema colaborativo. Diante dessa perspectiva, considera-se importante elucidar na descrição do modelo proposto a existência de outras estratégias que podem ser utilizadas para agregar o potencial da colaboração dos atores sociais e das práticas folksonômicas no processo de atribuição, recomendação e validação de etiquetas e conceitos relacionados ao domínio da música (além daquelas apresentadas na proposta). Para tanto, os gestores devem analisar: os elementos constituintes de um modelo colaborativo de representação da informação; a capacidade de aplicação desses elementos no sistema a partir de suas funcionalidades; a capacidade de manutenção da integridade do vocabulário; a interoperabilidade dos dados ou integração dessas ferramentas com outros sistemas de informação.

Na etapa de planejamento, especificamente no que concerne à definição dos usuários, considera-se importante ressaltar a necessidade de gestores realizarem estudos de usuários com vistas a identificar os seus perfis, níveis de especialidade, interesses e motivações nas práticas colaborativas de representação da informação. Além dos músicos, docentes, pesquisadores e público em geral (mencionados no modelo), existem outros tipos de usuários no domínio da música que carecem ser mencionados na proposta, são eles: graduandos em instrumentos, estudantes de conservatórios e de escolas de música com ênfase prática ou teórica. As necessidades informacionais desse grupo estão interligadas, essencialmente, às informações sonoras, como dificuldade de execução, instrumentos musicais, além de informações de produção, como compositor, banda e intérprete.

Em linhas gerais, reitera-se que a pesquisa doutoral apresenta uma proposta inovadora e construtivista para a Organização e Representação do Conhecimento, bem como para unidades e sistemas de recuperação da informação no domínio da música. Diante do exposto, considero o modelo proposto “**Aprovado**”.

#### **PARECER: Especialista 4**

**Título da pesquisa:** Proposta de modelo colaborativo para a construção e publicação de tesouros no contexto do *Linked Open Data*: aplicação no domínio da Música

**a) O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?**

O modelo proposto descreve com clareza as relações entre as dimensões de colaboração, os princípios que norteiam a publicação de tesouros em *Linked Open Data* e as etapas necessárias à construção desses tesouros. Os atores sociais e os recursos envolvidos nas práticas de colaboração estão suficientemente indicados na proposta. Os princípios norteadores estão bem sintetizados e encontram respaldo na literatura internacional. No decorrer da apresentação da proposta, a menção às práticas e aos princípios é recorrente.

**b) As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do *Linked Open Data*?**

As etapas para a construção e publicação do tesouro estão dispostas de maneira lógica, sendo os resultados de cada etapa frequentemente evidenciados como insumo para a etapa seguinte. Embora se destinem a uma abordagem mais teórica e generalista dos processos, provendo conhecimentos amplos acerca dos agentes, ferramentas e dados envolvidos – uma vez que se trata de um modelo, e não de um manual – essas etapas apresentam os elementos necessários até mesmo para subsidiar projetos práticos, no nível operacional, para a criação e publicação de tesouros como *Linked Open Data*.

Essas etapas do modelo estão embasadas na literatura nacional e internacional – normas, ensaios teóricos, relatos de experiência e de iniciativas, tanto sobre as linguagens documentais quanto sobre *Linked Open Data* – o que indica que sua suficiência e viabilidade têm sido objeto de estudo, ainda que possam apresentar limitações ou pontos de divergência.

**c) De modo geral, o modelo atende as necessidades do domínio de aplicação?**

O modelo foi proposto para um domínio específico, o da Música. Os conceitos e os termos deste domínio são a matéria-prima para o tesouro e os atores sociais do domínio fazem parte de diversas etapas do modelo, contribuindo nos processos de sua competência. No decorrer da apresentação das etapas, nota-se uma preocupação em trazer elementos do domínio

da Música para a discussão e/ou exemplificação dos processos, atores, instrumentos, etc., com isso, a proposta busca demonstrar que atende as necessidades do domínio.

**d) O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?**

As complexidades envolvendo os termos e os conceitos da Música, bem como seus atores e instrumentos, possuem equivalentes em outros domínios: variações léxicas, subordinação, superordenação de conceitos, etc. existem em todas as áreas do conhecimento, assim como usuários leigos, usuários especializados e fontes de informação com diferentes níveis de especificidade (dicionários, enciclopédias, ontologias, etc.). Assim, entendo que o domínio abordado na proposta atua mais como um exemplo de aplicação do que um limitador do modelo. Desse modo, o modelo pode ser aplicado em outros contextos ou domínios, seja em sua totalidade, nos casos em que ainda não há um tesouro, ou parcialmente, nos casos em que a pretensão é publicar como *Linked Open Data* um tesouro já existente.

**[resultado da avaliação e sugestões]**

Considerando os pontos destacados, **aprovo o modelo** e, na condição de sugestões para melhoria da descrição de suas etapas, proponho as seguintes alterações:

- a) termos e conceitos: revisar os trechos em que “termos” e “conceitos” são utilizados em conjunto de modo a evidenciar a distinção entre eles, visto que essa distinção pode não ser evidente para o leitor; ignore essa sugestão se a tese já dedicar outra seção para essa distinção;
- b) vocabulários de propriedades e formatos de serialização: revisar as menções aos vocabulários de propriedades (SKOS, SKOS-XL) e aos formatos de serialização dos dados (XML, CSV, JSON, N-Triples, Notation-3, etc.); é importante esclarecer ao leitor o que opera no nível do modelo de dados (ou seja, na “modelagem” do tesouro) e o que opera no nível das aplicações de informática (os dados modelados com o SKOS pode ser exportados em XML, JSON, N-Triples, etc., em apenas um ou em mais de um formato, conforme for definido nas “camadas” de saída/exportação dos dados); as figuras construídas para a apresentação do modelo estão muito didáticas, talvez, uma figura possa ser usada para explicar a serialização;
- c) definição dos URIs: entendo que a definição dos URIs é algo estratégico da publicação de *Linked Open Data*, portanto, deveria preceder a conversão dos registros, logo após a modelagem e o mapeamento (inclusive, dependendo das ferramentas utilizadas, pode ser necessário já dispor dos URIs antes do mapeamento).

## PARECER: Especialista 5

**Título da pesquisa:** Proposta de modelo colaborativo para a construção e publicação de tesouros no contexto do *Linked Open Data*: aplicação no domínio da Música

**a) O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?**

As relações são bem descritas e representadas. Há no trabalho apresentado uma riqueza de detalhes com a apresentação de figuras, fluxogramas e suficiência descritiva em cada etapa do processo.

**b) As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do *Linked Open Data*?**

O conjunto de atividades indicadas “Planejamento”, “Aquisição”, “Seleção”, “Preparação dos dados”, “Interligação”, “Validação”, “Publicação”, “Feedback”, são bastante pertinentes e estão organizados de maneira lógica. As etapas de “Preparação dos dados”, “Interligação dos Dados” e “Validação dos Dados” contextualizam o modelo no *Linked Open Data*.

**c) De modo geral, o modelo atende as necessidades do domínio de aplicação?**

Em termos teóricos o modelo apresentado é consistente e pode atender o domínio da Música, desde que a etapa de planejamento, com a definição do público alvo, nível de especialidade, fontes de referência e escolha dos conjuntos de interligação, seja bem delimitado.

O domínio da música é extremamente complexo, talvez até mais complexo que outros domínios como a mecânica, física, biologia etc.

A variação terminológica é tanta que confunde até mesmo os profissionais da área, exemplo disso a icônica entrevista do Jô Soares à Zeca Pagodinho no qual temos um questionamento ao músico sobre a diferença entre samba de roda, pagode, partido alto e sambarenredo.

Exemplo da complexidade terminológica do domínio da música pode ser notado no “Quadro 7 - Critérios para a definição de termos preferenciais do tesouro” temos por exemplo, nos “Aspectos linguísticos e culturais”

Exemplo:

Termo preferido: Acordeão

Usado para Sanfona

Usado para Gaita

Aqui: Acordeão, Sanfona e Gaita podem ser utilizados para identificar um mesmo instrumento, mas também podem ser utilizados para designar instrumentos bem diferentes.

E no “Neologismos, gírias e jargões”

Exemplos:

Termo preferido: Música popular

Usado para Música pop

Aqui: Música popular e música pop não são conceitos sinônimos ou quase-sinônimos.

Fato: há ganhos sensíveis na interligação de dados do Modelo de Tesouros proposto, trazendo múltiplas relações para a Comunidade de Usuários. Contudo, minha desconfiança na aplicação do tesouro LOD no domínio da Música é que sendo o domínio “complexo” e as plataformas MusicBrainz e do Wikidata colaborativas, a reconciliação de dados automatizada, gerem uma alta revocação quando da recuperação de informações para o usuário.

#### **d) O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?**

O modelo tem enorme potencial para aplicação em outros contextos e domínios, inclusive, valeria pensar a contribuição desse modelo na construção de ontologias, avançando na contribuição de Boccato e Ramalho (2008).

BOCCATO, Vera R. C.; Ramalho, R. A. S.; Fujita, M. S. L. A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. // Ibersid. (2008) 199-209.

#### **[resultado da avaliação e sugestões]**

O modelo apresentado é uma enorme contribuição para a área, e o considero **APROVADO**. Destaco a seguir algumas pontuações sobre o modelo que me fizeram pensar, mas que de modo algum desqualificam o que foi apresentado.

A “Figura 1 - Aspectos gerais das práticas colaborativas na construção dos vocabulários controlados do domínio da Música”, constrói uma representação de um sistema de mediação, de tal modo que, articula três categorias: de um lado você tem um público “comunidade de usuários”, de outro um conjunto de (meta)informações, fontes de referência, no caso o “domínio da música”, e no meio um “terceiro” que media as relações entre as duas categorias, a “Equipe editorial”. Essa representação poderia, talvez, ser pensada em “outros termos”:

A relação entre “Equipe editorial” e “Domínio da música”, provavelmente é bidirecional. O Domínio aciona e age sobre a Equipe Editorial. O Domínio está em constante expansão, o que gera a necessidade da Equipe Editorial revisar periodicamente a compilação

do “Vocabulário Controlado”. Talvez possamos dizer que aqui, o “Domínio da Música”, imponha uma relação de “Produção literária” (Dispersão terminológica) à Equipe editorial.

A relação entre “Comunidade de Usuários” e “Equipe editorial”, provavelmente é bidirecional. A Equipe Editorial atua sobre a seleção dos termos e, portanto, atua sob a Comunidade de Usuários. Talvez possamos dizer que a “Equipe editorial”, ofereça uma nova “Garantia estrutural”, à Comunidade de usuários.

Nessa representação, a terminologia “Equipe Editorial”, parece-me insuficiente. O que se vê é um “Dispositivo Editorial / Terminológico”, no qual há uma equipe e recursos tecnológicos que atuam de maneira sistemática e sistêmica, para criar e manter o Vocabulário Controlado, um dispositivo terminológico.

A “Figura 2 - Modelo colaborativo para a construção e publicação de tesouros do domínio da Música como Linked Open Data” representa o relacionamento entre os princípios, atividades e atores sociais. Os princípios colocados “Consenso, Controle editorial, Representação multilíngue, Interoperabilidade semântica, Avaliação da qualidade, Acesso, Reuso”, são extremamente pertinentes, mas parecem que compõe um sistema cíclico em torno do “Tesouros como Linked Open Data”, talvez seria o caso de pensar que esses princípios atuam sob a forma de rizoma, atrelados a um “nó” central, a: “Gestão do Tesouros como Linked Open Data”.

Ademais, destaco a algumas pontuações sobre o modelo a fim de enriquecê-lo, mas que de modo algum desqualificam o que foi apresentado.

Adicionar os linguistas em todas as etapas do modelo. Digo isso, por que nas representações (sob forma de figuras, que iniciam cada seção, de maneira muito didática), consta acertadamente os atores sociais “profissionais da informação e especialistas em Tecnologias de informação”, mas os linguistas tem também grande importância, por deterem um conhecimento em terminologia muito mais apurado que os demais profissionais indicados.

Adicionar no planejamento, ao identificar as primeiras definições, tais como: “público alvo, objetivos e escopo do tesouro”, algumas indicações que parecem óbvias, mas que tem relevância, tais como: “Qual será o software/sistema empregado”, “Quais serão as diretrizes gerais: de relacionamento entre termos, redação de notas, idiomas cobertos etc.”, “Qual o cronograma das ações”.

Nas Atividades, valeria pensar a substituição do termo “Feedback” por “Retroalimentação” ou “Avaliação”, apenas por uma questão de padronização linguística, uma vez que todas as outras etapas estão em português.



### **PARECER: Especialista 6**

Parecer sobre a pesquisa Proposta de modelo colaborativo para a construção e publicação de tesouros no contexto do *Linked Open Data*: aplicação no domínio da Música

**a) O modelo descreve com clareza as relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção?**

A presente pesquisa é muito importante no que diz respeito às questões ligadas a Organização do conhecimento na Web, sobretudo porque disserta acerca das questões que favorecem a recuperação da informação e reuso de dados na Web. O modelo proposto descreve com clareza e leva a compressão das relações entre as dimensões de colaboração, os princípios norteadores e as atividades de construção, evidenciando como é fundamental a existência e funcionamento de cada uma das partes envolvidas.

**b) As etapas foram estruturadas de maneira lógica? Elas atendem as necessidades de construção de tesouros no contexto do *Linked Open Data*?**

As etapas propostas foram estruturadas de maneira lógica, pois seguem uma ordem de desenvolvimento do fluxo de trabalho que atendem as demandas exigidas para a criação de tesouro de maneira geral. No contexto do *Linked Open Data* elas atendem as necessidades de construção de tesouros, visto que estão sua ordenação e descrição estão voltadas para facilitar que a organização da informação funcione em um ambiente web.

**c) De modo geral, o modelo atende as necessidades do domínio de aplicação?**

Acreditamos que o instrumento proposto atende as necessidades do domínio de aplicação, visto que as fontes consultadas para a coleta de termos são especializadas no domínio e também o pesquisador buscou apoio das instituições e profissionais que trabalham no domínio. Esses passos favoreceram o enriquecimento de termos e atividades para a elaboração da proposta.

**d) O modelo pode ser adaptado para a aplicação em outros contextos ou domínios?**

O instrumento pode ser adaptado ao outro domínio, desde que ao se fazer as adaptações os profissionais sigam os mesmos passos para a elaboração de sua proposta, como a consulta as fontes e instituições ligadas ao domínio no qual queira fazer a adaptação.

**[resultado da avaliação e sugestões]**

Por fim, **aprovo o modelo proposto**, visto que fica claro a forma como foi construído e enquanto produto final para o domínio da música dentro do Linked Data, pode suprir as necessidades existentes de organização e recuperação da informação.