



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO  
CONHECIMENTO

Murilo Silveira Gomes

**UM MODELO DE AVALIAÇÃO DE CONFORMIDADE DE PORTAIS DE DADOS  
ABERTOS DE GOVERNO**

Florianópolis  
2021

Murilo Silveira Gomes

**UM MODELO DE AVALIAÇÃO DE CONFORMIDADE DE PORTAIS DE DADOS  
ABERTOS DE GOVERNO**

Tese submetida ao Programa de Pós-graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do título de doutor em Engenharia e Gestão do Conhecimento.

Orientador: Prof. Fernando Álvaro Ostuni Gauthier, Dr.

Coorientador: Prof. José Leomar Todesco, Dr.

Florianópolis

2021

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

Silveira Gomes, Murilo  
UM MODELO DE AVALIAÇÃO DE CONFORMIDADE DE PORTAIS DE  
DADOS ABERTOS DE GOVERNO / Murilo Silveira Gomes ;  
orientador, Fernando Álvaro Ostuni Gauthier,  
coorientador, José Leomar Todesco, 2021.  
263 p.

Tese (doutorado) - Universidade Federal de Santa  
Catarina, Centro Tecnológico, Programa de Pós-Graduação em  
Engenharia e Gestão do Conhecimento, Florianópolis, 2021.

Inclui referências.

1. Engenharia e Gestão do Conhecimento. 2. Portal de  
Dados Abertos. 3. Avaliação. 4. Automatização. 5.  
Conformidade. I. Ostuni Gauthier, Fernando Álvaro. II.  
Leomar Todesco, José. III. Universidade Federal de Santa  
Catarina. Programa de Pós-Graduação em Engenharia e Gestão  
do Conhecimento. IV. Título.

Murilo Silveira Gomes

**UM MODELO DE AVALIAÇÃO DE CONFORMIDADE DE PORTAIS DE DADOS  
ABERTOS DE GOVERNO**

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

Prof. Rogério Cid Bastos, Dr.  
Universidade Federal de Santa Catarina

Prof. Alexandre Augusto Biz, Dr.  
Universidade Federal de Santa Catarina

Prof. José Maria Parente de Oliveira, Dr  
Instituto Tecnológico de Aeronáutica

Prof. Mario Antônio Ribeiro Dantas, Dr  
Universidade Federal de Juiz de Fora

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Engenharia e Gestão do Conhecimento.

---

Coordenação do Programa de  
Pós-Graduação

---

Prof. Fernando Álvaro Ostuni Gauthier, Dr.  
Orientador

Florianópolis, 2021.

Este trabalho é dedicado à minha esposa Lidiane, aos meu pais Levy e Kátia e aos meus avós João e Nilceia.

## **AGRADECIMENTOS**

A realização deste trabalho teria sido ainda mais difícil se não fosse o apoio que recebi. Deste modo, quero agradecer a algumas das pessoas e instituições que me deram este suporte.

Primeiramente agradecer a Deus pelas oportunidades que me fizeram chegar até aqui e os desafios que foram impostos a mim, de modo a proporcionar uma evolução pessoal, emocional e profissional. Agradeço especialmente e imensamente a minha esposa Lidiane, meu amor, que se não fosse pelo suporte, auxílio e incentivo não teria chegado até o final desta jornada. Aos meus pais Levy e Kátia por ensinarem a terminar sempre o que começou e meus avós João e Nilceia que me ajudaram ao longo desta vida.

Deixo aqui o meu registro do agradecimento a Universidade Federal de Santa Catarina e ao Departamento de Engenharia e Gestão do Conhecimento e toda sua equipe técnica, docente e discente que me proporcionaram muitas oportunidades e desafios que no qual tive orgulho de participar. A CAPES por proporcionar o auxílio financeiro que proporcionou chegar a conclusão desta pesquisa.

Meu especial agradecimento ao professor Fernando Gauthier pela confiança, oportunidades e por me escolher como orientando dentre muitos e ao coorientador professor José Leomar Todesco (Tite) por todo o auxílio e conversas que me ajudaram na construção deste trabalho.

Não poderia deixar de agradecer aos amigos que fiz durante esta jornada, aos colegas dos laboratórios Laboratório de Engenharia do Conhecimento (LEC), Núcleo de Estudos em Inteligência, Gestão e Tecnologia para Inovação (IGTI), Laboratório de Educação a Distância (LED) e dentre muitos outros deixo o meu muito obrigado, pelas nossas conversas, cafés e confraternizações. Não posso me esquecer do Righetto um amigo que proporcionou na medida do possível o uso da infraestrutura de Tecnologia da Informação (TI) para que pudesse realizar os experimentos e a pronta disposição para solucionar os problemas que iam surgindo. A todos vocês conto com vocês no futuro.

Agradecer a banca de qualificação que contribuiu com a lapidação da estrutura deste trabalho, bem como agradecer previamente a banca de defesa que aceitou o desafio de contribuir e auxiliar no refinamento final da pesquisa realizada.

*"O homem científico não pretende  
alcançar um resultado imediato.  
Ele não espera que suas ideias  
avançadas sejam imediatamente  
aceitas.  
Seus trabalhos são como  
sementes para o futuro.  
Seu dever é lançar as  
bases para aqueles que  
estão por vir e apontar  
o caminho. "*  
*(Nikola Tesla)*

## RESUMO

Dados tem grande relevância para a sociedade, pois através destes é possível detectar falhas, fraudes, índice de crescimento populacional, entre outros. Assim, percebe-se que a publicação dos dados tem se tornado fundamental para os cidadãos. Os dados disponibilizados em portais de organizações públicas e privadas contemplam diferentes categorias, estas despertam interesses de diversos setores, para vários objetivos de uso, tais como: novas pesquisas, realização de análises, visualização ou mesmo a criação de novos negócios. Deste modo, após anos do início do movimento de dados abertos, a publicação dos dados está aquém das expectativas, visto que, geralmente os portais de dados abertos funcionam como repositórios para viabilizar apenas a transparência. Observam-se diversos problemas de conformidade, tais como: metadados inconsistentes, baixa qualidade dos conjuntos de dados, falta de padronização, ausência de gerenciamento e manutenção. O presente trabalho tem por objetivo compor um modelo para avaliação de portais de dados abertos considerando as boas práticas e os aspectos existentes na literatura para a verificação da conformidade utilizando a *Design Science Research Methodology* (DSRM). Deste modo, o modelo desenvolvido permite avaliar elementos relacionados aos conjuntos de dados e seus recursos, bem como da estrutura do portal. Também foi definido o processo de avaliação que auxilia na aplicação do modelo, destacando-se que é possível fazer uso de diferentes parâmetros conforme o cenário a ser avaliado. Para auxiliar na coleta, aplicação do modelo e análise dos resultados foi desenvolvido um sistema computacional. Foram obtidas as informações para realização da prova de conceito considerando três portais: governo federal, do estado do Rio Grande do Sul e do município de São Paulo. Através dos resultados obtidos percebeu-se que o modelo permite a avaliação da conformidade de portais de dados abertos. Também foi possível o monitoramento, através das coletas realizadas em diferentes períodos de um portal, onde evidenciou-se a variação de algumas dimensões. Para complementar a avaliação do modelo e do sistema computacional foram consultados especialistas da área de dados abertos e evidenciou-se que o modelo de referência e o sistema desenvolvido possuem relevância para a área. Apresentada a estrutura do modelo, os resultados obtidos através dos cenários de uso e finalizada a consulta aos especialistas concluiu-se que a pesquisa desenvolvida contribuiu fornecendo uma percepção das limitações dos portais de dados abertos perante as recomendações existentes.

**Palavras-chave:** Portal de Dados Abertos. Avaliação. Automatização. Conformidade

## ABSTRACT

Data has great relevance to society, as through them it is possible to detect failures, fraud, population growth rate, among others. Thus, it is clear that the publication of data has become essential for citizens. The data made available on portals of public and private organizations include different categories, these arouse interests from different sectors, for various purposes of use, such as: new research, analysis, visualization or even the creation of new businesses. Thus, years after the beginning of the open data movement, the publication of data is below expectations, since generally open data portals work as repositories to enable only transparency. There are several compliance issues, such as: inconsistent metadata, poor quality of data sets, lack of standardization, lack of management and maintenance. The present work aims to compose a model for evaluating open data portals considering good practices and existing aspects in the literature for the verification of compliance using the DSRM. Thus, the developed model allows evaluating elements related to data sets and their resources, as well as the structure of the portal. The evaluation process that assists in the application of the model was also defined, highlighting that it is possible to use different parameters according to the scenario to be evaluated. To assist in the collection, application of the model and analysis of the results, a computational system was developed. Information was obtained to carry out the proof of concept considering three portals: the federal government, the state of Rio Grande do Sul and the municipality of São Paulo. Through the results obtained, it was noticed that the model allows the assessment of compliance of open data portals. Monitoring was also possible, through collections carried out at different periods of a portal, where the variation of some dimensions was evidenced. To complement the evaluation of the modelo and the computacional system, specialists in the area of open data were consulted and it became evident that the reference modelo and the developed system are relevant to the area. After presenting the structure of the model, the results obtained through the use scenarios and after consulting the experts, it was concluded that the developed research contributed by providing a perception of the limitations of open data portals in view of the existing recommendations.

**Keywords:** Open Data Portal. Assessment. Automation. Conformity.

## LISTA DE FIGURAS

Figura 1 – Etapas da Metodologia <i>Design Science Research</i> . . . . .	32
Figura 2 – Etapas da DSRM x Estágios da Pesquisa . . . . .	35
Figura 3 – Vantagens da utilização dos dados . . . . .	42
Figura 4 – Nuvem LOD em 2021 . . . . .	45
Figura 5 – Dados na Web Vs Dados Abertos Vs Dados Conectados . . . . .	46
Figura 6 – Classificação com base em cinco estrelas . . . . .	47
Figura 7 – Arquitetura da Web Semântica . . . . .	52
Figura 8 – Portal de Dados Abertos Governo Federal do Brasil - CKAN . . . . .	61
Figura 9 – Tela da plataforma Socrata . . . . .	62
Figura 10 – Tela da plataforma Opendatasoft . . . . .	63
Figura 11 – Classificação dos artigos encontrados na literatura . . . . .	65
Figura 12 – Complementação da classificação dos artigos encontrados na literatura	72
Figura 13 – A estrutura do modelo . . . . .	81
Figura 14 – Estrutura de Dimensão e indicadores do modelo . . . . .	88
Figura 15 – Processo de avaliação utilizando do modelo . . . . .	93
Figura 16 – Tela coleta - SiAPDA . . . . .	94
Figura 17 – Tela nova coleta - SiAPDA . . . . .	94
Figura 18 – Tela aplicar modelo - SiAPDA . . . . .	95
Figura 19 – Tela lista padrões - SiAPDA . . . . .	96
Figura 20 – Tela definir padrões - SiAPDA . . . . .	97
Figura 21 – Tela resultados sem uso de padrões - SiAPDA . . . . .	101
Figura 22 – Tela resultado da avaliação - SiAPDA . . . . .	102
Figura 23 – Arquitetura do software - SiAPDA . . . . .	105
Figura 24 – Modulo Coleta - SiAPDA . . . . .	106
Figura 25 – Modulo Modelo - SiAPDA . . . . .	107
Figura 26 – Nuvem de palavras-chave usadas no portal do governo federal . . .	232
Figura 27 – Nuvem de palavras-chave usadas no portal do estado do Rio Grande do Sul . . . . .	232
Figura 28 – Nuvem de palavras-chave usadas no portal do município de São Paulo	233

## LISTA DE GRÁFICOS

Gráfico 1 – Resultado - Dimensão Acesso . . . . .	112
Gráfico 2 – Comparação do Score - Dimensão Acesso . . . . .	113
Gráfico 3 – Resultados Variáveis - Dimensão Formatos . . . . .	115
Gráfico 4 – Comparação do Score - Dimensão Formatos . . . . .	116
Gráfico 5 – Resultados Variáveis – Dimensão Direitos . . . . .	117
Gráfico 6 – Comparação do Score – Dimensão Direitos . . . . .	119
Gráfico 7 – Resultado indicadores – Dimensão Semântica . . . . .	120
Gráfico 8 – Comparação do Score – Dimensão Semântica . . . . .	121
Gráfico 9 – Resultado indicadores – Dimensão Estrutura . . . . .	125
Gráfico 10 – Comparação do Score – Dimensão Estrutura . . . . .	126
Gráfico 11 – Resultado indicadores – Dimensão Apresentação . . . . .	128
Gráfico 12 – Comparação do Score - Dimensão Apresentação . . . . .	128
Gráfico 13 – Resultado Dimensões . . . . .	131
Gráfico 14 – Resultado Score Final – (Escala de 0 – 10) . . . . .	132
Gráfico 15 – Respostas critério 1 – Questão 1 . . . . .	133
Gráfico 16 – Respostas critério 1 – Questão 2 . . . . .	133
Gráfico 17 – Respostas Critério 1 – Questão 3 . . . . .	134
Gráfico 18 – Respostas Critério 1 – Questão 4 . . . . .	134
Gráfico 19 – Respostas Critério 1 - Questão 5 . . . . .	135
Gráfico 20 – Respostas Critério 1 - Questão 6 . . . . .	135
Gráfico 21 – Respostas Critério 2 - Questão 1 . . . . .	136
Gráfico 22 – Respostas Critério 2 - Questão 2 . . . . .	136
Gráfico 23 – Respostas Critério 2 - Questão 3 . . . . .	137
Gráfico 24 – Respostas Critério 2 - Questão 4 . . . . .	137
Gráfico 25 – Queda no quantitativo de conjuntos de dados disponibilizados ao longo dos anos . . . . .	140
Gráfico 26 – Série histórica dos conjuntos de dados e recursos - Portal Governo Federal . . . . .	225
Gráfico 27 – Série histórica dos conjuntos de dados e recursos - Portal Estado do Rio Grande do Sul . . . . .	226
Gráfico 28 – Série histórica dos conjuntos de dados e recursos - Portal Município de São Paulo . . . . .	226
Gráfico 29 – Informações complementares sobre o indicador versionamento . . . . .	236
Gráfico 30 – Informações complementares sobre o indicador URI - Gráficos HTTP Status Code . . . . .	238
Gráfico 31 – Informações complementares sobre o indicador Formato de dados . . . . .	240
Gráfico 32 – Informações complementares sobre o indicador Licença . . . . .	242

Gráfico 33 – Temas x Conjuntos de dados - Portal governo federal . . . . .	252
Gráfico 34 – Temas x Conjuntos de dados - Portal estado do Rio Grande do Sul	253
Gráfico 35 – Temas x Conjuntos de dados - Portal município de São Paulo . . .	254

## LISTA DE QUADROS

Quadro 1 – Trabalhos que abordam discussões sobre dados abertos e dados abertos conectados . . . . .	29
Quadro 2 – Trabalhos que abordam discussões sobre indicadores . . . . .	30
Quadro 3 – Trabalhos que abordam discussões sobre portais . . . . .	30
Quadro 4 – Áreas e Setores que pesquisam sobre dados abertos . . . . .	42
Quadro 5 – Tipos licença <i>Creative Commons</i> . . . . .	50
Quadro 6 – Tipos licença <i>Open Data Commons</i> . . . . .	51
Quadro 7 – Desafios e boas práticas de dados na web . . . . .	54
Quadro 8 – Trabalhos relacionados . . . . .	67
Quadro 9 – Grupo de conceitos frequentes na literatura . . . . .	75
Quadro 10 – Grupo de conceitos eventuais na literatura . . . . .	78
Quadro 11 – Concepção dos indicadores do modelo de referência . . . . .	81
Quadro 12 – Conceitos da literatura não utilizados . . . . .	84
Quadro 13 – Definição das dimensões . . . . .	86
Quadro 14 – Dimensão Acesso com seus indicadores e variáveis . . . . .	89
Quadro 15 – Dimensão Formato com seus indicadores e variáveis . . . . .	90
Quadro 16 – Dimensão Direitos com seus indicadores e variáveis . . . . .	90
Quadro 17 – Dimensão Semântica com seus indicadores e variáveis . . . . .	91
Quadro 18 – Dimensão Estrutura com seus indicadores e variáveis . . . . .	91
Quadro 19 – Dimensão Apresentação com seus indicadores e variáveis . . . . .	92
Quadro 20 – Perfis de usuários e permissões . . . . .	105
Quadro 21 – Portais coletados para avaliação e análise . . . . .	108
Quadro 22 – Requisitos técnicos da lei brasileira e o mapeamento dos princípios da OGD . . . . .	172
Quadro 23 – O framework de benchmarking para avaliação de qualidade de portais de dados abertos . . . . .	174
Quadro 24 – Quadro de avaliação de envolvimento do usuário . . . . .	179
Quadro 25 – Dimensões e critérios MELODA . . . . .	184
Quadro 26 – Qualificação global baseada em resultados de avaliação . . . . .	187
Quadro 27 – Descrição das variáveis, indicador atualização . . . . .	190
Quadro 28 – Descrição das variáveis, indicador download em massa . . . . .	191
Quadro 29 – Descrição das variáveis, indicador Metadados . . . . .	193
Quadro 30 – Descrição das variáveis, indicador metadados descritivos . . . . .	194
Quadro 31 – Descrição das variáveis, indicador metadados estruturais . . . . .	197
Quadro 32 – Descrição das variáveis, indicador versionamento . . . . .	199
Quadro 33 – Descrição das variáveis, indicador URI . . . . .	200
Quadro 34 – Descrição das variáveis, indicador formato de dados . . . . .	202

Quadro 35 – Descrição das variáveis, indicador licenças . . . . .	204
Quadro 36 – Descrição das variáveis, indicador proveniência . . . . .	206
Quadro 37 – Descrição das variáveis, indicador vocabulário . . . . .	207
Quadro 38 – Descrição das variáveis, indicador pesquisa . . . . .	209
Quadro 39 – Descrição das variáveis, indicador conta usuário . . . . .	210
Quadro 40 – Descrição das variáveis, indicador recomendação . . . . .	211
Quadro 41 – Descrição das variáveis, indicador acessibilidade . . . . .	212
Quadro 42 – Descrição das variáveis, indicador multi-idioma . . . . .	212
Quadro 43 – Descrição das variáveis, indicador visualização de dados . . . . .	213
Quadro 44 – Descrição das variáveis, indicador temática . . . . .	214
Quadro 45 – Descrição das variáveis, indicador API   SPARQL . . . . .	215
Quadro 46 – Descrição das variáveis, indicador Interatividade . . . . .	217
Quadro 47 – Descrição das variáveis, indicador compreensão . . . . .	219
Quadro 48 – Descrição das variáveis, indicador quantitativos . . . . .	220
Quadro 49 – Questionário utilizado na pesquisa de campo . . . . .	261

## LISTA DE TABELAS

Tabela 1 – Resultado das busca nas quatro bases em detalhes . . . . .	37
Tabela 2 – Exemplificação do cálculo da avaliação . . . . .	98
Tabela 3 – Exemplificação dos cálculos de indicadores . . . . .	99
Tabela 4 – Exemplificação cálculo dimensões . . . . .	99
Tabela 5 – Aplicação modelo - Dimensão Acesso . . . . .	109
Tabela 6 – Aplicação modelo - Dimensão Formatos de Dados . . . . .	114
Tabela 7 – Aplicação modelo - Dimensão Direitos . . . . .	117
Tabela 8 – Aplicação modelo - Dimensão Semântica . . . . .	119
Tabela 9 – Aplicação modelo - Dimensão Estrutura . . . . .	122
Tabela 10 – Aplicação modelo - Dimensão Apresentação . . . . .	127
Tabela 11 – Aplicação modelo - Dimensões . . . . .	129
Tabela 12 – Quantitativo de conjuntos de dados adicionados, obtidos através das coletas realizadas no portal do governo federal . . . . .	139
Tabela 13 – Demais Portais Coletados . . . . .	141
Tabela 14 – Monitoramento Portal de Santa Catarina . . . . .	144
Tabela 15 – Variáveis Indicador Atualização . . . . .	223
Tabela 16 – Variáveis Indicador Download em Massa . . . . .	227
Tabela 17 – Variáveis Indicador Metadados . . . . .	228
Tabela 18 – Variáveis Indicador Metadados Descritivos . . . . .	230
Tabela 19 – Variáveis Indicador Metadados Estrutural . . . . .	233
Tabela 20 – Variáveis Indicador Versionamento . . . . .	235
Tabela 21 – Variáveis Indicador URI . . . . .	236
Tabela 22 – Variáveis Indicador Formatos de dados . . . . .	239
Tabela 23 – Variáveis Indicador Licenças . . . . .	241
Tabela 24 – Variáveis Indicador Proveniência . . . . .	243
Tabela 25 – Variáveis Indicador Vocabulários . . . . .	244
Tabela 26 – Variáveis Indicador Pesquisa . . . . .	245
Tabela 27 – Variáveis Indicador Conta Usuário . . . . .	246
Tabela 28 – Variáveis Indicador Recomendação . . . . .	247
Tabela 29 – Variáveis Indicador Acessibilidade . . . . .	248
Tabela 30 – Variáveis Indicador Multi-Idioma . . . . .	249
Tabela 31 – Variáveis Indicador Visualização dos Dados . . . . .	249
Tabela 32 – Variáveis Indicador Temática . . . . .	251
Tabela 33 – Variáveis Indicador API SPARQL . . . . .	255
Tabela 34 – Variáveis Indicador Interatividade . . . . .	256
Tabela 35 – Variáveis Indicador Compreensão . . . . .	258
Tabela 36 – Variáveis Indicador Quantitativos . . . . .	259

## LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
AHP	<i>Analytic Hierarchy Process</i>
APA	<i>American Psychological Association</i>
API	<i>Application Programming Interface</i>
CC	<i>Creative Commons</i>
CKAN	<i>Comprehensive Knowledge Archive Network</i>
CSV	<i>Comma Separated Values</i>
DC	<i>Dublin Core</i>
DCAT	<i>Data Catalog Vocabulary</i>
DSR	<i>Design Science Research</i>
DSRM	<i>Design Science Research Methodology</i>
EC	Engenharia do Conhecimento
EUA	Estados Unidos da América
FTP	<i>File Transfer Protocols</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IGTI	Núcleo de Estudos em Inteligência, Gestão e Tecnologia para Inovação
IHC	Interação Humano-Computador
INDA	Infraestrutura Nacional de Dados Abertos
INPI	Instituto Nacional da Propriedade Industrial
IRI	<i>Internacional Resource Identifier</i>
JSON	<i>Javascript Object Notation</i>
LAI	Lei de Acesso à Informação
LEC	Laboratório de Engenharia do Conhecimento
LED	Laboratório de Educação a Distância
LGPD	Lei Geral de Proteção de Dados
LOD	<i>Linked Open Data</i>
MELODA	Métricas para Liberar Dados Abertos
MLA	<i>Modern Language Association</i>
MRAPDA	Modelo de Referência de Avaliação de Portais dos Dados Abertos
MS	Microsoft
ODbL	<i>Open Database License</i>
ODI	<i>Open Data Institute</i>
ODM	<i>Open Data Monitor</i>

ODS	<i>Open Document Spreadsheets</i>
ODT	<i>Open Document</i>
OGD	<i>Open Government Data</i>
OKF	<i>Open Knowledge Foundation</i>
OWL	<i>Web Ontology Language</i>
PDF	<i>Portable Document Format</i>
POC	<i>Proof of Concept</i>
PPEGC	Programa de Pós- graduação em Engenharia e Gestão do Conhecimento
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
RIF	<i>Rule Interchange Format</i>
SiAPDA	Sistema de avaliação de portais de dados abertos
SKOS	<i>Simple Knowledge Organization System</i>
SPARQL	<i>Protocol and RDF Query Language</i>
SQL	<i>Structured Query Language</i>
SROI	<i>Social Return on Investmen</i>
TI	Tecnologia da Informação
TIC	Tecnologia de Informação e Comunicação
TICs	Tecnologias de Informação e Comunicação
UFSC	Universidade Federal de Santa Catarina
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web</i>
WCAG	<i>Web Content Accessibility Guidelines</i>
XML	<i>Extensible Markup Language</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>21</b>
1.1	CONTEXTUALIZAÇÃO	21
1.2	PROBLEMA DE PESQUISA	22
1.3	OBJETIVOS	24
<b>1.3.1</b>	<b>Objetivo Geral</b>	<b>24</b>
<b>1.3.2</b>	<b>Objetivos Específicos</b>	<b>24</b>
1.4	JUSTIFICATIVA	24
1.5	CONTRIBUIÇÕES E INEDITISMO	26
1.6	LIMITAÇÃO DE PESQUISA	26
1.7	ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO DE ENGENHARIA E GESTÃO DO CONHECIMENTO	28
1.8	METODOLOGIA DE PESQUISA	31
<b>1.8.1</b>	<b>Metodologia de apoio à construção de indicadores</b>	<b>32</b>
<b>1.8.2</b>	<b>Guia metodológico de investigação social - Corbetta</b>	<b>33</b>
<b>1.8.3</b>	<b>Desenvolvimento de pesquisa</b>	<b>34</b>
1.8.3.1	Identificação do problema e motivação	36
1.8.3.2	Definir objetivos para uma solução	36
1.8.3.3	Design e desenvolvimento	37
1.8.3.4	Demonstração	37
1.8.3.5	Avaliação	38
1.8.3.6	Comunicação	39
1.9	ESTRUTURA DO DOCUMENTO	39
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>40</b>
2.1	DADOS ABERTOS	40
2.2	BENEFÍCIOS NO USO DE DADOS ABERTOS	41
2.3	DADOS CONECTADOS	43
2.4	DADOS ABERTOS CONECTADOS	44
2.5	PADRÕES NA WEB	46
2.6	FORMATO DE DADOS E LICENÇA ABERTA	48
2.7	TECNOLOGIAS DA WEB SEMÂNTICA	51
2.8	BOAS PRÁTICAS DE DADOS NA WEB	54
2.9	PORTAIS DE DADOS ABERTOS	60
<b>2.9.1</b>	<b>Comprehensive Knowledge Archive Network - CKAN</b>	<b>60</b>
<b>2.9.2</b>	<b>Socrata</b>	<b>62</b>
<b>2.9.3</b>	<b>OpenDataSoft</b>	<b>63</b>
2.10	MODELOS DE AVALIAÇÃO DE PORTAIS DE DADOS ABERTOS	64
<b>2.10.1</b>	<b>Classificação dos estudos</b>	<b>64</b>

2.10.2	<b>Análise dos trabalhos relacionados</b> . . . . .	67
2.10.3	<b>Complemento aos trabalhos relacionados</b> . . . . .	69
2.11	CONSIDERAÇÕES FINAIS . . . . .	73
<b>3</b>	<b>O MODELO DE REFERÊNCIA PARA AVALIAÇÃO DE PORTAIS DE DADOS ABERTOS</b> . . . . .	<b>74</b>
3.1	ORGANIZAÇÃO DOS CONCEITOS DA LITERATURA . . . . .	74
3.2	ESTRUTURA DO MODELO . . . . .	80
<b>3.2.1</b>	<b>Indicadores</b> . . . . .	<b>81</b>
<b>3.2.2</b>	<b>Dimensões</b> . . . . .	<b>86</b>
<b>3.2.3</b>	<b>Variáveis</b> . . . . .	<b>88</b>
3.3	PROCESSO DE AVALIAÇÃO UTILIZANDO DO MODELO . . . . .	93
<b>3.3.1</b>	<b>Coletar</b> . . . . .	<b>93</b>
<b>3.3.2</b>	<b>Definir Padrões</b> . . . . .	<b>95</b>
<b>3.3.3</b>	<b>Efetuar Cálculos</b> . . . . .	<b>97</b>
<b>3.3.4</b>	<b>Analisar</b> . . . . .	<b>100</b>
3.4	CONSIDERAÇÕES FINAIS . . . . .	103
<b>4</b>	<b>PROVA DE CONCEITO</b> . . . . .	<b>104</b>
4.1	SISTEMA DESENVOLVIDO - SIAPDA . . . . .	104
4.2	APLICAÇÃO DO MODELO EM TRÊS CENÁRIOS DE USO . . . . .	108
<b>4.2.1</b>	<b>Análise dos resultados - Dimensão Acesso</b> . . . . .	<b>109</b>
<b>4.2.2</b>	<b>Análise dos resultados - Dimensão Formatos</b> . . . . .	<b>114</b>
<b>4.2.3</b>	<b>Análise dos resultados - Dimensão Direitos</b> . . . . .	<b>116</b>
<b>4.2.4</b>	<b>Análise dos resultados - Dimensão Semântica</b> . . . . .	<b>119</b>
<b>4.2.5</b>	<b>Análise dos resultados - Dimensão Estrutura</b> . . . . .	<b>121</b>
<b>4.2.6</b>	<b>Análise dos resultados - Dimensão Apresentação</b> . . . . .	<b>126</b>
<b>4.2.7</b>	<b>Análise dos resultados - Dimensões e Score Final</b> . . . . .	<b>129</b>
4.3	AVALIAÇÃO COM ESPECIALISTAS . . . . .	132
<b>4.3.1</b>	<b>Análise das Respostas às Perguntas do Critério 1</b> . . . . .	<b>133</b>
<b>4.3.2</b>	<b>Análise das Respostas às Perguntas do Critério 2</b> . . . . .	<b>136</b>
4.4	APLICAÇÃO DO MODELO EM DIVERSOS PORTAIS . . . . .	141
4.5	MONITORAMENTO DO PORTAL DE DADOS DO ESTADO DE SANTA CATARINA . . . . .	143
4.6	CONSIDERAÇÕES FINAIS . . . . .	144
<b>5</b>	<b>CONSIDERAÇÕES FINAIS DA PESQUISA</b> . . . . .	<b>146</b>
5.1	CONCLUSÕES . . . . .	146
5.2	TRABALHOS FUTUROS . . . . .	148
	<b>REFERÊNCIAS</b> . . . . .	<b>150</b>
	<b>APÊNDICE A – MODELOS DE AVALIAÇÃO DE PORTAIS DE DADOS ABERTOS</b> . . . . .	<b>172</b>

<b>A.0.1</b>	<b>Proposta de Andreiwid Sh Corrêa <i>et al.</i> (2017)</b>	<b>172</b>
<b>A.0.2</b>	<b>Proposta de Máchová e Lnénicka (2017)</b>	<b>173</b>
<b>A.0.3</b>	<b>Proposta de Sisto <i>et al.</i> (2018)</b>	<b>176</b>
<b>A.0.4</b>	<b>Proposta de Xiaohua Zhu e Freeman (2019)</b>	<b>178</b>
<b>A.0.5</b>	<b>Proposta de ODM (2019)</b>	<b>182</b>
<b>A.0.6</b>	<b>Proposta de Abella <i>et al.</i> (2019)</b>	<b>184</b>
	<b>APÊNDICE B – DEFINIÇÕES EM DETALHES DO MODELO DE REFERÊNCIA PROPOSTO</b>	<b>189</b>
<b>B.0.1</b>	<b>Dimensão Acesso</b>	<b>189</b>
B.0.1.1	Indicador Atualização e suas variáveis	189
B.0.1.2	Indicador Download em Massa e suas variáveis	191
B.0.1.3	Indicador Metadados e suas variáveis	192
B.0.1.4	Indicador Metadados Descritivos e suas variáveis	194
B.0.1.5	Indicador Metadados Estrutural e suas variáveis	196
B.0.1.6	Indicador Versionamento e suas variáveis	198
B.0.1.7	Indicador URI e suas variáveis	200
<b>B.0.2</b>	<b>Dimensão Formatos</b>	<b>201</b>
B.0.2.1	Indicador Formatos de dados e suas variáveis	201
<b>B.0.3</b>	<b>Dimensão Direitos</b>	<b>203</b>
B.0.3.1	Indicador Licença e suas variáveis	203
<b>B.0.4</b>	<b>Dimensão Semântica</b>	<b>205</b>
B.0.4.1	Indicador Proveniência e suas variáveis	205
B.0.4.2	Indicador Vocabulários e suas variáveis	206
<b>B.0.5</b>	<b>Dimensão Estrutura</b>	<b>209</b>
B.0.5.1	Indicador Pesquisa e suas variáveis	209
B.0.5.2	Indicador Conta Usuário e suas variáveis	210
B.0.5.3	Indicador Recomendação e suas variáveis	210
B.0.5.4	Indicador Acessibilidade e suas variáveis	211
B.0.5.5	Indicador Multi-Idioma e suas variáveis	212
B.0.5.6	Indicador Visualização dos dados e suas variáveis	213
B.0.5.7	Indicador Temática e suas variáveis	214
B.0.5.8	Indicador API SPARQL e suas variáveis	214
B.0.5.9	Indicador Interatividade e suas variáveis	215
<b>B.0.6</b>	<b>Dimensão Apresentação</b>	<b>218</b>
B.0.6.1	Indicador Compreensão e suas variáveis	218
B.0.6.2	Indicador Quantitativos e suas variáveis	219
	<b>APÊNDICE C – RESULTADO DETALHADO DOS PORTAIS ANALISADOS</b>	<b>223</b>
C.1	DIMENSÃO ACESSO	223

C.1.1	<b>Indicador Atualização</b>	223
C.1.2	<b>Indicador Download em Massa</b>	227
C.1.3	<b>Indicador Metadados</b>	227
C.1.4	<b>Indicador Metadados Descritivos</b>	229
C.1.5	<b>Indicador Metadados Estrutural</b>	233
C.1.6	<b>Indicador Versionamento</b>	234
C.1.7	<b>Indicador URI</b>	236
C.2	DIMENSÃO FORMATOS	238
C.2.1	<b>Indicador Formatos de dados</b>	238
C.3	DIMENSÃO DIREITOS	240
C.3.1	<b>Indicador Licenças</b>	240
C.4	DIMENSÃO SEMÂNTICA	242
C.4.1	<b>Indicador Proveniência</b>	242
C.4.2	<b>Indicador Vocabulários</b>	243
C.5	DIMENSÃO ESTRUTURA	245
C.5.1	<b>Indicador Pesquisa</b>	245
C.5.2	<b>Indicador Conta Usuário</b>	246
C.5.3	<b>Indicador Recomendações</b>	247
C.5.4	<b>Indicador Acessibilidade</b>	248
C.5.5	<b>Indicador Multi-Idioma</b>	248
C.5.6	<b>Indicador Visualização dos dados</b>	249
C.5.7	<b>Indicador Temática</b>	250
C.5.8	<b>Indicador API SPARQL</b>	255
C.5.9	<b>Indicador Interatividade</b>	255
C.6	DIMENSÃO APRESENTAÇÃO	257
C.6.1	<b>Indicador Compreensão</b>	257
C.6.2	<b>Indicador Quantitativos</b>	258
	<b>APÊNDICE D – QUESTIONÁRIO APLICADO NO GRUPO DE PROFISSIONAIS DA ÁREA DE DADOS ABERTOS</b>	261

## 1 INTRODUÇÃO

Neste capítulo será apresentada a contextualização, evidenciando benefícios de dados abertos para a sociedade e para as organizações que disponibilizam dados, bem como alguns dos benefícios que são obtidos com a difusão de dados abertos.

Na sequência, apresenta-se o problema de pesquisa delimitado pelo contexto dos portais de dados abertos, considerando dados disponibilizados ao longo de toda existência do portal. Também destaca-se a pergunta de pesquisa, os objetivos (geral e específicos), a justificativa da realização desta pesquisa, as contribuições e ineditismo. Salientam-se ainda as limitações, a aderência ao Programa de Pós-graduação em Engenharia e Gestão do Conhecimento (PPEGC) e a metodologia utilizada para a realização desta tese.

### 1.1 CONTEXTUALIZAÇÃO

Dados tem grande relevância para a sociedade, em destaque para a sociedade do conhecimento. Neste sentido, dados abertos é visto como um movimento para compartilhamento de conhecimento, beneficiando a sociedade e as organizações que realizam a abertura de seus dados (BAUER; KALTENBÖCK, 2011). Este movimento ganhou ênfase, visto que as organizações aumentam a responsabilidade, a transparência, a até mesmo a eficiência, através da disponibilização dos dados, além de incentivar o uso e o reuso de dados e informações para diversos propósitos (ATTARD *et al.*, 2015; KAPOOR *et al.*, 2015).

Neste sentido, dados abertos precisam atender alguns critérios, dentre estes destacam-se dois: **i)** ser tecnicamente abertos, ou seja, fazerem uso de formatos abertos que possam, ser recuperados, lidos e preferencialmente processados, desta forma viabilizando o seu uso; **ii)** ser legalmente aberto, ou seja fazer uso de licença aberta, permitindo o seu uso gratuito, desta forma explicitando uma garantia legal sobre o uso dos dados (ARMAN *et al.*, 2015; DIN *et al.*, 2015).

Observando os critérios para dados abertos, destaca-se que estes podem beneficiar diversas áreas, tais como: educação, transporte, eletricidade, petróleo, saúde, e finanças (MANYIKA *et al.*, 2013; MACHADO *et al.*, 2019). Assim como enfatiza-se que o uso de dados abertos tem um grande valor econômico potencial, pois melhora a transparência, aborda e amplia um grande volume de dados para os mais diversificados setores, permitindo a experimentação e possibilitando oportunidades de negócios (MANYIKA *et al.*, 2013; LÓSCIO, B. F. *et al.*, 2016; MACEDO; LEMOS, 2021).

Este aumento na disponibilidade de dados é em decorrência dos diferentes benefícios que dados abertos podem proporcionar, destacando a inovação e geração de receitas, fazendo os governos perceberem seus dados como um recurso de relevância estratégica (BATES, 2014; CAVANILLAS *et al.*, 2016) ou mesmo por decorrência das

leis que impõem aos governos a abertura e disponibilização dos dados (NEUMAIER, 2015). Por conta deste cenário, nos últimos anos, os governos se destacaram na criação e na disponibilização de dados abertos (MÁCHOVÁ; LNÉNICKA, 2017).

As Tecnologias de Informação e Comunicação (TICs) e a Web proporcionaram o desenvolvimento e a propagação de dados por meio das plataformas digitais, facilitando o seu acesso e possibilitando a participação ativa da sociedade (ROTTA, 2018). Notoriamente o interesse em usufruir dos benefícios dos dados, motivou a necessidade de se ter políticas e padrões para disponibilizar dados de maneira adequada, para potencializar seu uso e reuso. Desta forma, diversas iniciativas se propõem a avaliar e estimular a conformidade da estrutura e dos dados disponibilizados pelas organizações (SILVA *et al.*, 2020), com o intuito de impulsionar dados abertos.

Portanto, percebe-se a relevância da disponibilização de dados abertos e os esforços que existem para alavancar este tema. Contudo, existem questões que ainda dificultam o uso, o acesso, a localização e a exploração dos dados. Neste âmbito, a seção a seguir aborda alguns destes aspectos.

## 1.2 PROBLEMA DE PESQUISA

Desde o início do movimento de dados abertos em 2009 pelo governo dos Estados Unidos da América (EUA), percebe-se uma crescente adesão por diversas organizações públicas e privadas, refletindo em um grande volume de dados disponibilizados. Algumas dessas organizações efetuam a publicação de dados com o único objetivo de atender a uma legislação, isso revela um problema, pois os dados ali publicados muitas vezes não são atrativos aos usuários que visam utilizá-los (MACHADO *et al.*, 2018). Acarretando em que a maioria dos portais de dados abertos torna-se um simples repositório de dados, perdendo diversas características de dados abertos (LOURENÇO, 2015).

Pesquisas realizadas apontam alguns problemas em relação aos portais de dados abertos, como exemplo: metadados inconsistentes, baixa qualidade dos conjuntos de dados, baixa padronização dos dados, falta de gerenciamento e manutenção, dados não legíveis por máquinas, dificuldade na visualização dos dados, falta de alinhamento com o interesse da sociedade, o pouco tempo para o tratamento dos dados a serem disponibilizados, são alguns problemas apontados pela literatura (LOURENÇO, 2015; SAYOGO *et al.*, 2014; ARQUERO AVILES; MARCO CUENCA, 2014; KUBLER *et al.*, 2016; MÁCHOVÁ *et al.*, 2018; SISTO *et al.*, 2018). Neste sentido, é possível perceber que estes problemas são reflexo da ausência de padrões, sendo estes responsáveis por proporcionar o mínimo de conformidade nos dados e possibilitar seu uso pela sociedade.

Além disso, os órgãos públicos enfrentam algumas dificuldades ao trabalhar com os dados abertos, tais como, a dispersão dos dados, a falta de confiança dos

usuários em relação aos dados e até mesmo a dificuldade técnica e tecnológica das organizações em publicar dados (SALM JÚNIOR, 2012).

Desta forma, percebe-se que existem boas práticas, guias metodológicos, recomendações, entre outros, acerca de padrões, processos e infraestrutura para dados abertos. No entanto, se faz necessário compreender e utilizar de maneira adequada os padrões e recomendações para que se tenha sucesso com a abertura dos dados e eficiência no uso, ou seja, o uso de padrões viabilizam o reuso, a compreensão, facilidade de processamento, a facilidade de descoberta, a confiança, a capacidade de conexão, facilidade de acesso e a interoperabilidade (LÓSCIO, B. F. *et al.*, 2017).

Percebe-se que vários estudos foram desenvolvidos com base em avaliar como os padrões e as recomendações estão sendo aplicados acerca de dados abertos, mas em sua maioria realizam a coleta e avaliação de maneira manual, devido a sua natureza. No entanto, observa-se que este tipo de avaliação pode acarretar em erros ou inconsistências por falha humana. Desta forma, sendo uma questão emergente de pesquisa (SILVA *et al.*, 2020).

Destaca-se ainda que ao avaliar portais de dados abertos como um todo, de modo manual, podem ser ocasionados ruídos, como: respostas incoerentes em decorrência do cansaço, interpretações incorreta, dentre outros fatores, por reflexo do grande volume de dados que um portal pode ter, visto que podem haver dezenas ou milhares de dados.

Neste sentido, sugeriram iniciativas de organizações governamentais e não governamentais que disponibilizam uma série de informações e relatórios que se propõem a monitorar a disponibilização de dados, em destaque cita-se: Open Data Barometer<sup>1</sup> ; Open Data Monitor<sup>2</sup> ; e Global Open Data Index<sup>3</sup>. No entanto, nenhuma destas iniciativas tem como foco avaliar e analisar em detalhes os portais de dados abertos, considerando as boas práticas de dados e as recomendações para o portal de dados abertos. Além de não proporcionarem a avaliação utilizando de pesos na estrutura do modelo, o que garante maior flexibilidade e adequação aos diferentes cenários.

Deste modo, a maneira em que são realizadas as avaliações sobre os portais de dados podem ser aprimoradas, ampliadas e combinadas através de um modelo de referência que proporcione avaliar a conformidade dos padrões dos dados disponibilizados e potencializar o seu uso e reuso, dentre outros benefícios que podem ser alcançados. Proporcionando uma avaliação independente do nível de governo ou país, além de possibilitar a automatização do processo de coleta e avaliação, bem como a disponibilização de um índice geral.

Dessa forma, a questão que essa pesquisa propõe responder é: **Como avaliar**

<sup>1</sup> Disponível em: [https://opendatabarometer.org/?\\_year=2017&indicator=0DB](https://opendatabarometer.org/?_year=2017&indicator=0DB)

<sup>2</sup> Disponível em: <https://www.opendatamonitor.eu/frontend/web/index.php?r=dashboard%2Findex>

<sup>3</sup> Disponível em: <https://index.okfn.org/>

## **de forma automatizada a conformidade de portais de dados abertos considerando os diferentes cenários?**

Para isso, tem-se o objetivo geral e os objetivos específicos destaca-dos abaixo.

### 1.3 OBJETIVOS

#### 1.3.1 Objetivo Geral

Propor um modelo de referência para avaliação de portais de dados abertos que sintetize as boas práticas e conceitos de avaliações existentes.

#### 1.3.2 Objetivos Específicos

- Identificar os conceitos relacionados a boas práticas de dados na Web;
- Identificar os conceitos oriundos da literatura para avaliação de portais de dados abertos;
- Compor um modelo de referência baseado nos conceitos e boas práticas identificados;
- Desenvolver um sistema computacional de apoio a aplicação do modelo e avaliação de portais de dados abertos;
- Realizar a prova de conceito do modelo em diferentes cenários de uso;
- Avaliar junto a especialistas a adequação do modelo de referência e o sistema de apoio para a avaliação de portais de dados abertos.

### 1.4 JUSTIFICATIVA

A publicação de dados abertos em diferentes frentes traz vários benefícios, dentre estes, a geração de conhecimento (COLPAERT *et al.*, 2013; BOGDANOVIĆ-DINIĆ *et al.*, 2014) e geração de inovação, que se reflete em novos negócios e, por consequência, um impacto positivo da economia (LNENICKA, 2015; EUROPEAN DATA PORTAL, 2018).

No entanto, para que a sociedade possa fazer uso dos dados disponibilizados nos portais ou mesmo ter uma participação ativa, é necessário que tanto os dados quanto os portais sigam os padrões e recomendações apresentadas por estudos acadêmicos, bem como pela *World Wide Web* (W3C).

Além disso, o *Open Data Leaders Network*, coordenado pelo *Open Data Institute* (ODI) apresentou 12 recomendações para os responsáveis sobre as publicações

de dados abertos, para que estes, realizem uma gestão de mudança cultural no governo (CEWEB, 2016), sendo destacada o item 12, correspondente a avaliação de dados abertos:

- [...]12-]Crie métricas para a avaliação regular das atividades de dados abertos para medir o progresso, comparar com padrões de sucesso e identificar áreas para melhorias.

Há autores que destacam desafios relacionados ao acesso e ao uso dos dados abertos, sendo que estes devem ser considerados na avaliação de portais. Dentre estes, destacam-se:

- **Desafios técnicos:** questões e desafios relacionados à tecnologia, dados e formatos ou infraestrutura necessária para a publicação (KUCERA *et al.*, 2015).
- **Desafios de infraestrutura dos dados:** A infraestrutura e o gerenciamento de dados se mostram deficientes e inconsistentes devido a alterações frequentes nos conjuntos de dados (MACEDO; LEMOS, 2021).
- **Desafios de acesso aos dados e usabilidade:** A interface dos portais afeta a capacidade de localização, consumo e interação. Os portais geralmente possuem uma abordagem técnica e orientada a desenvolvedores, não contemplando os usuários finais e cidadãos comuns (MACEDO; LEMOS, 2021).
- **Desafios de publicação:** Os conjuntos de dados são geralmente incompletos, sem informações adicionais, com falta de padrão e muitas vezes sem metadados significativos, evidenciando problemas de interoperabilidade e em fontes de referência. As publicações de forma geral não utilizam todo o potencial da Web de Dados, essenciais para conectar conjunto de dados heterogêneos e descobrir recursos relacionados (MACEDO; LEMOS, 2021).
- **Desafios legais e de privacidade:** A natureza dos dados abertos impõe novas demandas para prover privacidade, segurança e impedir o uso indevido das informações. Devem existir políticas sobre quais dados devem ser compartilhados e suas circunstâncias adequadas (MACEDO; LEMOS, 2021).

Várias pesquisas multidisciplinares emergiram sobre dados abertos ao longo do tempo (BARRIO *et al.*, 2016; HOUNSELL *et al.*, 2016; MEDJROUBI *et al.*, 2017). Contudo como pode ser observado existem lacunas a serem preenchidas, dentre estas, estão os aspectos tecnológicos e de conformidade, sendo estes necessários para fomentar o uso dos dados abertos na sociedade (ENRIÉQUEZ-REYES *et al.*, 2021). Além da necessidade do desenvolvimento de ferramentas que possibilitem a automatização do processo de coleta e avaliação das atividades de dados abertos (CORREA *et al.*, 2019).

Desta forma, a realização deste trabalho se justifica na construção de um modelo abrangente que considere aspectos relacionados aos conjuntos de dados, aos recursos, metadados e a estrutura do portal. Este modelo viabilizará a coleta e avaliação automatizada dos portais de dados abertos, independente do seu contexto e do nível de governo.

## 1.5 CONTRIBUIÇÕES E INEDITISMO

Percebe-se o esforço da comunidade acadêmica para avaliar os portais a fim de atingir a excelência de dados abertos, contudo, salienta-se que os modelos existentes não conseguem contribuir efetivamente neste sentido. Visto que portais de dados são ambientes que possuem milhares de dados, dificultado a coleta e análise total dos dados disponibilizados.

Destaca-se que a realização deste estudo torna-se original e inédita devido a ser abordado diferentes dimensões, indicadores e variáveis, sendo um estudo apoiado por um sistema computacional, que viabiliza a coleta, a análise e o monitoramento de diferentes portais, independente do volume de dados que constam no portal.

Isso se dá, devido ao modelo de referência ser abrangente e por isso permite o desdobramento de novas estruturas com o intuito de obter avaliações pontuais sobre diferentes cenários e aplicações independentes do tamanho ou da esfera do portal de dados abertos. Além da formalização do processo de avaliação, que possibilita criar padrões que viabilizam diferentes perspectivas de avaliação, bem como a disponibilização do score final sobre o portal avaliado, contribuindo com a continuidade de pesquisas na área e com o próprio portal.

Ainda, com o intuito de auxiliar na aplicação do modelo de referência e viabilizar a análise dos resultados, foi desenvolvido um sistema computacional para a coleta e avaliação dos portais de dados abertos, o que permite apresentar os resultados de maneira interativa, facilitando a sua compreensão. Assim possibilitando a automatização do processo de coleta e avaliação de portais de dados abertos.

Assim observa-se que o modelo de referência poderá ser utilizado como base para o desenvolvimento de novos modelos, bem como servirá para impulsionar o uso dos dados em conformidade com recomendações e para gerar evidências que possibilitem o aprimoramento da abertura dos dados e dos portais.

## 1.6 LIMITAÇÃO DE PESQUISA

Para a realização desta pesquisa e proposição do modelo de referência foram considerados estudos no âmbito de dados abertos que se propõem à avaliar os portais de dados abertos. As boas práticas sobre dados na Web e o uso da literatura também auxiliaram na composição do modelo proposto.

Referente ao tema de dados abertos, será abordado o contexto de dados abertos conectados, tendo em vista que, ambos compõem parte do universo de dados na Web (LÓSCIO, B. *et al.*, 2016) e dão suporte a questões semânticas.

Na composição do modelo foram considerados aspectos relacionados aos conjuntos de dados e seus recursos, bem como questões da estrutura e da apresentação do portal de dados. Contudo, aspectos relacionados ao conteúdo dos recursos (os dados) não foram considerados. Outro ponto considerado na composição do modelo foi o emprego de uma metodologia para a definição das dimensões, indicadores e suas variáveis.

Quanto à prova de conceito, foi aplicada em cenários de uso real através de portais de dados abertos que se remetem aos diferentes níveis do governo brasileiro (Federal, Estadual e Municipal). É ressaltado que os dados coletados são dados públicos e disponibilizados abertamente para a sociedade. E também destaca-se que o modelo apresentado pode ser utilizado em outros cenários desde que façam o uso da plataforma *Comprehensive Knowledge Archive Network* (CKAN) e as suas extensões.

Nesta pesquisa serão recorrentemente serão utilizados alguns conceitos, sendo estes:

- **Modelo de Referência:** Para definir modelo de referência primeiramente é necessário definir modelo, sendo este uma representação de parte da realidade, como vista pelas pessoas que desejam fazer uso deste para entender, mudar, gerenciar e controlar parte da realidade de uma forma ou outra (PIDD, 1999).  
Usando deste pressuposto, o modelo de referência tem por objetivo proporcionar uma solução inicial dos processos de negócio da empresa, para que este possa ser especificado e detalhado, sendo um modelo específico e particular. Estes modelos podem ser especializados para algum mercado, segmento ou tipologias específicas, como exemplo, setor automotivo, tecnológico, dentre outros (BREMER; LENZA, 2000).
- **Conformidade:** Do inglês *Compliance* é um termo amplamente utilizado no âmbito organizacional, jurídico e político (SANTOS *et al.*, 2012; GÓIS, 2016; COELHO, 2017) referindo-se a um conjunto de disciplinas que objetivam-se em cumprir e executar as normas legais e regulamentares, as políticas e as diretrizes estabelecidas para negócios e atividades institucionais ou empresariais, possibilitando atuar contra desvios ou inconformidades que possam acontecer (GÓIS, 2016). No âmbito desta tese, conformidade não será considerada no contexto legal, mas de padrões e boas práticas aplicados a dados abertos.
- **Dimensão:** É a divisão dos componentes que englobam o significado do conceito e que irão representar as dimensões de análise (CORBETTA, 2003).

- **Indicador:** Segundo Van Bellen (2005) é a maneira de categorizar e quantificar informações de modo que seu significado seja de fácil compreensão, simplificando as informações sobre fenômenos complexos.
- **Variáveis:** É responsável pela mensuração na prática do indicador (CORBETTA, 2003).
- **Conjuntos de dados (*dataset*):** É uma parcela de dados - por exemplo, podem ser as estatísticas de crimes de uma região, os números de gastos de um departamento do governo ou as leituras de temperatura de várias estações meteorológicas (CKAN, 2019). E este pode conter diversos recursos.
- **Recursos:** São os responsáveis por manter os dados em si. Um recurso pode ser uma planilha CSV ou Excel, arquivo XML, documento PDF, arquivo de imagem, dados vinculados no formato RDF etc. Os recursos são abordados pelo W3C como distribuição (LÓSCIO, B. F. *et al.*, 2017).
- **Dados:** Menor unidade armazenada em um determinado recurso. Encontrada em grandes quantidades nos portais de dados.

## 1.7 ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO DE ENGENHARIA E GESTÃO DO CONHECIMENTO

Esta tese esta contextualizada na área de concentração de Engenharia do Conhecimento (EC) do PPEGC da Universidade Federal de Santa Catarina (UFSC). Estando na linha de pesquisa Engenharia do Conhecimento aplicada às organizações (EGC, 2021). Esta linha estuda a concepção, desenvolvimento e implantação de soluções da Engenharia do Conhecimento nas organizações públicas e privadas.

A EC é amparada pela visão epistemológica do cognitivismo, sendo que esta visão surgiu com o início dos computadores e da inteligência artificial. Portanto, o cognitivismo percebe que o processo de adquirir o conhecimento pode ser entendido por métodos científicos, podendo ser processado por máquina, ou seja, existe a percepção que o conhecimento pode ser processado por computadores, armazenados em banco de dados, arquivos, manuais ou rotinas, assim possibilitando o fácil compartilhamento em uma organização (PACHECO, 2016).

Ao considerar todos os dados, as informações e o potencial conhecimento que estão presentes nos portais de dados abertos, a EC contribui com métodos e técnicas para armazenar, instanciar e disseminar as informações obtidas através dos portais de dados abertos, possibilitando potencializar e promover os avanços da área para as diferentes organizações públicas e privadas.

Neste sentido, foram realizadas buscas no banco de tese e dissertações do PPEGC <sup>4</sup>, com o objetivo de encontrar trabalhos similares a esta tese e que evidenciem contribuições dos trabalhos existentes, além de manter e evidenciar a memória do PPEGC.

Para a apresentação dos estudos similares a esta tese serão considerados nos quadros subsequentes os cinco primeiros trabalhos recuperados no banco de teses e dissertação do PPEGC, sendo apresentadas pesquisas que não se repetem nas diferentes buscas realizadas.

Em um primeiro momento foram buscados trabalhos que abordam o tema dados abertos e dados abertos conectados, sendo encontrados doze trabalhos. O Quadro 1 apresenta alguns dos trabalhos.

Quadro 1 – Trabalhos que abordam discussões sobre dados abertos e dados abertos conectados

<b>Referências</b>	<b>Título</b>	<b>Área de Concentração</b>	<b>T/D</b>
(CASAES, 2019)	Governança de Dados Abertos Governamentais: Framework Conceitual para as Universidades Federais, Baseado em uma Visão Sistêmica.	Engenharia do Conhecimento	Tese
(CAMPOS, 2018)	Dados Abertos Governamentais: Desafios na Publicação	Engenharia do Conhecimento	Dissertação
(PEREIRA, 2017)	OGDPub: Uma ontologia para Publicação de Dados Abertos Governamentais	Engenharia do Conhecimento	Dissertação
(GOMES, M. S., 2017)	Proposta de Arquitetura para Ecosystema de Inovação em Dados Abertos.	Engenharia do Conhecimento	Dissertação
(MOURA SPERONI, 2016)	Modelo de Referência Para Indicadores de Inovação Regional Suportado por Dados Ligados	Engenharia do Conhecimento	Tese

Fonte: do Autor

Por sua vez, realizou-se uma busca sobre trabalhos que abordam estudos que fazem uso de indicadores e foram encontrados quarenta e uma teses e dissertações. O Quadro 2 apresenta alguns dos trabalhos .

<sup>4</sup> <http://btd.egc.ufsc.br>

Quadro 2 – Trabalhos que abordam discussões sobre indicadores

<b>Referências</b>	<b>Título</b>	<b>Área de Concentração</b>	<b>T/D</b>
(CUNHA, 2019)	Rankings e Indicadores para Smart Cities: uma proposta de cidades inteligentes autopoieticas	Gestão do Conhecimento	Dissertação
(ALMEIDA NASCIMENTO, 2019)	Acessibilidade e Usabilidade em Plataformas MOOC: Indicadores de Boas Práticas	Mídia do Conhecimento	Dissertação
(JUNIOR, 2016)	Ontologia para Representação Semântica de Indicadores de Desempenho Considerando Aspectos de Vaguidade, Temporalidade e Relacionamento entre Indicadores	Engenharia do Conhecimento	Tese
(SILVA, D. B. L. da, 2016)	Indicadores para Avaliação da Influência dos Ambientes de Empreendedorismo Inovador na Geração de Capital Social	Gestão do Conhecimento	Tese
(SILVA, A. W. L., 2014)	Governança de Sistemas de Indicadores de Sustentabilidade em Processos de Avaliação Ambiental Estratégica sob Mediação da Gestão do Conhecimento	Gestão do Conhecimento	Tese

Fonte: do Autor

Por fim, realizou-se uma última busca sobre estudos que abordam o contexto de portais e foram encontrados dez trabalhos. O Quadro 3 apresenta alguns dos trabalhos recuperados.

Quadro 3 – Trabalhos que abordam discussões sobre portais

<b>Referências</b>	<b>Título</b>	<b>Área de Concentração</b>	<b>T/D</b>
(SANTOS ALBUQUERQUE, 2020)	Modelo de Apoio a Gestão da Qualidade da Informação em Portais do Poder Judiciário Utilizando Fuzzy Quality Function Deployment	Engenharia do Conhecimento	Tese
(MEZZARROBA, 2019)	Framework para Avaliação de Portais do Poder Judiciário Brasileiro a Partir de Mecanismos de Gestão do Conhecimento.	Engenharia do Conhecimento	Tese

(MEZZAROBA, 2015)	Requisitos para a Avaliação de Portais de Governo Eletrônico do Poder Judiciário a partir das Resoluções de Metas do CNJ.	Engenharia do Conhecimento	Dissertação
(ALMEIDA, 2017)	O Compartilhamento do Conhecimento em Portais de E-Participação Legislativa	Engenharia do Conhecimento	Dissertação
(BIZ, 2009)	Avaliação dos Portais Turísticos Governamentais Quanto ao Suporte a Gestão do Conhecimento	Engenharia do Conhecimento	Tese

Fonte: do Autor

Deste modo, estes estudos prévios demonstram a aderência do presente trabalho ao objeto de pesquisa do Programa de Pós-graduação em Engenharia e Gestão do Conhecimento.

Deste modo, nota-se que há oportunidades de pesquisa neste meio visando avaliar os portais de dados abertos através de um modelo de referência. Este trabalho de tese diferencia-se dos demais trabalhos que foram desenvolvidos no PPEGC, por convergir aspectos acadêmicos teóricos que serão oriundos da pesquisa na literatura e aspectos técnicos oriundos das boas praticas de dados na Web recomendadas pelo W3C. Salienta-se ainda que o modelo de referência construído possibilita a automatização dos processos de coleta e avaliação, além da disponibilização de um sistema computacional que permite instanciar o modelo de referência.

Com isso, ressalta-se a importância da relação entre os temas escolhidos e a contribuição com os avanços de pesquisas no programa e nas áreas correlatas aos temas.

## 1.8 METODOLOGIA DE PESQUISA

Nesta seção apresentam-se as etapas utilizadas na realização desta pesquisa de tese. Um procedimento metodológico em qualquer trabalho de pesquisa é importante, pois possibilita apresentar com clareza e de modo sistêmico os procedimentos, técnicas e métodos para a elaboração de conhecimento científico confiável (DE HAES *et al.*, 2013).

Dentre as diferentes metodologias existente na literatura, destaca-se a *Design Science Research Methodology* (DSRM), visto que, possui uma abordagem diferenciada em relação aos métodos de pesquisa tradicionais por possibilitar a criação de artefatos e geralmente com algum propósito prático (DRESCH *et al.*, 2015). Outro ponto relevante sobre a DSRM é o fato de ser classificada como tecnológica, pois,

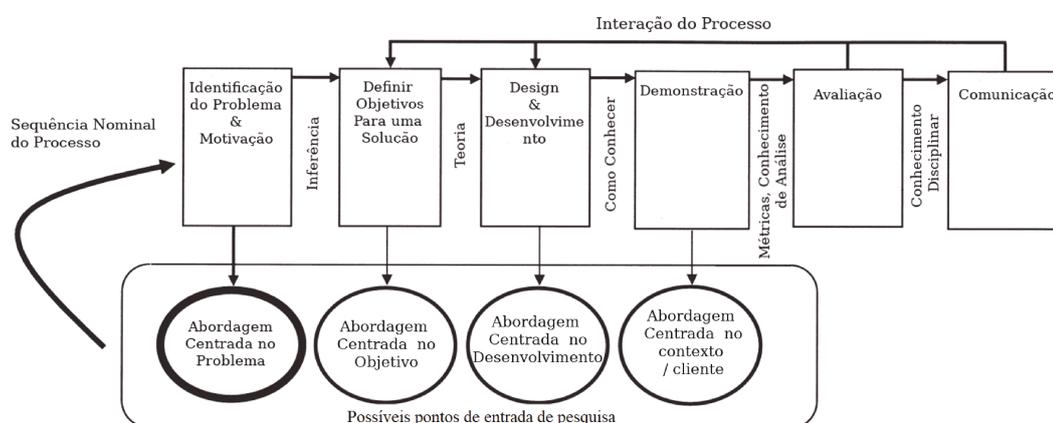
seu objetivo está na solução de problemas específicos e pontuais, tendo como foco o artefato a ser desenvolvido (JUNIOR *et al.*, 2017).

Na literatura é possível ver a relevância e a evolução das pesquisas sobre a DSRM, no trabalho apresentado por Le Moigne (1995), o autor apresenta um quadro dos principais autores que contribuíram com a evolução da *Design Science Research* (DSR) até chegar na DSRM, sendo estes: Herbert Alexander Simon, March e Smith, dentre outros.

Naturalmente, com a evolução das teorias acerca da DSRM, foram sendo aprimorados os métodos para auxiliar no desenvolvimento de pesquisas tecnológicas (VON ALAN *et al.*, 2004; JAPPUR, 2014; WIERINGA, 2014; DRESCH *et al.*, 2015).

Dentre as diversas ramificações da DSRM, o método proposto por Peffers *et al.* (2007) se destaca por possuir quatro tipos de entradas para o processo de pesquisa e por não exigir uma entrada preferencial, podendo iniciar o processo em qualquer uma das quatro entradas que antecedem a etapa de avaliação, como pode ser observado na Figura 1.

Figura 1 – Etapas da Metodologia *Design Science Research*



Fonte: (PEFFERS *et al.*, 2007)

Além da metodologia DSRM, esta pesquisa utilizou-se de uma metodologia de apoio à construção de indicadores. A próxima seção irá apresentar a metodologia utilizada, desta forma, foi possível compor a estrutura do modelo.

### 1.8.1 Metodologia de apoio à construção de indicadores

Os indicadores são amplamente utilizados por diversas áreas possibilitando, por exemplo, monitorar ou avaliar a condição de uma rede de computadores, a fim de identificar falhas, interrupções ou lentidão. Isso possibilita aos administradores de rede tomarem ações de reparos preventivos ou corretivos. Torna-se ainda mais relevante na medida que o tempo passa, pois com os registros anteriores "histórico" é possível

detectar padrões, conhecer o ambiente, fortalecer os processos, parcerias e dentre outros (FIEP, SESI-PR, 2010).

O principal objetivo de utilizar-se dos indicadores é categorizar e quantificar informações de modo que seu significado seja de fácil compreensão, simplificando as informações sobre fenômenos complexos (VAN BELLEN, 2005). As informações obtidas pelos indicadores ajudam aos analistas, administradores e demais interessados a se orientar no processo de tomada de decisão, auxiliando em tarefas eficientes, eficazes e específicas (FIEP, SESI-PR, 2010).

Na literatura existem algumas metodologias que apoiam a construção de indicadores e dentre estas foi optada pela metodologia apresentada por Corbetta (2003), pois esta permite uma flexibilidade na construção da estrutura de indicadores e a possibilidade da criação de um índice(score final), bem como por possuir uma etapa que fundamenta a análise da viabilidade de mensuração na prática dos indicadores, no caso específico deste trabalho foi considerada a mensuração através de um sistema computacional.

### **1.8.2 Guia metodológico de investigação social - Corbetta**

No guia metodológico de investigação social proposto por Corbetta (2003), ao se utilizar de indicadores, torna-se um conceito de alto nível mais específico e simples. Isso se reflete em uma tradução e em uma compreensão mais clara do conceito em termos empíricos. Esta afirmação se resume, portanto, a em modificar os conceitos gerais para conceitos específicos, criando um elo entre si através de uma afinidade por significado.

É importante ressaltar que a relação entre indicadores e conceito é apenas parcial. Contudo, um conceito(geral) suporta mais de um indicador(específico), por outro lado, um indicador pode ser utilizado para especificar mais de um conceito (CORBETTA, 2003).

Outro ponto destacado por Corbetta (2003) é a arbitrariedade da escolha de um indicador, pois o indicador representa apenas parte de um conceito, podendo haver vários indicadores para o mesmo conceito. Portanto, o pesquisador vai eleger os indicadores que irá utilizar em sua pesquisa conforme seus critérios.

Na metodologia proposta por Corbetta (2003) a criação de indicadores é dividida em quatro etapas, sendo estas descritas a seguir.

- 1) **Divisão do conceito em dimensões:** Esta etapa tem por objetivo a reflexão teórica sobre os componentes que englobam o significado do conceito e que irão representar as dimensões de análise;
- 2) **A escolha de indicadores:** Esta etapa tem por objetivo a identificação dos indicadores de cada dimensão. Mesmo que os indicadores sejam apenas conceitos,

é possível iniciar o planejamento do problema de observação;

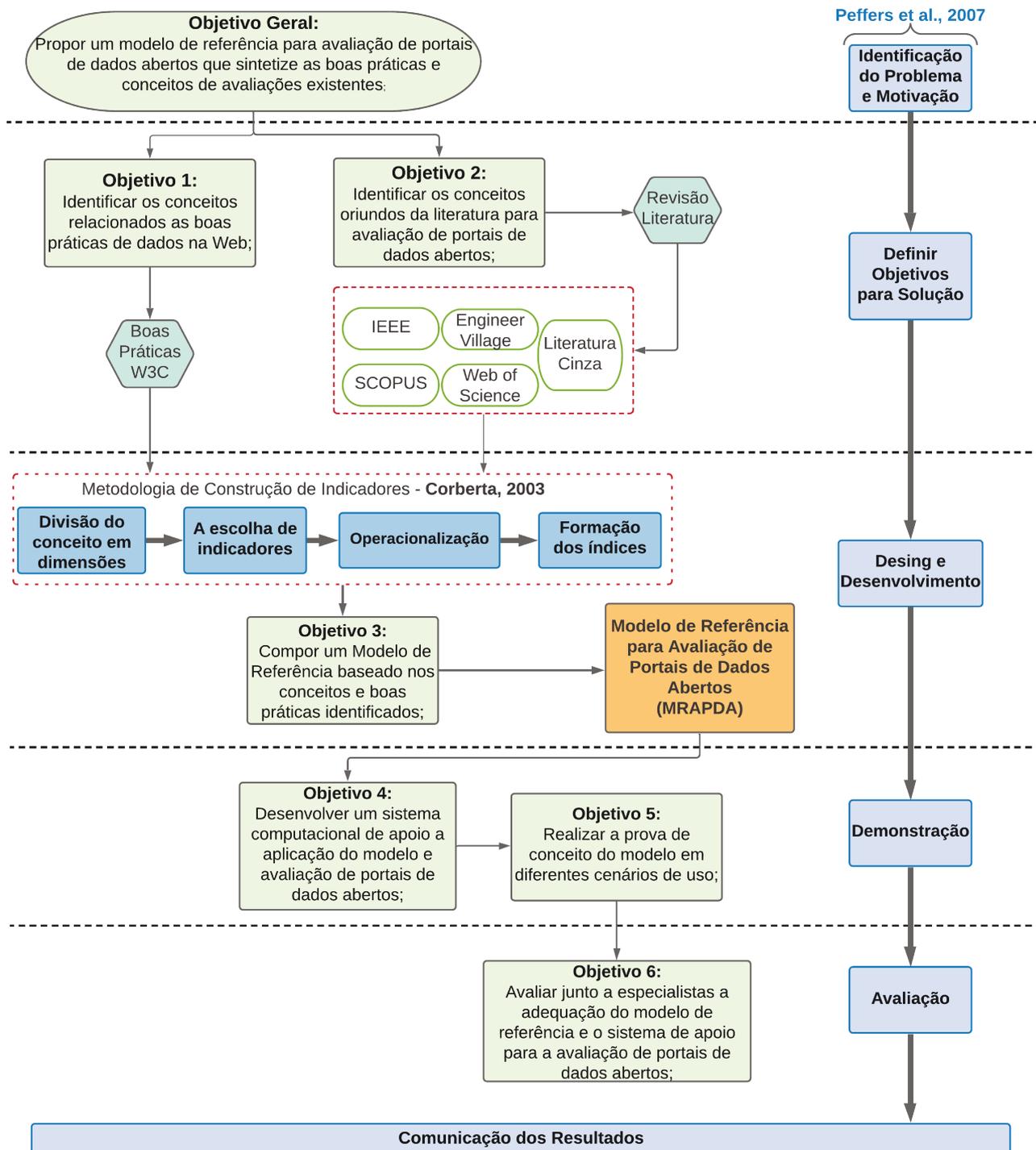
- 3) **Operacionalização:** Esta etapa tem por objetivo transformar os indicadores em variáveis, ou seja, como será mensurado na prática os indicadores;
- 4) **Formação dos índices:** Esta etapa tem por objetivo sintetizar múltiplas variáveis em um índice geral;

Segundo Corbetta (2003) o processo de tradução (construção dos indicadores) de um conceito de operações empíricas que é realizado através das quatro fases apresentadas previamente nem sempre são executadas em sequência. O pesquisador, quando não usar de evidências próprias, mas sim de evidências já existentes, não pode planejar a definição dos indicadores segundo sua análise conceitual. Isso ocorre, pois, o pesquisador irá somente utilizar de evidências já existentes oriundas da recompilação de outros pesquisadores.

### 1.8.3 Desenvolvimento de pesquisa

Considerando a importância do emprego de uma metodologia para pesquisa, este trabalho de tese foi amparado pela metodologia de Peffers *et al.* (2007). Neste sentido, visando a reprodutibilidade do estudo, apresenta-se a seguir cada etapa da metodologia DSRM descrevendo os estágios realizados na condução desta pesquisa. A figura 2 apresenta a relação das etapas da DSRM com a pesquisa.

Figura 2 – Etapas da DSRM x Estágios da Pesquisa



Fonte: do Autor

### 1.8.3.1 Identificação do problema e motivação

Durante o processo de pesquisa sobre o contexto de dados abertos e discussões junto aos integrantes do laboratório de pesquisa, observou-se a falta de interesse dos cidadãos em utilizar de dados abertos, abrindo margem para algumas indagações, porque os dados abertos não são amplamente utilizados pelos cidadãos? O que falta para despertar o interesse da sociedade em utilizar e gerar valor econômico através dos dados publicados? O dados que estão sendo publicados estão seguindo padrões? Os dados publicados estão disponíveis? Neste sentido, foram analisados trabalhos sobre dados abertos, assim como os portais de dados e percebeu-se a falta de padrões, além de algumas inconsistências, visto as recomendações e padrões existentes na literatura.

Desta forma, iniciou-se a estruturação do objetivo geral deste estudo, contudo, para refinar e evidenciar o problema, foram realizadas buscas nas bases de dados onde foi possível observar uma lacuna de pesquisa que envolve essas indagações.

### 1.8.3.2 Definir objetivos para uma solução

Com o intuito de definir os objetivos e propor a solução, foram realizadas buscas nas seguintes bases de dados: IEEE Xplore , *Web of science*, SCOPUS e *Engenning Village*, utilizando-se de três sintaxes de busca e considerando os últimos cinco anos (2014-2019):

- "*Open Data*"AND "*Portal*"\*AND "*Assessment*";
- "*Open Data*"AND "*Portal*"\*;
- *Open Data*"AND "*Portal*"\*AND *Monitoring*";

Como pode ser observado nas sintaxes, a presença do termo *Open Data* que se remete a Dados abertos e *Portal* que remete-se a portais web, juntamente com outros três diferentes termos, sendo, *Assessment* e *Evaluation* que correspondem a avaliação e *Monitoring* correspondente a monitoramento.

Optou-se por fazer uso dos termos que se remetem a avaliação e monitoramento, pois, acredita-se que um artefato que identifique as deficiências relacionada aos dados possa contribuir para a realização de ações a fim de melhorar o processo de publicação, contribuir com a conformidade dos dados e por consequência trazer bons resultados para a sociedade.

Como consequência desta busca, obteve-se um total em relação as quatro bases de 204 documentos(Recuperados) e com o auxílio do software EndNote<sup>5</sup> foi possível identificar 35 documentos duplicados(Dup.), chegando à um total de 169 documentos(T.Bases) correspondente as quatro bases utilizadas.

<sup>5</sup> Mais informações em <<https://endnote.com/>>

Tabela 1 – Resultado das busca nas quatro bases em detalhes

Bases	Recuperados	Dup.	T. Base	Dup. T.	S/Aderência	T. Analisar
IEEE	41	2	169	69	73	27
Engennier Village	42	14				
SCOPUS	88	11				
Web of Science	33	8				

Fonte: do Autor

Por sua vez, verificou-se pela segunda vez a presença de documentos duplicados, no total das bases correspondentes a soma de todas as bases(T.Bases), foram encontrados assim 69 documentos duplicados(DupT.). Por fim, efetuou-se uma análise dos 100 documentos restantes, levando em conta o título e o resumo, resultando em 73 documentos sem aderência à pesquisa(S/aderência). Com isso, restaram 27 documentos(T. Analisar) para serem analisado na integra. A seguir a Tabela 1 apresenta em detalhes os documentos recuperados das bases de dados.

Após a análise dos 27 documentos, observou-se que apenas quatro possuem relação direta com esta pesquisa. No entanto, a fim de complementar o resultado foi efetuada uma nova busca utilizando da literatura cinza (*grey literature*) e através desta busca foi possível encontrar mais dois trabalhos relacionados. Na seção 2.10 são apresentados os seis trabalhos, bem como uma classificação dos documentos recuperados.

Nesta etapa foi realizada a identificação dos trabalhos encontrados na literatura, sendo realizada a organização dos conceitos encontrados nos trabalhos selecionados e nas boas práticas, sendo elaborado o agrupamento destes, por conta das diferentes nomenclaturas encontradas.

### 1.8.3.3 Design e desenvolvimento

Desta maneira, foi composto o modelo seguindo a metodologia de construção de indicadores de (CORBETTA, 2003), sendo definido os indicadores e suas variáveis e por fim a definição das dimensões. Durante a composição das variáveis foi verificada a viabilidade técnica para possibilitar a implementação de um sistema computacional ou mesmo a criação de outro tipo de ferramenta tecnológica. Na seção 3.1 é apresentado em detalhes como foi feita a composição do modelo proposto.

### 1.8.3.4 Demonstração

A etapa de demonstração do modelo se deu por meio do desenvolvimento de um sistema computacional que auxilia na aplicação do modelo. Foram escolhidos três

cenários para a aplicação do modelo, sendo o portal de dados abertos do governo federal, do estado do Rio Grande do Sul e da município de São Paulo. Estes portais foram escolhidos por possuírem um grande volume de dados e por representar os três níveis de governos (federal, estadual e municipal). Contudo, é importante destacar que o sistema desenvolvido aplicou o modelo em 25 portais de dados abertos<sup>6</sup> escolhidos aleatoriamente.

Desta forma, foi possível visualizar o funcionamento do processo de coleta e avaliação automática utilizando do modelo, mais detalhes sobre a aplicação do modelo esta na seção 4.2.

Os resultados relacionados à aplicação do modelo estão disponíveis em: <https://muriologomes.studiokem.ufsc.br/acesso>, sendo apenas necessário o cadastro de usuário no sistema para visualizar os resultados.

#### 1.8.3.5 Avaliação

A avaliação deste trabalho é dada por intermédio de um questionário com o intuito de avaliar a estrutura do modelo e o sistema computacional. Foram avaliados por um grupo de especialistas que atuam no âmbito de dados abertos, sendo estes: um professor e um administrador, ambos com grau de instrução de doutor; um analista de TI, mestre; e um auditor fiscal com especialização em gestão pública.

Foram convidados 30 especialistas, contudo somente quatro responderam aos questionamentos. É importante destacar a dificuldade com relação à comunicação com os especialistas, devido a sobrecarga que existe por conta da pandemia do Covid-19 e a realização de trabalho remoto. Acredita-se que em decorrência disso não obteve-se um numero maior de participação.

Para auxiliar na avaliação foi gravado um vídeo<sup>7</sup> expondo o contexto e a importância em avaliar dados abertos, a estrutura do modelo e as etapas de avaliação. Também foi gravado um segundo vídeo<sup>8</sup> que apresenta o sistema computacional desenvolvido, além do envio de material extra de apoio.

O questionário (Apêndice D) foi elaborado com objetivo de avaliar a adequação do modelo em relação às boas práticas apresentadas pela literatura, bem como, avaliar o software desenvolvido para a aplicação do modelo. O questionário é composto por dois critérios de análise, sendo **I-** Em relação ao Modelo de Referência e **II-** Em Relação ao Sistema de avaliação de portais de dados abertos (SiAPDA). Além de conter um campo descritivo para sugestões, críticas ou elogios e até mesmo compreender a visão do profissional sobre a sua avaliação. Para mais detalhes ver seção 4.3.

<sup>6</sup> Encontrados em: <https://datashades.info> - sendo este um Web site que informa os portais em nível global que utilizam do CKAN

<sup>7</sup> Link vídeo apresentação modelo: <[https://youtu.be/Y\\_kYkZWxFJc](https://youtu.be/Y_kYkZWxFJc)>

<sup>8</sup> Link vídeo sistema Web: <<https://youtu.be/YbmR-MeOVTM>>

### 1.8.3.6 Comunicação

A comunicação desta pesquisa é destinada à comunidade científica e sociedade civil com a apresentação dos resultados. Serão escritos e enviados artigos à periódicos e congressos internacionais relacionados a área de estudo.

Foi realizado o registro do sistema computacional desenvolvido e a disponibilização de um Web site <sup>9</sup> onde é apresenta toda a pesquisa, bem como os trabalhos realizados ao longo do doutorado.

Também foi divulgado o estudo realizado a um grupo de especialistas na área de dados abertos em uma reunião *on-line* no ano de 2021, sendo que este grupo esta desenvolvendo uma proposta de similar ao estudo elaborado. Nesta reunião fizeram-se presentes membros da W3C Brasil, Open Knowledge Brasil, Ceweb.br, NIC.br, entre outros.

## 1.9 ESTRUTURA DO DOCUMENTO

Este trabalho possui uma estrutura composta de cinco capítulos, sendo, o **primeiro capítulo** composto pela introdução que objetiva-se em apresentar a visão geral sobre a pesquisa, seus objetivos, a justificativa, as limitações, a aderência ao programa de pós-graduação EGC e a metodologia utilizada.

Por sua vez, o **capítulo dois** apresenta a fundamentação teórica, sendo estes relacionados a dados abertos, os benefícios do uso de dados abertos, dados abertos conectados, os padrões empregados, as boas práticas recomendadas pela W3C sobre dados na Web, os tipos de plataformas utilizadas em portais de dados abertos e os modelos de avaliação encontrados na literatura que auxiliaram na construção do modelo.

O **capítulo três** relata o modelo de referência, a organização dos conceitos que o compõem, a definição de sua estrutura e do processo de avaliação utilizado.

O **capítulo quatro** aborda a prova de conceito evidenciando a arquitetura e os módulos que compõem o sistema computacional, a explicitação dos resultados oriundos da avaliação automatizada do sistema e expõe a opinião dos especialistas sobre o trabalho desenvolvido.

Por fim, no **capítulo cinco** encontram-se as considerações finais, as contribuições teóricas e práticas e as sugestões para trabalhos futuros.

---

<sup>9</sup> <https://murilogomes.studiokem.ufsc.br>

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo em evidenciar os conceitos que sustentam esta tese. São abordadas questões acerca de dados abertos, dados conectados, dados abertos conectados, os padrões que amparam estas abordagens, as plataformas comumente utilizadas que dão suporte aos portais de dados, além das boas práticas de dados na Web. Busca-se também apresentar e discutir os modelos de avaliação de portais de dados abertos encontrados na literatura com o intuito de identificar os conceitos que possam auxiliar na construção do modelo de referência.

### 2.1 DADOS ABERTOS

Diversas mudanças tecnológicas ocorreram nas últimas décadas, dentre estas o crescimento da Web e o massivo uso de Tecnologia de Informação e Comunicação (TIC) promoveu a criação exponencial de dados. Todo este crescimento é reflexo de oferta e procura, criando um ciclo virtuoso sobre disponibilização e uso de dados na Web (CAI; ZHU, Y., 2015; ÁVILA, 2017).

Por sua vez o movimento de dados abertos veem crescendo e tornando-se evidente perante às organizações públicas, privadas e aos cidadãos. Dados abertos são reflexo de outro movimento existente a tempos, denominado de *openness*, que é definido por Schlagwein *et al.* (2017) como:

Terminologia *openness* pode se referir a um conceito de ordem superior (por exemplo, a "filosofia de abertura"); a natureza dos recursos (por exemplo, "dados abertos"); a natureza dos processos (por exemplo, "inovação aberta"); ou os efeitos em domínios específicos (por exemplo, "educação aberta") [...] Os princípios normalmente usados para caracterizar este conceito de ordem superior são: acesso à informação e outros recursos; participação de forma inclusiva e frequentemente colaborativa; transparência de recursos e ações; e democracia ou "democratização", como a quebra de estruturas de exclusão.

Contudo, os dados abertos ganharam ênfase em 2009, quando o governo americano começou a abertura de seus dados com o propósito de obter maior transparência, colaboração e participação da sociedade. Além dos dados publicados por órgãos públicos também existem publicações por parte de empresas privadas. Todo esse movimento de dados abertos se reflete em oportunidades de novos serviços e benefícios à sociedade (KUBLER *et al.*, 2018).

A fim de conceituar dados abertos, este foi definido como: "dados que qualquer um pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, exigências que visem preservar sua proveniência e sua abertura" (OPEN DEFINITION, 2021).

Desde o surgimento dos dados abertos, entidades e indivíduos se uniram para criar recomendações sobre o uso de dados abertos, promover sua publicação, o seu

uso e reuso. Em 2007 um grupo da Califórnia definiu oito princípios de dados abertos (DADOS ABERTOS, 2018), sendo estes:

- **Completos:** Todos os dados públicos devem ser disponibilizados. Dado público é aquele que não está sujeito a restrições de privacidade, segurança ou outros privilégios.
- **Primários:** São apresentados tal como colhidos da fonte, com o maior nível possível de granularidade, sem agregação ou modificação.
- **Atuais:** Devem ser publicados o mais rápido possível para preservar seu valor. Em geral, têm periodicidade: quanto mais recentes e atuais, mais úteis.
- **Acessíveis:** São disponibilizados para a maior quantidade possível de pessoas, atendendo, assim, aos mais diferentes propósitos.
- **Compreensíveis por máquina:** Devem estar estruturados de modo razoável, possibilitando que sejam processados automaticamente.
- **Não discriminatórios:** Devem estar disponíveis para qualquer pessoa, sem necessidade de cadastro ou qualquer outro procedimento que impeça o acesso.
- **Não proprietários:** Nenhuma entidade ou organização deve ter controle exclusivo sobre os dados disponibilizados.
- **Livres de licenças:** Não devem estar submetidos a *copyrights*, patentes, marcas registradas ou regulações de segredo industrial. Restrições razoáveis quanto a privacidade, segurança e outros privilégios são aceitas, desde que transparentes e bem justificadas.

Os dados abertos vem ganhando destaque com o passar do tempo. As instituições, ao publicarem dados, contribuem significativamente com a sociedade, refletindo no impacto direto na economia, em aspectos sociais e no desempenho de setores públicos. A seção a seguir apresenta os benefícios no uso de dados abertos.

## 2.2 BENEFÍCIOS NO USO DE DADOS ABERTOS

A disponibilização de dados abertos traz uma série de benefícios a todos os setores da sociedade (INFORMATION SOCIETY *et al.*, 2015). Os principais favorecidos são os órgãos públicos, refletindo na qualidade das equipes, de serviços prestados e no conhecimento sobre gargalos e obstáculos internos.

A ciência e as pesquisas se beneficiam através de dados abertos, isso porque, o grande volume de dados disponível sobre as mais diversas áreas, tais como: saúde, clima, geográficos e dentre outros, faz com que os dados sejam um grande atrativo à pesquisa. O setor privado também pode obter benefícios ao utilizar dados abertos, isso ocorre porque permite aos empreendedores a criação e idealização de novas tecnologias e novos negócios (PIRES, 2019).

Um estudo divulgado pelo portal europeu de dados abertos<sup>1</sup>, apresenta benefícios no uso de dados, sendo estes ligados à três pilares, **i) Desempenho:** que reflete diretamente na eficiência dos serviços públicos, **ii) Economia:** que remete-se ao acesso dos dados, informações, conteúdos e ao conhecimentos, isso contribui com

<sup>1</sup> Disponível em: <https://www.europeandataportal.eu/pt/using-data/benefits-of-open-data>

a geração de serviços e tecnologias inovadoras e **iii) Social:** que fomenta a colaboração, participação e a inovação social, em virtude das informações mais transparentes e acessíveis (INFORMATION SOCIETY *et al.*, 2015). A Figura 3 exibe os três pilares e seus benefícios.

Figura 3 – Vantagens da utilização dos dados



Fonte: (EUROPEAN DP, 2019), tradução do autor

Há iniciativas que incentivam o uso de dados abertos, exemplos destas são: os *hackathons* ou apenas o dia dos dados abertos (KITSIOS; KAMARIOTOU, 2018). Estes eventos são promovidos por instituições públicas ou privadas de diferentes proporções, como é o caso do *Open Data Day*. Este evento, ocorre em uma escala global, envolvendo diversos países, sempre tendo como foco algumas áreas, como exemplo, a edição de 2018 que teve como temas: pesquisa acadêmica, rastreamento de dinheiro público, meio ambiente e dados abertos para o direito (ODD, 2019).

As iniciativas que incentivam o uso de dados abertos chamam a atenção de várias áreas e setores, sendo que estas obtêm valor através do uso de dados abertos, sendo algumas áreas apresentadas no Quadro 4.

Quadro 4 – Áreas e Setores que pesquisam sobre dados abertos

Áreas/Setor	Referência
Agricultura	(JAROLIÉMEK; MARTINEC, 2016; CHARVAT <i>et al.</i> , 2014)
Automobilística	(KEGELMAN <i>et al.</i> , 2017)
Ciências Naturais	(MASUZZO; MARTENS <i>et al.</i> , 2015; BERNASCONI <i>et al.</i> , 2017; WIENER <i>et al.</i> , 2016)
Energia	(KORFIATI <i>et al.</i> , 2016; MEDJROUBI <i>et al.</i> , 2017)
GeoEspaciais/GeoDados	(CUCA, 2016; HARRIS; BAUMANN, 2015; BENITEZ-PAEZ <i>et al.</i> , 2018)
Negócios	(ALBINO, 2017; HUSEIN <i>et al.</i> , 2015; ROSSI <i>et al.</i> , 2007; SUSHA <i>et al.</i> , 2015a)

Saúde	(OWOLABI <i>et al.</i> , 2015; HAO <i>et al.</i> , 2017; DEMSKI <i>et al.</i> , 2016)
Turismo	(MCNAUGHTON <i>et al.</i> , 2016; CAO <i>et al.</i> , 2011; HOUNSELL <i>et al.</i> , 2016; DE VOCHT <i>et al.</i> , 2016)
Tecnológica	(PINHEIRO <i>et al.</i> , 2016; ROCCA <i>et al.</i> , 2016; ASAT <i>et al.</i> , 2017; BARRIO <i>et al.</i> , 2016; DERGUECH <i>et al.</i> , 2014; THOMA <i>et al.</i> , 2014)
Portuario	(CHEN <i>et al.</i> , 2014; KAZEMI <i>et al.</i> , 2013)

Fonte: do Autor

Contudo, não existem somente os dados abertos que geram e trazem benefícios. Existem também os dados conectados e dados abertos conectados, sendo estes, voltados a difusão da Web semântica ou comumente conhecida como Web de dados. Na próxima seção serão apresentados e contextualizados os dados conectados e dados abertos conectados.

### 2.3 DADOS CONECTADOS

A Web Semântica, também conhecida como Web de Dados, é uma extensão da Web atual, também denominada de Web 3.0 (SANTOS; NICOLAU, 2012). A Web atual tem foco em documentos voltados ao uso dos seres humanos e a Web Semântica incluiu dados e informações de modo que softwares inteligentes possam raciocinar sobre os dados e informações (SHADBOLT *et al.*, 2006; PEDRINACI; DOMINGUE *et al.*, 2010).

Os dados conectados são definidos como: "um conjunto de boas práticas para conectar e publicar dados estruturados na Web de diferentes fontes" (BERNERS-LEE, 2011), com o objetivo de torna-los legíveis por máquinas. Há quatro recomendações básicas para a publicação e o consumo de dados conectados, segundo Berners-Lee (2011), sendo estas:

- I) Use *Uniform Resource Identifier* (URI) como nome para as coisas;
- II) Use *Hypertext Transfer Protocol* (HTTP) URI para que as pessoas possam procurar esses nomes;
- III) Quando alguém procura um URI, fornecer informações úteis, usando de padrões (*Resource Description Framework* (RDF) ou *Protocol and RDF Query Language* (SPARQL));
- IV) Incluir links para outros URIs. Para que se possa descobrir mais informações

Por sua vez, os dados conectados também proporcionam alguns benefícios, sendo estes apresentados por Auer *et al.* (2012).

- **Uniformidade:** Com a utilização do modelo de dados RDF toda a informação publicada é representada por termos consistindo por sujeito, predicado e objeto. Deste modo todos os dados são baseados em um modelo uniforme.
- **Capacidade de ser referenciado:** Utiliza-se de URIs para identificar entidades e permite a localização, recuperação de recursos descritos e representados na Web por estas entidades assim como uma *Uniform Resource Locator* (URL).
- **Coerência:** As triplas RDF estabelecem uma ligação entre as entidades identificadas pelo sujeito e o objeto sendo interligadas por um predicado que representa a ligação entre eles.
- **Integrabilidade:** Todas as fontes de dados conectados são baseadas em um único mecanismo para representar a informação, sendo de fácil integração semântica por diferentes conjuntos de dados.
- **Pontualidade:** Todas as publicações e atualizações dos dados tendem a reutilização, facilitando a disponibilização.

Para suportar dados conectados em um formato aberto, em 2007 foi criado o projeto de dados abertos conectados. Na próxima seção, será apresentado este contexto.

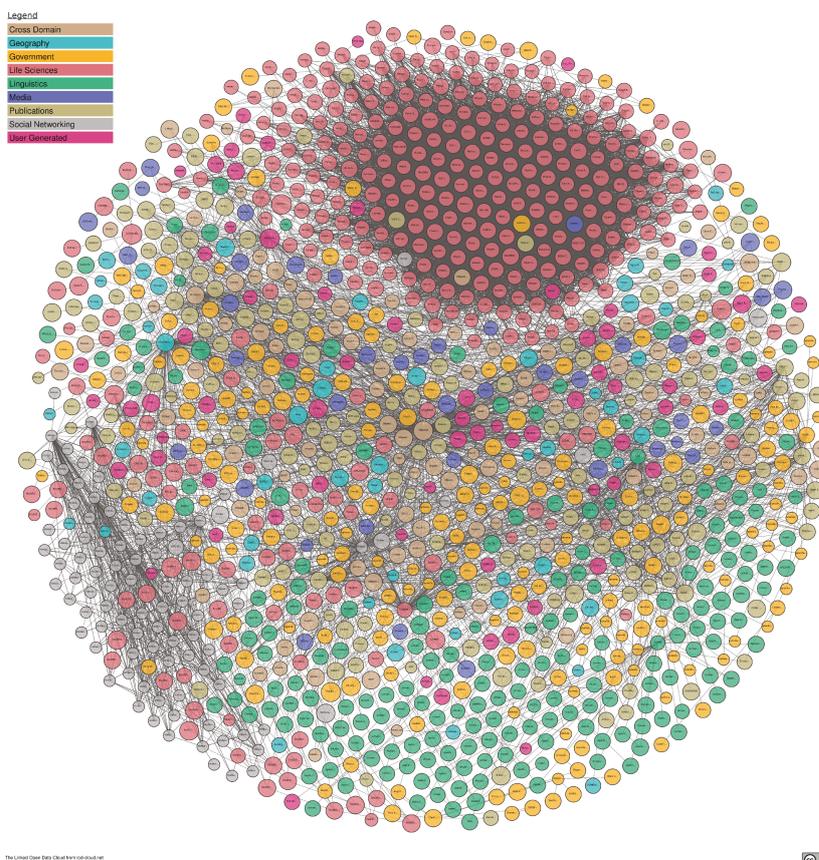
## 2.4 DADOS ABERTOS CONECTADOS

Em 2007 Tim Berners-Lee iniciou o projeto de dados abertos conectados (*Linked Open Data (LOD)*), com objetivo de construir a Web de dados utilizando dos conjuntos de dados (*datasets*). Utilizando-se de licença aberta e transformando-os em RDF, com auxílio de vocabulários para seu uso e entendimento, mas levando em consideração todas as recomendações de dados conectados para sua publicação (BIZER *et al.*, 2011).

Os dados conectados e dados abertos conectados utilizam tecnologias pertencentes à arquitetura da Web Semântica (RDF, URI, Ontologias, SPARQL e entre outras) (W3C, 2019b). O uso dessas tecnologias proporcionam uma padronização dos dados, facilitando o processamento e a conexão dos conjuntos de dados. Deste modo, viabiliza a associação entre os conjuntos de dados, formando assim a Web de dados (ALSHEHHI *et al.*, 2013). A Figura 4 apresenta o estado da Web de dados no ano de 2021.

Como pode ser observado na Figura 4, existe um grande volume de conjuntos de dados, dividido em nove grupos (domínio cruzado, geografia, governo, ciência da vida,

Figura 4 – Nuvem LOD em 2021

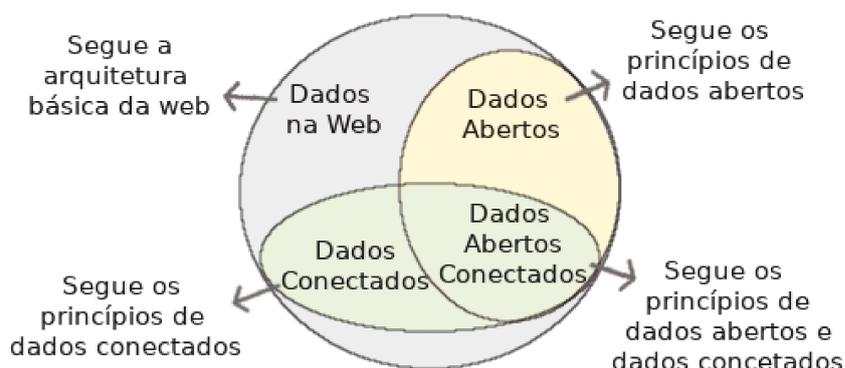


Fonte: (MCCRAE; ABELE, 2021)

linguística, mídia de comunicação, publicações, redes sociais e gerado por usuários). A nuvem LOD começou a ser construída em 2007 junto ao lançamento do projeto, na época havia cerca de 12 conjuntos de dados publicados, em 2010 havia cerca de 203 conjuntos de dados e atualmente há cerca de 1301 conjuntos de dados.

Pode-se afirmar que a Web é um universo de dados, por sua vez, dados abertos e dados abertos conectados também estão englobados neste contexto em um mesmo grupo, diferenciando-os apenas pelo contexto semântico. Contudo, dados conectados também fazem parte da Web, mas seguindo apenas seus princípios sem a necessidade de aderir as definições de dados abertos. O estudo realizado por Bernadette Lóscio *et al.* (2016) define bem esta afirmação, sendo que a Figura 5 representa este contexto.

Figura 5 – Dados na Web Vs Dados Abertos Vs Dados Conectados



Fonte: (LÓSCIO, B. *et al.*, 2016), tradução do autor

No entanto, para utilizar-se destes conceitos, se faz necessário seguir alguns padrões. A seção seguinte aborda alguns dos padrões acerca de dados, como por exemplo: formatos, licença, URI, dentre outros.

## 2.5 PADRÕES NA WEB

O uso de padrões está presente em várias áreas, auxiliando a organizar e igualar propósitos para que seja atingindo um objetivo comum. Existem diversas entidades responsáveis por padronizar os diferentes segmentos, contudo, destaca-se o W3C, o *Creative Commons* (CC) e o *Open Knowledge Foundation* (OKF) estas são algumas das organizações que estão a frente em propor e criar padrões para a Web, dados abertos, licenças e dentre outras abordagens tecnológicas.

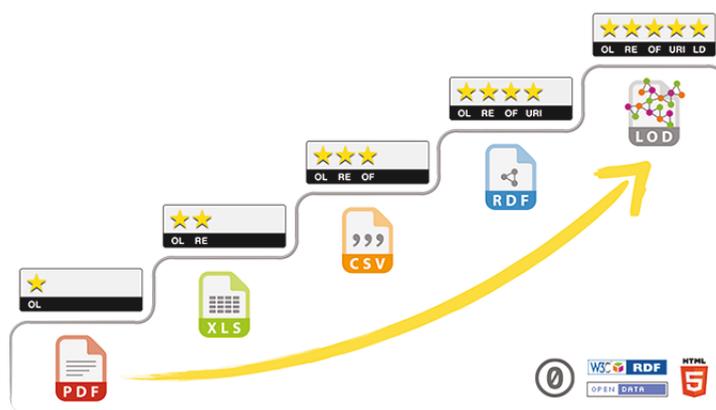
O W3C é a organização que propõe padrões sobre a Web, que consiste em um consórcio internacional integrando empresas, órgãos governamentais e organizações independentes, sendo fundada por Tim Berners-Lee em 1994 (W3C, 2021). Esta organização é responsável por algumas recomendações no âmbito de dados na Web, dados conectados, dados abertos conectados e questões que englobam a Web semântica, por exemplo (W3C, 2019a).

A CC é uma organização não governamental sem fins lucrativos, que tem por objetivo ampliar a quantidade de obras criativas disponíveis por meio de seus diferentes tipos de licenças, que permite a cópia e o compartilhamento com menos restrições do que o tradicional *Copyright* © (CC, 2021).

A OKF também é uma organização sem fins lucrativos que tem como objetivo fomentar a abertura de todas as informações não pessoais, possibilitando a utilização, desenvolvimento e compartilhamento por qualquer pessoa. Esta organização está presente em 40 países, incluindo o Brasil, onde propõe em analisar e promover a transparência, participação social, dados abertos, além de capacitar a sociedade civil (OKF, 2021)

Um das principais iniciativas que visa agregar padronização aos dados é a classificação de cinco estrelas proposta por Tim Berners-Lee (BERNERS-LEE, 2010), sendo uma das mais conhecidas e utilizadas. No entanto, essa classificação leva em conta apenas os critérios de ser legível por máquina, licença aberta e o formato dos dados disponibilizados. A Figura 6 apresenta esta classificação.

Figura 6 – Classificação com base em cinco estrelas



Fonte: (BERNERS-LEE, 2010)

Neste esquema de classificação de cinco estrelas, temos o **I**- primeira estrela onde objetiva-se em disponibilizar na Web dados de qualquer formato desde que faça uso de licença aberta (ver Quadro 5 e 6) para que seja considerado um dado aberto; **II**- segunda estrela refere-se a disponibilizar dados estruturados legíveis por máquinas (por exemplo Microsoft (MS) Excel <sup>2</sup>) em vez de digitalizar uma tabela, por exemplo; **III**- terceira estrela disponibilizar dados em formato de máquina estruturado, contudo, utilizando de formato não proprietário (por exemplo *Comma Separated Values* (CSV)); **IV**- quarta estrela possibilita disponibilizar os dados em si, possibilitando vincula-los com outros dados na Web. Desta forma, torna-se possível a construção da Web de dados (dados conectados), mas é importante salientar o uso dos padrões abertos recomendados pela W3C (RDF e SPARQL); e **V**- quinta estrela proporciona vincular os dados disponibilizados no quarto nível com outros dados na Web permitindo fornecer o contexto destes dados (dados conectados abertos).

Vale ressaltar que a escala de estrelas são acumulativas, ou seja, se está no nível três significa que os dados disponibilizados possuem **(I)**licença aberta, **(II)**fazem uso de um formato estruturado legível por máquina e **(III)**fazem uso de um formato não proprietário.

<sup>2</sup> Editor de planilha eletrônica da Microsoft

## 2.6 FORMATO DE DADOS E LICENÇA ABERTA

Uma das recomendações mais citadas é a atribuição da licença aberta, esta viabiliza o uso legal dos dados. No entanto, também é importante fazer uso de formatos de dados adequados. Porém, salienta-se que o formato não pode possuir qualquer tipo de restrição, como exemplo: monetário e de utilização, espera-se que no mínimo o formato permita o processamento por ao menos uma ferramenta de software livre ou gratuita (OPEN DEFINITION, 2018).

Existem diversos formatos de dados, sendo estes considerados abertos ou proprietários (fechados). Contudo sobre a ótica de dados abertos, deve-se utilizar preferencialmente de formatos abertos e preferivelmente formatos abertos legíveis por máquinas (ZHU, X.; FREEMAN, 2019).

É importante salientar que **formatos abertos** são aqueles que não necessitam de uma ferramenta de software específica para acesso aos dados, ou seja, está disponível a qualquer pessoa e livre de cobrança. Por sua vez, **formato proprietário** é aquele que necessita a compra da ferramenta de software ou o seu acesso é limitado (OKF, 2020).

Alguns dos formatos abertos comumente utilizados, PDF<sup>3</sup>, ODT, ODS, TXT e dentre outros. Alguns dados em formato aberto legível por máquina, que são apresentados de forma estruturada, como por exemplo, *CSV*, *Javascript Object Notation (JSON)*, *Extensible Markup Language (XML)* e *RDF* (OPENDATA, 2019; OKF, 2020).

Uma das preocupação ao fazer uso de dados abertos é a definição de uma licença. Para isso, a OKF através do projeto Open Definition (2021), apresenta as condições necessárias e aceitáveis para atribuição da licença aberta:

### Condições necessárias:

- **Uso:** A licença deve permitir a livre utilização e usufruto da obra licenciada;
- **Redistribuição:** O licença deve permitir a redistribuição da obra licenciada, incluindo a venda, seja por conta própria seja como parte de uma coleção feita a partir obras de diferentes fontes;
- **Modificação:** A licença deve permitir a criação de derivativos da obra licenciada, e permitir a distribuição de tais derivativos sob os mesmos termos que a obra original licenciada;
- **Separação:** O licença deve permitir que qualquer parte da obra seja livremente utilizada, distribuída ou modificada, separadamente de qualquer outra parte da obra ou de qualquer coleção de obras na qual tenha sido originalmente distribuída. Todos aqueles que recebem qualquer distribuição de qualquer parte de uma obra dentro dos termos da licença original, devem ter os mesmos direitos que aqueles que foram garantidos em conjunção com a obra original;
- **Compilação:** A licença deve permitir que a obra licenciada seja distribuída com outras obras distintas, sem colocar restrições sobre tais obras;

<sup>3</sup> O formato PDF foi considerado aberto a partir da versão 1.7 em 2008 quando a Adobe abriu mão dos seus direitos afim de promover a adoção generalizada do formato. Mais detalhes na *PDF Association* <<https://www.pdfa.org/about-us/the-portable-document-format/>>

• **Não-discriminação:** A licença não deve discriminar qualquer pessoa ou grupo.

• **Propagação:** Os direitos associados à obra devem aplicar-se a todos a quem é redistribuída, sem a necessidade de concordar com quaisquer termos legais adicionais;

• **Aplicação para Qualquer Finalidade:** O licença deve permitir o uso, redistribuição, modificação e compilação para qualquer finalidade. A licença não deve restringir ninguém de fazer uso da obra em um campo específico de atuação;

• **Sem custo:** A licença não deve impor qualquer arranjo de taxa, direitos autorais, ou outra compensação ou remuneração monetária, como parte de suas condições;

#### Condições aceitáveis:

• **Atribuição:** A licença pode exigir que as distribuições da obra incluam atribuição de colaboradores, de titulares de direitos, de patrocinadores e criadores, desde que tais prescrições não sejam onerosas;

• **Integridade:** A licença pode exigir que versões modificadas da obra licenciada tenham um nome ou número de versão diferente da obra original, ou de outro modo indicar que mudanças foram feitas;

• **Share-alike:** A licença pode requerer que cópias ou derivativos da obra licenciada permaneçam sob licença igual ou semelhante à original;

• **Aviso:** A licença pode requerer a retenção de avisos de direitos autorais e identificação da licença.

• **Fonte:** A licença pode requerer que obras modificadas sejam disponibilizadas de forma específica, que garanta modificações subsequentes;

• **Proibição de Restrição Técnica:** A licença pode proibir a distribuição da obra em caso de medidas técnicas imporem restrições aos exercício dos direitos, que sem tais medidas são permitidos;

• **Não-agressão:** A licença pode exigir modificadores para conceder as permissões públicas adicionais (por exemplo, licenças de patentes), como necessários para o exercício dos direitos concedidos pela licença. A licença pode também condicionar permissões à não-agressão contra licenciados, com relação ao exercício de qualquer direito permitido (novamente, por exemplo, litígios de patentes).

Mesmo fazendo uso de licença aberta, se faz necessária a discriminação detalhada quando se publicam dados abertos. Com isso, foram criadas dois tipos de licenças, sendo essas:

(i) Licença sobre conteúdos criativos (textos, imagens, dentre outros) denominada de *Creative Commons*;

(ii) Licença voltada a banco de dados, denominada de *Open Data Commons*.

O Quadro 5 apresenta com detalhes as licenças do tipo *Creative Commons*.

Quadro 5 – Tipos licença *Creative Commons*

<b>Licença</b>	<b>Descrição</b>
Atribuição CC-Zero	Esta licença permite que outros distribuam, remixem, adaptem e criem a partir dos dados, mesmo para fins comerciais. É a licença mais flexível de todas as licenças disponíveis. É denominada de licença universal ou de dedicação ao domínio público.
Atribuição CC BY	Esta licença permite que outros distribuam, remixem, adaptem e criem a partir dos dados, mesmo para fins comerciais, desde que atribuam o devido crédito pela criação original. É recomendada para maximizar a disseminação e uso dos materiais licenciados.
Atribuição-Compartilha Igual CC BY-SA	Esta licença permite que outros remixem, adaptem e criem a partir dos dados, mesmo para fins comerciais, desde que lhe atribuam o devido crédito e que licenciem as novas criações sob termos idênticos. Esta licença costuma ser comparada com as licenças de software livre e de código aberto "copyleft". Todos os trabalhos novos baseados no seu terão a mesma licença, portanto quaisquer trabalhos derivados também permitirão o uso comercial. Esta é a licença usada pela Wikipédia e é recomendada para materiais que seriam beneficiados com a incorporação de conteúdos da Wikipédia e de outros projetos com licenciamento semelhante.
Atribuição-Sem Derivações CC BY-ND	Esta licença permite a redistribuição, comercial e não comercial, desde que o trabalho seja distribuído inalterado e no seu todo, com o devido crédito atribuído .
Atribuição-Não Comercial CC BY-NC	Esta licença permite que outros remixem, adaptem e criem a partir dos seus dados para fins não comerciais, e embora os novos trabalhos tenham que atribuir o devido crédito e não possam ser usados para fins comerciais, os usuários não têm de licenciar esses trabalhos derivados sob os mesmos termos.
Atribuição-Não Comercial- Compartilha Igual CC BY-NC-SA	Esta licença permite que outros remixem, adaptem e criem a partir dos dados para fins não comerciais, desde que atribuam o devido crédito e que licenciem as novas criações sob termos idênticos.

Atribuição-Sem Derivações Sem Derivados CC BY-NC-ND	Esta é a mais restritiva das seis licenças principais, só permitindo que outros façam download dos dados e os compartilhem desde que atribuam o crédito, mas sem que possam alterá-los de nenhuma forma ou utilizá-los para fins comerciais.
---	--

Fonte: (CC, 2019)

Vale destacar que nem todas as licenças descritas no Quadro 5, são licenças abertas. Apenas as licenças **Atribuição CC-Zero**, **Atribuição CC BY** e **Atribuição Compartilhável CC BY-SA** podem ser categorizadas como licenças abertas. O Quadro 6 apresenta as licenças do tipo *Open Data Commons*.

Quadro 6 – Tipos licença *Open Data Commons*

Licença	Descrição
<i>Public Domain Dedication and License (PDDL)</i>	Permite o compartilhamento, a modificação e o uso livremente do banco de dados, para qualquer finalidade e sem quaisquer restrições.
Attribution License (ODC-By)	Permite o compartilhamento, a modificação e o uso livremente do banco de dados, desde que seja a autoria identificada.
Open Database License (ODC-ODbL)	Permite o compartilhamento, a criação e adaptação no banco de dados, no entanto, requer atribuição de autoria e compartilhamento sob a mesma licença.

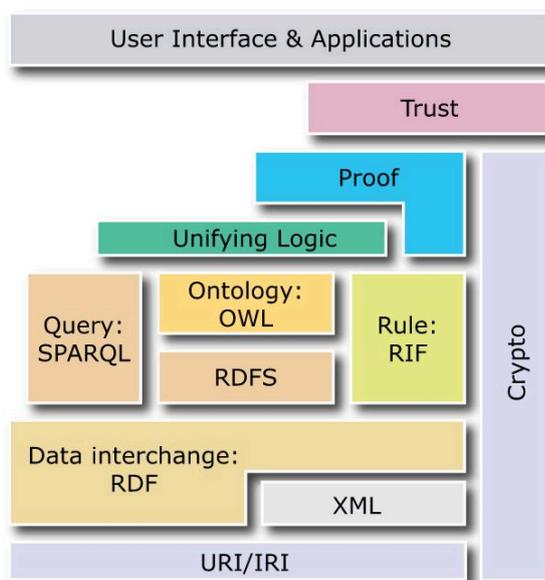
Fonte: (OPC, 2019)

## 2.7 TECNOLOGIAS DA WEB SEMÂNTICA

Existem padrões e tecnologias presentes na Web semântica, estes formam a arquitetura da Web semântica, sendo proposta pelo W3C e apresentada na Figura 7.

Observando a representação da arquitetura da Web Semântica fica evidente a utilização de camadas que proporcionam a interoperabilidade. Os padrões são amplamente utilizados por serem consolidados e recomendados pelo W3C. A seguir apresenta-se um breve detalhamento sobre algumas das camadas que compõem esta arquitetura.

Figura 7 – Arquitetura da Web Semântica



Fonte: (W3C, 2019b)

- URI / IRI: *URI* de forma simples e extensível à identificação de recursos, usando de uma sentença compacta de caracteres, assegurando a sua unicidade (BERNERS-LEE *et al.*, 1998; GOÑI *et al.*, 2002). A *Internacional Resource Identifier (IRI)* é um novo protocolo que complementa a URI, utilizando de caracteres do *Universal Character*<sup>4</sup>(W3C, 2016).
- XML: *XML* é classificada como linguagem de marcação, é baseada na representação de informações estruturadas utilizando apenas textos simples (W3C, C., 2016). O XML possui padrões relacionados, como os *namespace (NS)*, usando de esquemas para estruturar dados na Web sem nenhuma comunicação e nenhum significado. Sendo a base para outras tecnologias semânticas (AGHAEI *et al.*, 2012).
- RDF: *RDF* é um modelo de dados que objetiva representar informações, utilizando URIs para identificar recursos baseados na Web e descrever relações entre os recursos. Esta estrutura pode melhorar o processamento das informações por máquina, suportando normalmente a interoperabilidade a nível semântico. Quando se faz uso de RDF, conseqüentemente há uma semântica, deste modo os agentes de software são capazes de fazer inferências lógicas a fim de executar funções baseadas em metadados (LIMA; CARVALHO, 2005; AGHAEI *et al.*, 2012). A estrutura das informações declaradas são em forma de triplas, representadas por Sujeito, Predicado e Objeto. Tanto sujeito como objeto denotam

<sup>4</sup> Mais detalhes em <http://www.unicode.org/>

um recurso na Web, enquanto o predicado representará a relação entre os dois, sendo identificada por URIs (ANTONIOU, 2004; HARTH *et al.*, 2011).

- RDFS: *Resource Description Framework Schema (RDFS)* é responsável por definir vocabulários no modelo RDF, descrevendo as propriedades e classes de recursos específicos que serão aplicadas em determinado domínio. Possibilitando assim uma estrutura simples de raciocínio para realizar inferências nos tipos de recursos (AGHAEI *et al.*, 2012; GOMES, F. C. d. R., 2013; HARTH *et al.*, 2011).
- Ontologia: Tem por objetivo especificar formalmente um domínio ou discurso, tendo como principal característica uma lista finita de termos e seus relacionamentos (ANTONIOU, 2004). A ontologia é representada computacionalmente utilizando o *Web Ontology Language (OWL)*, permitindo descrever os aspectos semânticos e relacionamentos entre os conceitos de um domínio (RAMALHO; OUCHI, 2011a; GOMES, F. C. d. R., 2013). Apresentado também como complemento para documentos RDF e RDFS, possibilitando aumentar o nível de expressividade e a realizações de inferências (WAGNER FILHO; LÓSCIO, 2016).
- RIF: O *Rule Interchange Format (RIF)* é uma linguagem derivada do XML e um complemento para o OWL, sendo um padrão para o intercâmbio de regras orientadas a lógica de primeira ordem, possibilitando uma interoperabilidade entre as regras lógicas utilizadas em outros sistemas (gomes2013arquitetura; DOMINGUE *et al.*, 2011; RAMALHO; OUCHI, 2011b).
- SPARQL: é um protocolo utilizado para consulta declarativa semelhante ao *Structured Query Language (SQL)*, utilizado sobre dados RDF. Não possui meios para realizar inferências, sendo assim somente possibilita a manipulação das informações armazenadas (gomes2013arquitetura; DOMINGUE *et al.*, 2011; RAMALHO; OUCHI, 2011b).
- *Unifying Logic*: tem por objetivo facilitar a realização de inferências por agentes de software. Esta camada contém as regras que poderão ser utilizadas na execução de serviços inteligentes, podendo auxiliar no relacionamento e processamento das informações contidas em estrutura de dados (gomes2013arquitetura; AGHAEI *et al.*, 2012).
- *Proof*: se refere a prova, sendo responsável pela verificação da consistência dos dados que serão acessados pela Web Semântica. Esta verificação se dá através da execução das regras de inferências definidas na camada lógica, visando garantir a segurança e a veracidade ao conteúdo publicado (gomes2013arquitetura; AGHAEI *et al.*, 2012).

- *Trust*: é responsável por analisar a confiabilidade da prova dada (KOIVUNEN; MILLER, 2001). Esta análise se dá avaliando com auxílio da assinatura digital, a autenticidade das fontes e a confiabilidade das informações consultadas pelos agentes (**gomes2013arquitetura**; AGHAEI *et al.*, 2012).

Contudo, mesmo com tais tecnologias disponíveis, é recomendado que os dados estejam completos e adequados para que estes possam se utilizados de maneira confiável. A subseção seguinte aborda as boas práticas tanto para a publicação como para o uso de dados na Web.

## 2.8 BOAS PRÁTICAS DE DADOS NA WEB

Desde a difusão dos dados na Web, existem dificuldades, como: grande heterogeneidade, falta de padrões para a descrição e acesso aos conjuntos de dados, refletindo diretamente na complexidade do processo de publicação, compartilhamento e uso dos dados. Contudo, existem esforços para estreitar o entendimento comum entre os atores envolvidos em todo o processo, desde a abertura até o seu consumo. Promovendo uma relação de confiança e aumentando o potencial dos dados. Caso não exista esta confiança e a sincronia entre os atores envolvidos, todos os esforços feitos pelos publicadores de dados serão em vão, decorrente da incompatibilidade das necessidades dos consumidores (LÓSCIO, B. F. *et al.*, 2016, 2015)

A fim de buscar esta sincronia entre os atores e o entendimento comum, a W3C juntamente com um grupo de pesquisadores, compilaram um conjunto de doze casos de uso, que representam cenários reais e comuns de como os dados são publicados e utilizados na Web (DEIRDRE LEE *et al.*, 2019). A partir deste estudo, identificaram-se os principais desafios e os requisitos que publicadores e consumidores enfrentam e desejam. Deste modo, foi criado um documento aplicável a qualquer domínio, podendo ser complementado quando necessário por outros documentos ou normas mais específicas. Este documento conta com abordagens técnicas, exemplo: formato de dados, e não-técnicas, por exemplo: o uso e escolha de licenças (LÓSCIO, B. F. *et al.*, 2017). O quadro 7 apresenta os desafios de dados na Web categorizados, bem como as boas práticas.

Quadro 7 – Desafios e boas práticas de dados na web

<b>Desafio</b>
<b>Metadados:</b> Permitir que os seres humanos entendam os metadados, interpretando a natureza e a estrutura dos dados, e que as máquinas também possam processá-los
<b>Boas Práticas</b>
<b>BP1 - Fornecer metadados:</b> Os seres humanos serão capazes de compreender os metadados, e os agentes de software serão capazes de processá-los.

<b>B2 - Fornecer metadados descritivos:</b> Os seres humanos serão capazes de interpretar a natureza do conjunto de dados e suas distribuições, e os agentes de software serão capazes de descobrir automaticamente conjuntos de dados e distribuições.
<b>B3 - Fornecer metadados estruturais:</b> Os seres humanos serão capazes de interpretar o esquema de um conjunto de dados, e os agentes de software serão capazes de processar automaticamente os dados das distribuições.
<b>Desafio</b>
<b>Licença:</b> Permitir que os seres humanos compreendam as informações da licença e que as máquinas possam detectar automaticamente
<b>Boas Práticas</b>
<b>B4 - Fornecer informações sobre a licença de dados:</b> Os seres humanos serão capazes de compreender a licença de dados, descrevendo eventuais restrições impostas à utilização de certos dados, agentes de software serão capazes de detectar automaticamente a licença de dados de uma distribuição.
<b>Desafio</b>
<b>Proveniência:</b> Permitir que os seres humanos conheçam a origem ou o histórico do conjunto de dados e que as máquinas possam processar automaticamente tais informações.
<b>Boas Práticas</b>
<b>BP5 - Fornecer informações de proveniência dos dados:</b> Os seres humanos serão capazes de identificar a origem dos conjuntos de dados, e agentes de software serão capazes de processar automaticamente informações de proveniência.
<b>Desafio</b>
<b>Qualidade:</b> Documentar a qualidade dos dados, para facilitar o processo de seleção dos conjuntos de dados e de reutilização
<b>Boas Práticas</b>
<b>BP6 - Fornecer informação de qualidade de dados:</b> Os seres humanos e os agentes de software serão capazes de avaliar a qualidade e, portanto, a adequação de um conjunto de dados para a sua aplicação.
<b>Desafio</b>
<b>Versionamento:</b> Permitir que versões dos dados sejam geradas e seja possível o acesso a cada versão
<b>Boas Práticas</b>
<b>BP7 - Fornecer indicador de versão:</b> Os seres humanos e os agentes de software poderão facilmente determinar qual versão de um conjunto de dados.
<b>BP8 - Fornecer o histórico de versões:</b> Os seres humanos e os agentes de software serão capazes de entender como o conjunto de dados muda de versão para versão e como quaisquer duas versões específicas diferem.
<b>Desafio</b>
<b>Identificação:</b> Fornecer identificadores únicos para os conjuntos de dados e distribuições
<b>Boas Práticas</b>

<p><b>BP9 - Usar URIs persistentes como identificadores de conjuntos de dados:</b> Os conjuntos de dados ou informações sobre conjuntos de dados poderão ser descobertos e citados ao longo do tempo, independentemente da sua disponibilidade ou do formato dos dados.</p>
<p><b>BP10 - Usar URIs persistentes como identificadores dentro de conjuntos de dados:</b> Os itens de dados serão relacionados em toda a Web criando um espaço global de informação acessível a humanos e máquinas.</p>
<p><b>BP11 - Atribuir URIs para as versões dos conjuntos de dados e séries:</b> Os seres humanos e os agentes de software serão capazes de referenciar versões específicas de um conjunto de dados, séries de conjunto de dados, bem como a versão mais recente de um conjunto de dados.</p>
<p><b>Desafio</b></p>
<p><b>Formato:</b> Escolher formatos que permitam o uso e o reuso</p>
<p><b>Boas Práticas</b></p>
<p><b>BP12 - Usar formatos de dados padronizados legíveis por máquina:</b> Máquinas serão capazes de ler e processar dados publicados na Web e os seres humanos serão capazes de usar ferramentas computacionais para manipular os dados.</p>
<p><b>BP13 - Usar representações de dados que sejam independentes de localidade (locais neutros):</b> Os seres humanos e os agentes de software serão capazes de interpretar o significado do conjunto de caracteres (strings) que representam datas, horas, moedas, números, entre outros, com precisão</p>
<p><b>BP14 - Fornecer dados em vários formatos:</b> Tantos usuários quanto possível serão capazes de utilizar os dados sem ter que transformá-los em seu formato preferido.</p>
<p><b>Desafio</b></p>
<p><b>Vocabulários:</b> A fim de melhorar a interoperabilidade e manter terminologia comum entre os produtores e consumidores</p>
<p><b>Boas Práticas</b></p>
<p><b>BP15 - Reutilizar vocabulários, dando preferência aos padronizados:</b> Interoperabilidade e consenso entre os publicadores e consumidores de dados serão reforçados.</p>
<p><b>BP16 - Escolher o nível de formalização adequado:</b> Os casos de aplicação mais prováveis serão apoiados com não mais complexidade do que o necessário.</p>
<p><b>Desafio</b></p>
<p><b>Acesso:</b> Permitir o fácil acesso aos dados usando a infraestrutura da Web tanto para seres humanos quanto para máquinas.</p>
<p><b>Boas Práticas</b></p>
<p><b>BP17 - Fornecer 'bulk download':</b> transferências de arquivos grandes, ou seja, que exigem mais tempo do que um usuário típico consideraria razoável, serão possíveis por meio de protocolos de transferência de arquivos dedicados.</p>

<p><b>BP18: Fornecer subconjuntos para conjuntos de dados grandes:</b> Os seres humanos e as aplicações serão capazes de acessar subconjuntos de um conjunto de dados, em vez de todo o conjunto. Isso proporcionará aos consumidores o acesso aos dados com uma elevada proporção de dados que são realmente necessários em comparação aos dados desnecessários. Conjuntos de dados estáticos considerados muito grandes poderão ser recuperados em porções menores. APIs podem ser usadas para filtrar os dados disponíveis. A granularidade de acesso aos dados poderá ser definida de acordo com as necessidades do domínio e as demandas de desempenho das aplicações.</p>
<p><b>BP19 - Usar <i>content negotiation</i> para servir os dados disponíveis em vários formatos:</b> O <i>Content negotiation</i> permitirá que diferentes recursos ou representações diferentes de um mesmo recurso possam ser servidas de acordo com a requisição feita pelo cliente.</p>
<p><b>BP20 - Fornecer acesso em tempo real:</b> Aplicações serão capazes de acessar os dados em tempo real ou quase em tempo real; em tempo real significa um intervalo de milissegundos até alguns segundos após a criação de dados.</p>
<p><b>BP21 - Fornecer dados atualizados:</b> Os dados na Web serão atualizados em tempo hábil para que os dados disponíveis on-line reflitam os dados mais recentes divulgados através de qualquer outro canal. Quando novos dados estiverem disponíveis, logo que possível, serão publicados na Web.</p>
<p><b>BP22 - Fornecer uma explicação para os dados que não estão disponíveis:</b> Os consumidores saberão que os dados que são referenciados a partir do conjunto de dados não estão disponíveis ou se estão disponíveis sob diferentes condições.</p>
<p><b>BP23 - Tornar os dados disponíveis através de uma API:</b> Os desenvolvedores terão acesso aos dados para uso em seus próprios aplicativos, com dados atualizados e sem a necessidade de esforço por parte dos consumidores. As aplicações serão capazes de obter dados específicos por meio de consultas à API.</p>
<p><b>BP24 - Usar padrões Web como base para construção de APIs:</b> Desenvolvedores que tenham alguma experiência com APIs baseadas em padrões Web, tais como o REST, já deverão ter um conhecimento inicial de como usar a API. Além disso, será mais fácil dar manutenção na API.</p>
<p><b>BP25 - Fornecer documentação completa de sua API:</b> Os desenvolvedores poderão obter informações detalhadas sobre cada chamada para a API, incluindo os parâmetros necessários e o retorno esperado, ou seja, todo o conjunto de informações relacionadas à API. O conjunto de valores - como usá-lo, avisos de alterações recentes, informações de contato e assim por diante - deve ser descrito e facilmente navegável na Web. Ele também permite que as máquinas acessem a documentação da API para ajudar os desenvolvedores a criar o software cliente da API.</p>

<p><b>BP26 - Evitar alterações que afetem o funcionamento de sua API:</b> O código do desenvolvedor deve continuar válido após alterações na API. Os desenvolvedores devem ser notificados das melhorias feitas na API e devem ser capazes de fazer uso delas. Quebrar alterações em sua API será raro e, se ocorrer, os desenvolvedores terão tempo e informações suficientes para adaptar o seu código, aumentando a confiança na API. Alterações na API deverão ser anunciadas no site da documentação da API.</p>
<p><b>BP26 - Evitar alterações que afetem o funcionamento de sua API:</b> O código do desenvolvedor deve continuar válido após alterações na API. Os desenvolvedores devem ser notificados das melhorias feitas na API e devem ser capazes de fazer uso delas. Quebrar alterações em sua API será raro e, se ocorrer, os desenvolvedores terão tempo e informações suficientes para adaptar o seu código, aumentando a confiança na API. Alterações na API deverão ser anunciadas no site da documentação da API.</p>
<p><b>Desafio</b></p>
<p><b>Preservação:</b> A fim de indicar corretamente se os dados foram removidos ou arquivados</p>
<p><b>Boas Práticas</b></p>
<p><b>BP27 - Preservar identificadores:</b> A URI de um conjunto de dados sempre levará para o conjunto propriamente dito ou então redirecionará para um recurso com informações sobre ele.</p>
<p><b>BP28 - Avaliar a cobertura do conjunto de dados:</b> Os usuários serão capazes de fazer uso de dados arquivados no futuro.</p>
<p><b>Desafio</b></p>
<p><b>Feedback:</b> Receber <i>feedback</i> dos consumidores e assegurar que os dados atendem as necessidades dele.</p>
<p><b>Boas Práticas</b></p>
<p><b>BP29 - Coletar <i>feedback</i> dos consumidores de dados:</b> Os consumidores de dados serão capazes de fornecer <i>feedback</i> e avaliações sobre conjuntos de dados e distribuições.</p>
<p><b>BP30 - Compartilhar o <i>feedback</i> disponível:</b> Os consumidores serão capazes de avaliar os tipos de erros que afetam o conjunto de dados, avaliar experiências de outros usuários e ter a certeza de que o publicador trata os problemas de forma adequada. Os consumidores também serão capazes de determinar se outros usuários já fizeram comentários semelhantes, poupando-lhes a submissão de relatórios desnecessários e poupando os publicadores de terem que lidar com duplicatas.</p>
<p><b>Desafio</b></p>
<p><b>Enriquecimento:</b> Enriquecer, melhorar ou refinar os dados brutos agregando valor.</p>
<p><b>Boas Práticas</b></p>

<p><b>BP31 - Enriquecer dados por meio da geração de novos dados:</b> Os conjuntos de dados com valores nulos poderão ser "corrigidos" a partir do preenchimento de tais valores. Estrutura poderá ser conferida aos dados e sua utilidade poderá ser melhorada se forem adicionadas medidas ou atributos relevantes. Porém, tal adição só deverá ser feita se não alterar os resultados analíticos, o significado ou o poder estatístico dos dados.</p>
<p><b>BP32 - Fornecer visualizações complementares:</b> Complementar os conjuntos de dados com possíveis visualizações permitirá que os consumidores humanos tenham uma visão imediata sobre os dados, apresentando-os de forma que possam ser facilmente compreendidos.</p>
<p><b>Desafio</b></p>
<p><b>Republicação:</b> Permitir que os dados utilizados possam ser republicados</p>
<p><b>Boas Práticas</b></p>
<p><b>BP33 - Fornecer <i>feedback</i> para o publicador original:</b> Uma melhor comunicação entre publicadores e consumidores fará com que seja mais fácil para os publicadores originais determinar como os dados que eles publicam estão sendo usados. Isso ajudará a justificar o investimento na publicação dos dados. Os publicadores também serão informados de medidas que podem ser tomadas para melhorar a qualidade dos seus dados.</p>
<p><b>BP34 - Obedecer os termos de licença:</b> Os publicadores serão capazes de confiar que seu trabalho está sendo reutilizado de acordo com os seus requisitos de licenciamento, tornando-os mais propensos a continuar com a publicação dos dados. Reutilizadores de dados vão ser capazes de licenciar adequadamente os trabalhos derivados a partir de dados previamente publicados.</p>
<p><b>BP35 - Citar a publicação original do conjunto:</b> Os consumidores finais serão capazes de avaliar a origem dos dados e os esforços dos publicadores originais serão reconhecidos. A cadeia de proveniência para os dados na Web será rastreável de volta ao seu publicador original.</p>

Fonte: (LÓSCIO, B. F. *et al.*, 2017)

Estas boas práticas descritas no Quadro 7 foram desenvolvidas com o objetivo de orientar a área técnica na publicação de dados na Web bem como seu uso, contribuindo para melhorar a relação entre publicadores e consumidores de dados (LÓSCIO, B. F. *et al.*, 2016). Os autores apresentam o contexto de dados na Web, no entanto percebe-se que estes desafios e boas práticas se refletem também no âmbito de dados abertos e dados abertos conectados.

Neste sentido, os dados padronizados precisam estar disponíveis para acesso, sendo comumente feito o uso de portais para disponibilizá-los. Alguns destes portais possuem uma estrutura que viabiliza a catalogação dos dados, bem como sua visualização interativa. Portanto a seguir apresenta-se a contextualização dos portais.

## 2.9 PORTAIS DE DADOS ABERTOS

Os portais são amplamente utilizados por conta de sua praticidade, baixo custo, rapidez na disseminação de conteúdo e a difusão em escala global (SALES DOS SANTOS; GONCALVES DA SILVA, 2011). Com a evolução da tecnologia, um portal tornou-se mais que um centralizador de informações dispersas. Ampliando suas funcionalidades, possibilitando por exemplo: pesquisa de conteúdo e recursos, categorização dos conteúdos, conteúdo baseado em perfil de usuário, atividades colaborativas, como: os fóruns e vários tipos de tarefas intensivas em conhecimento (MACK *et al.*, 2001).

Portal de dados em um contexto tecnológico é definido como "qualquer sistema que permite ao usuário obter informações e conhecimento de que necessite, fazendo uso de um navegador" (CHAN; CHUNG, 2002).

O uso dos portais se destacou em diversos meios, como: jornalístico, entretenimento, governamental, dentre outros, sendo um canal direto com a sociedade (NOVA URIARTE, 2006). Portanto, é possível compreender o motivo porque é utilizado por diversas instituições que publicam dados abertos.

Para a publicação e compartilhamento dos dados, se faz uso de portais, alguns destes são desenvolvidos com o objetivo de facilitar todo o processo de publicação, compartilhamento e utilização dos dados. A seguir serão apresentadas algumas das plataformas mais conhecidas e utilizadas para implementação de portais.

### 2.9.1 Comprehensive Knowledge Archive Network - CKAN

O CKAN<sup>5</sup> é uma solução *open source* do tipo catálogo de dados, que oferece funções para publicação, armazenamento e gerenciamento dos conjuntos de dados. Por ser uma plataforma livre, seu uso vem se destacando em diversos países: Brasil, Uruguai, Canadá, Reino Unido, Estados Unidos, Romênia, Austrália, Holanda, Itália e Alemanha (CKAN, 2019).

O CKAN é uma plataforma com várias funcionalidades. Este permite a criação de metadados, uso de múltiplos formatos de arquivos, o uso de vocabulários, a divulgação de conjuntos de dados em redes sociais, bem como a utilização de *Application Programming Interface (API)*, permitindo aos desenvolvedores realizarem requisições diretamente a plataforma de maneira facilitada e possibilitar a coleta e utilização dos dados<sup>6</sup>.

O CKAN também se destaca por permitir a personalização de sua interface, deixando a critério da instituição aspectos visuais e de interatividade<sup>7</sup>. Outro ponto é a existência da comunidade CKAN<sup>8</sup>, que trabalha em prol da plataforma, desen-

<sup>5</sup> <<https://ckan.org/>>

<sup>6</sup> <<https://docs.ckan.org/en/2.8/api/index.html>>

<sup>7</sup> <<https://docs.ckan.org/en/2.8/theming/index.html>>

<sup>8</sup> <<https://extensions.ckan.org/>>

volvendo extensões, deixando-a mais robusta e completa. Porém, as extensões por serem desenvolvidas pela comunidade não são pré-configuradas ou instaladas por padrão (CKAN, 2019). A Figura 8 mostra um exemplo de portal de dados que utiliza o CKAN, neste caso, o portal de dados abertos do governo federal do Brasil.

Figura 8 – Portal de Dados Abertos Governo Federal do Brasil - CKAN

The screenshot shows the homepage of the Brazilian Open Data Portal (dados.gov.br). The header includes navigation links for 'Services', 'Simplifique!', 'Participe', 'Information access', 'Legislation', and 'Information channels'. The main content area features a search bar with the text 'Pesquisar conjuntos de dados...' and a 'PESQUISAR' button. Below the search bar, it indicates 'em 6455 conjuntos de dados com 33159 recursos (o que é isto?)'. There are three featured dataset sections: 'Conjuntos de dados em destaque', 'Serviços Públicos Federais para o Cidadão', and 'Orçamento Federal - Orçamento Fiscal e da Seguridade Social'. On the right side, there is a table titled 'Publicações mais recentes' with columns for 'Conjunto de dados', 'Data', and 'Responsável'.

Conjunto de dados	Data	Responsável
DOCENTES-CAMPOS BELOS	05/02/2019	Instituto Federal de Educação, Ciênc...
Coleta de Dados, Discentes dos Programas de Pós-Graduação stricto sensu no Brasil 2004 a 2012	05/02/2019	Coordenação de Aperfeiçoamento de Pe...
Coleta de Dados, Projetos dos Programas de Pós-Graduação stricto sensu no Brasil 2004 a 2012	05/02/2019	Coordenação de Aperfeiçoamento de Pe...
Coleta de Dados, Docentes dos Programas de Pós-Graduação stricto sensu no Brasil 2004	05/02/2019	Coordenação de Aperfeiçoamento de Pe...

Fonte: (PORTAL DE DADOS GOV, 2019)

Há limitações de funcionalidades em alguns portais, por exemplo, não faz-se o uso de vocabulários, como o *Data Catalog Vocabulary (DCAT)*. Além de não utilizar de versionamento nos conjuntos de dados e não possibilitar a visualização de formatos de dados, dentre outros.

O uso do DCAT, possibilita a compreensão dos significados referentes ao conjunto de dados vinculado, de modo que uma máquina possa interpretar o conteúdo sem dificuldades e realizar inferências automatizadas. Nesse sentido, destaca-se a importância de conhecer as funcionalidades que a plataforma dispõe, de modo a, contribuir com a semântica dos dados disponibilizados.

Outro ponto a ser destacado é o uso de API no CKAN, que ao contrário das extensões como o DCAT, por exemplo, é uma opção nativa da plataforma, ou seja, está habilitada ao instalar. Para efetuar algumas consultas o usuário pode utilizar alguns recursos padrões para coletar informações sobre os conjuntos de dados presentes no portal e esse tipo de consulta através de API é utilizada por agentes computacionais, que contribui com a legibilidade por máquina.

## 2.9.2 Socrata

O Socrata é uma plataforma baseada em nuvem, tendo como principal diferencial a visualização de dados. Esta plataforma é usada como portal oficial do Kenya, do Banco Mundial, em algumas cidades e estados dos EUA, como Chicago, Nova York, Austin e Maryland. Ao contrário do CKAN, o Socrata é pago (SOCRATA, 2019a).

Por ser uma plataforma em nuvem, não necessita de uma infraestrutura para a instalação em um servidor como acontece com o CKAN. O Socrata tem como o foco a simplicidade e para isso disponibiliza apoio necessário para a criação de um portal, inclusive há uma preocupação com a interface, de modo que seja amigável e simples para o usuário final (cidadãos) (SOCRATA, 2019b). A Figura 9 mostra um exemplo de portal de dados que utiliza o Socrata, neste caso, o portal de dados abertos da cidade de Virginia Beach.

Figura 9 – Tela da plataforma Socrata

The screenshot displays the Socrata data portal for the City of Virginia Beach. At the top, there is a search bar and a navigation menu with links like 'VBGov Home', 'Catalog', 'Suggest a Dataset', 'About this Site', 'Tutorials', 'Developers', 'Open Performance', 'Open Budget', and 'Open Spending'. Below the navigation, there is another search bar and a 'Sign In' button. The main content area shows 40 results, sorted by 'Most Relevant'. The first three results are:

Dataset Name	Category	Updated	Views
Police Incident Reports	Public Safety	February 11, 2019	31,536
Employee Salaries	Finance / Government	August 13, 2018	15,378
Police Calls for Service	Public Safety		

Fonte: (VIRGINIA BEACH, 2019)

### 2.9.3 OpenDataSoft

O OpenDataSoft é uma empresa que foi fundada em 2011 que desenvolveu uma plataforma para compartilhar e reutilizar os dados de empresas públicas e privadas (OPENDATASOFT, 2019a). A empresa destaca que seu diferencial está na facilidade de busca, navegação e no uso de filtros nos conjunto de dados ou mesmo a possibilidade de consultas à partir de SPARQL, dentre outras funcionalidades da plataforma (OPENDATASOFT, 2019c).

Assim como o Socrata, o OpenDataSoft (nome da plataforma, que coincide com o nome da empresa) utiliza do conceito de usuários não-técnicos, disponibilizando toda uma infraestrutura em nuvem e uma estrutura de suporte (OPENDATASOFT, 2019b). A Figura 10 mostra um exemplo de portal de dados que utiliza o OpenDataSoft, neste caso, o portal de dados abertos disponibilizado pela própria OpenDataSoft, onde este centraliza os conjuntos de dados dos seus clientes.

Figura 10 – Tela da plataforma Opendatasoft

The screenshot shows the OpenDataSoft Data Network interface. At the top, there is a navigation bar with links for Data, Maps, Charts, API, Help, Reference, and API Prime. The main content area is divided into several sections. On the left, there is a search bar and a list of datasets. The central part of the page features four dataset cards, each with a title, a brief description, and statistics. The cards are: 'V'Lille - Disponibilité en temps réel', 'Disponibilité des places de parkings publics en temps réel - Orléans', 'Ilévia Prochains passages bus et tramway', and 'Concentrations horaires de polluants dans l'air ambiant issues du réseau permanent de mesures automatiques - Poitiers'. Each card includes a title, a brief description, and statistics like '223 records' and 'Popularity 186.6'.

Fonte: (OPENDATASOFT, 2019)

Destaca-se que assim como o Socrata o OpenDataSoft também é uma plataforma paga.

As plataformas consideradas por este trabalho são as comumente utilizadas. Há uma ênfase sobre a plataforma CKAN, visto que esta é a única plataforma não paga e que viabiliza funcionalidades que permitem agregar semântica aos dados nela disponibilizados. Por este motivo, a plataforma CKAN será objeto de estudo neste trabalho.

A seção seguinte aborda os estudos encontrados na literatura que apresentam similaridade ao que se propõe nesta tese, com o intuito de fornecer subsídios para a avaliação de portais de dados abertos.

## 2.10 MODELOS DE AVALIAÇÃO DE PORTAIS DE DADOS ABERTOS

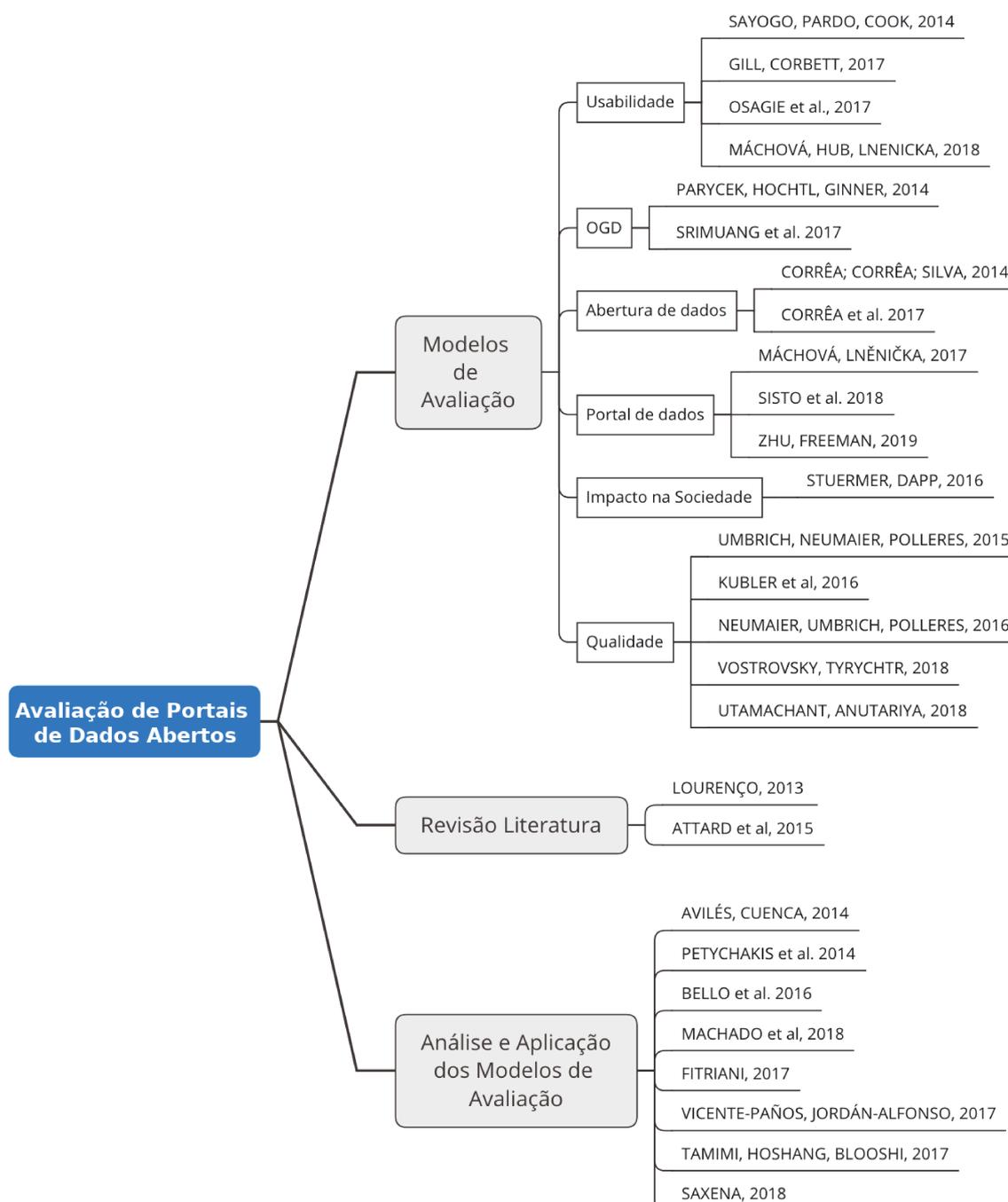
Esta seção apresenta os trabalhos correlatos à proposta elencada por este estudo, com objetivo de evidenciar os conceitos que a literatura utiliza para avaliar portais de dados abertos e seus conjuntos de dados. Para isso, primeiramente foi realizada a classificação dos trabalhos de modo a estruturar as pesquisas da área e a partir disso foi realizada a análise, com o propósito de selecionar os trabalhos que possam contribuir com a composição do modelo. Além disso, foi realizada uma complementação dos estudos elencados considerando novos trabalhos publicados entre o ano de 2019 a março de 2021.

### 2.10.1 Classificação dos estudos

Para a realização da classificação dos estudos encontrados na literatura, foram utilizados parâmetros de busca e refinamento apresentados na seção 1.8.3.2. A busca nas bases de dados resultou em um total de 27 documentos, sendo estes classificados em três grandes grupos: Modelo de Avaliação, Revisão de Literatura e Aplicação dos modelos de avaliação.

Esta classificação foi feita para mapear as pesquisas e auxiliar na identificação dos trabalhos que podem contribuir na composição do modelo. A Figura 11 apresenta a classificação dos 27 trabalhos encontrados na literatura.

Figura 11 – Classificação dos artigos encontrados na literatura



Fonte: do Autor

O grupo **modelos de avaliação** está subdividido em seis subgrupos. Sendo estes:

- **Usabilidade**: Envolve quatro trabalhos, cujo o foco destes estudos está relacionado ao usuário, como exemplo: o processo de aprendizagem do usuário ao usar o portal (GILL; CORBETT, 2017), na avaliação como os usuários não-técnicos utilizam o portal de dados abertos (OSAGIE *et al.*, 2017) e na avaliação utilizando de um *Framework* para verificar a usabilidade do portal (SAYOGO *et al.*, 2014;

MÁCHOVÁ *et al.*, 2018).

- OGD: Envolve os trabalhos relacionados a *Open Government Data* (OGD), considerando a avaliação da implementação de OGD (PARYCEK *et al.*, 2014) e abrangendo aspectos de avaliação da estrutura do OGD (SRIMUANG *et al.*, 2017).
- Impacto na Sociedade: Engloba o trabalho de Stuermer e Dapp (2016) que propõe monitorar o impacto que os dados abertos tiveram na sociedade através do uso do *Social Return on Investment* (SROI).
- Qualidade: Estes trabalhos focam na adequação dos dados, com objetivo de atingir uma qualidade desejável, como é o caso do trabalho de Umbrich *et al.* (2015), que avalia além da qualidade a evolução dos portais de dados abertos, considerando questões elencadas pelos autores, como (recuperabilidade, uso, completude, precisão, abertura e contactabilidade). Com isso, os autores realizaram uma análise temporal dos conjuntos de dados publicados permitindo avaliar a evolução do portal.

Por sua vez, o trabalho de Kubler *et al.* (2016) realiza a comparação da qualidade dos portais utilizando os critérios de qualidades elencados por (UMBRICH *et al.*, 2015) e fazem uso do método *Analytic Hierarchy Process* (AHP) para ranquear os portais. Toda esta avaliação se deu com base em portais que utilizaram o CKAN.

O trabalho de Neumaier *et al.* (2016), os autores criaram um processo automatizado baseado no DCAT, independente da plataforma utilizada, para que possa ser feita a coleta dos dados e informações permitindo avaliar a qualidade dos portais de dados. Este trabalho sucede o trabalho de Kubler *et al.* (2016).

O trabalho de Vostrovsky e Tyrychtr (2018) avalia a qualidade dos conjuntos de dados do setor de agricultura da República Tcheca, a fim de identificar inconsistências por parte dos publicadores. Nesta avaliação os autores utilizaram a ISO/IEC FDIS 25010:2011<sup>9</sup>.

Por fim, o trabalho Utamachant e Anutariya (2018) avalia a qualidade dos conjuntos de dados do portal da Tailândia. Os autores utilizam a ISO25012<sup>10</sup> nesta avaliação.

O grupo **revisão de literatura** possui dois trabalhos atribuídos, sendo Lourenço (2013) que aborda aspectos de avaliação em nível estratégico e Attard *et al.* (2015) que apresenta uma revisão sistemática da literatura com foco nas iniciativas de governo aberto.

<sup>9</sup> <<https://www.iso.org/standard/35733.html>>

<sup>10</sup> <<https://iso25000.com/index.php/en/iso-25000-standards/iso-25012>>

O terceiro e último grupo denominado de **análise e aplicação dos modelos de avaliação**, em como objetivo utilizar de instrumentos de avaliação, aplicando-os em cenários específicos para que sejam feitas análises e discussões sobre os resultados, sendo estes (ARQUERO AVILES; MARCO CUENCA, 2014; PETYCHAKIS *et al.*, 2014; BELLO *et al.*, 2016; MACHADO *et al.*, 2018; FITRIANI *et al.*, 2017; VICENTE-PANOS; JORDAN-ALFONSO, 2017; TAMIMI *et al.*, 2017; SAXENA, 2018).

Foram apresentados resumidamente os trabalhos que não serão utilizados esta pesquisa, devido à apresentarem particularidades que não correspondem ao que se propõem por este trabalho e além de apresentarem instrumentos desenvolvidos superficialmente, inacessíveis ou que inviabilizam a automatização da coleta de dados.

Os trabalhos do subgrupo Portais de Dados ((MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018; ZHU, X.; FREEMAN, 2019)), Abertura de Dados ((CORRÊA, Andreiwid Sh *et al.*, 2017)) e trabalhos encontrados em literatura cinza(*grey literature*) (ODM, 2019; ABELLA *et al.*, 2019), serão discutidos na sequência.

### 2.10.2 Análise dos trabalhos relacionados

A fim de apresentar de forma sintetizada os trabalhos relacionadas o Quadro 8 traz cada trabalho com seu objetivo, a escala de avaliação utilizada e a forma de avaliação. Os principais aspectos destes trabalhos são apresentados no apêndice A.

Quadro 8 – Trabalhos relacionados

Referências	Objetivo	Escalas de Avaliação	Avaliação
Andreiwid Sh Corrêa <i>et al.</i> (2017)	Avaliar a transparência dos portais de dados de 561 municípios brasileiro.	Sim, Não e Parcial, além da nota de acessibilidade informada pela plataforma, Access Monitor.	Cenários de Uso
Máchová e Lnénicka (2017)	Avaliar a qualidade de portais de dados abertos em escala global e compará-los.	Utiliza a escala Likert de 5 níveis.	Cenários de Uso

Sisto <i>et al.</i> (2018)	Avaliar os portais com uma perspectiva de conteúdo.	Utiliza de pontuação conforme os critérios forem atendidos. O somatório destes pontos possibilitam a definição de uma pontuação total sobre o portal de dados.	Cenários de Uso
Xiaohua Zhu e Freeman (2019)	Avaliar os portais de dados dos municípios dos EUA, com foco na interação e experiência do usuário em relação ao portal de dados.	Utiliza de variáveis binárias como respostas em relação às 42 perguntas.	Cenários de Uso
ODM (2019)	Monitora portais de dados da união europeia e também possibilita o comparativo entre os portais.	Quantificação dos critérios utilizados.	Não identificado
Abella <i>et al.</i> (2019)	Avalia os conjuntos de dados com objetivo de verificar a sua usabilidade.	Fórmula matemática própria.	Não identificado

Fonte: do Autor

Os trabalhos apresentados nesta seção são iniciativas que contribuem fornecendo diferentes perspectivas de avaliação sobre portais de dados abertos, no entanto, percebem-se algumas limitações.

Em relação à proposta de Andreiwid Sh Corrêa *et al.* (2017), observa-se que ao basear-se o estudo em uma lei, mesmo realizando uma co-relação entre aspectos da literatura, isso pode acarretar em uma oscilação na proposta à medida que a lei é atualizada. Contudo, este é o único estudo que considera questões de acessibilidade e apresenta uma plataforma que possibilita a avaliação de critérios determinados pela W3C. Além disso, o modelo apresentado é simples, mas consegue proporcionar uma visão geral sobre os portais de dados sob a ótica da OGD.

Por sua vez, observa-se que a proposta de Máchová e Lnénicka (2017) possui uma fragilidade por utilizar um questionário e um grupo de pessoas para avaliar uma quantidade considerável de portais de dados abertos. Isso porque, ao utilizar um questionário com um volume de critérios e vários portais a serem avaliados, tudo de

forma manual, podem ocorrer inconsistências. As inconsistências podem ser oriundas de ruídos nas repostas devido ao cansaço, a incorreta interpretação por parte dos participantes, entre outros fatores, refletindo em resultados muitas vezes distorcidos. No entanto, para amenizar a questão do ruído, os autores criaram um "roteiro" para os participantes com objetivo de facilitar a coleta e avaliação dos portais. No entanto, destaca-se que o modelo é um dos poucos que aborda questões sobre a estrutura do portal de dados e dos conjuntos de dados.

No que concerne à proposta de Sisto *et al.* (2018), percebe-se que ao empregar uma pontuação fixa conforme a quantidade de conjuntos de dados disponibilizados pode-se prejudicar a avaliação de portais de dados em fase inicial. Visto que portais novos possuem poucos conjuntos de dados e conseqüentemente teriam uma pontuação baixa, como consequência. Outro item a ser analisado é a pontuação fixa, tornando rígida a avaliação de maneira geral, sem considerar diferentes perspectivas para os diferentes portais de dados abertos. Contudo, o modelo apresentado contempla diversos aspectos que envolvem o portal e os conjuntos de dados, bem como é o único estudo que disponibiliza um total geral sobre a avaliação realizada no portal de dados.

Na proposta de ODM (2019), percebe-se que é o único estudo que utiliza uma estrutura automatizada para efetuar a coleta e monitoramento dos portais de dados, bem como a disponibilização dos resultados em uma plataforma que passa por atualização bimensal. No entanto, os critérios elencados são genéricos e superficiais, focando na análise e não em avaliação. Isso ocorre pois o objetivo desta proposta é compreender e comparar aspectos de dados abertos na Europa. Esta proposta foi selecionada porque contempla critérios que contribuem com a avaliação dos portais de dados.

Por sua vez, observa-se na proposta de Abella *et al.* (2019) uma avaliação superficial, visto o volume de dados disponibilizados nos portais, pois consideram-se perguntas onde não possui valores precisos sobre determinados aspectos dificultando uma avaliação pontual. O trabalho faz uso de um formulário para avaliar o quão reutilizável os dados são, sendo este objetivo e simples atendendo pequenos portais que desejam começar a disponibilizar dados.

### **2.10.3 Complemento aos trabalhos relacionados**

Com o intuito de verificar trabalhos que foram desenvolvidos em paralelo com esta pesquisa, foram realizadas novas buscas nas bases de dados, utilizando-se os mesmos critérios apresentados na seção 1.8.3.2, sendo considerado o período de 2019 a março de 2021.

Após a aplicação dos filtros encontraram-se 27 artigos, destes 14 artigos são focados no uso de dados abertos. Realizou-se uma nova seleção dos trabalhos que possuem semelhança a este estudo, chegando a um total de 7 artigos que são apre-

sentados na sequência. Estes estudos reforçam a relevância da área e sua importância perante a comunidade científica e a sociedade.

- No trabalho de Torino *et al.* (2019) são analisados 46 conjuntos de dados publicados no portal da CAPES, enquadrando-os com as boas práticas de dados na Web (LÓSCIO, B. F. *et al.*, 2017) e com o esquema de classificação de cinco estrelas (BERNERS-LEE, 2010).
- O estudo de Royo-Montanes e Benitez-Gomez (2019) propõe uma estrutura para avaliar portais de dados municipais, com a justificativa de que os estudos existentes não avaliam estes portais. Para isso, eles utilizam outros estudos pré-existentes na construção da sua metodologia. Contudo, a estrutura proposta avalia os portais de dados de modo simples e objetiva, sem considerar a maioria dos padrões de dados abertos.
- A proposta de Zhang e Xiao (2020) faz uma meta síntese baseada em trabalhos publicados entre 2009 e 2019. A partir disso, os autores propuseram um *framework* conceitual que possibilita avaliar a qualidade de OGD. Os autores salientam a importância de materializar está proposta com o intuito de testar o instrumento proposto na qual ainda não foi testado, apenas discutido teoricamente.
- O trabalho de Marmier e Mettler (2020) apresenta uma estrutura que considera os 10 princípios de OGD para avaliar a qualidade dos portais de dados suíços. Desta forma, é possível chegar a um índice através da média das pontuações obtidas por cada recurso da estrutura criada. De modo a avaliar a qualidade dos portais de OGD suíços. Para coletar, compilar e analisar os dados utilizaram o MS Excel "Power Pivot".
- O estudo de Ibáñez *et al.* (2019) propõe questões de dados conectados para avaliar portais de dados europeus. Para isso, eles se aprofundam em questões semânticas (por exemplo, vocabulários e proveniência) e realizam a análise de conteúdo para identificar e quantificar os conjuntos de dados e recursos que dispõem de acesso e fazem uso de tecnologias semânticas.
- No trabalho de Correa *et al.* (2019) foi criado um método automatizado para avaliar dados de OGD, sendo considerado apenas portais de dados abertos que não fazem uso de plataforma de catalogação de dados, por exemplo, CKAN. Proporcionando a avaliação automatizada de portais que utilizam de uma estrutura clássica, por exemplo, formulários.
- O estudo de SILVA *et al.* (2020) apresenta uma metodologia de avaliação focada nos conjuntos de dados e recursos, onde quantificam-se os resultados. E a coleta

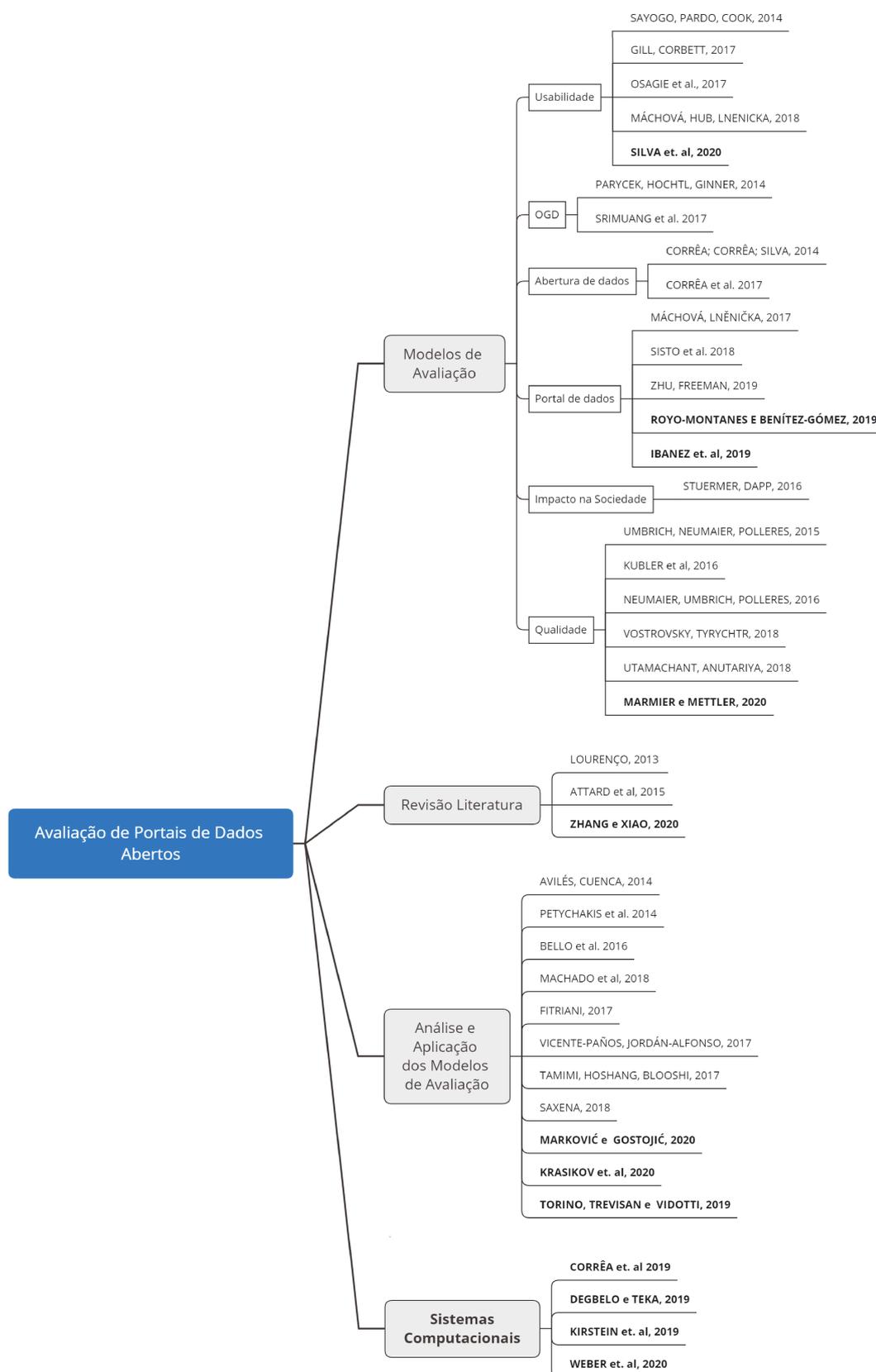
é realizada de modo manual, aplicada sobre uma pequena fração de conjuntos de dados presentes no portal de dados abertos do Brasil.

Observa-se que novas pesquisas sobre avaliação de dados, conjuntos de dados e portais vêm sendo desenvolvidas, bem como propostas para automatização do processo de avaliação. Percebe-se que alguns destes estudos encontram-se em fase inicial e como consequência não possuem uma estrutura de avaliação bem definida e clara, em sua maioria são avaliações focadas em um nível de governo ou mesmo proposições teóricas sem aplicabilidade comprovada.

Entende-se a importância dos estudos e estes contribuem com a evolução do tema. Contudo, nota-se a necessidade de iniciativas práticas e aplicáveis em cenários do cotidiano, fazendo que organizações tenham meios concretos, seguros e sustentados pela ciência para auxiliar na avaliação, considerando o âmbito de dados abertos.

O complemento dos trabalhos relacionados não foram considerados na composição do modelo de referência. No entanto, abordou-se este contexto para justificar a relevância do tema, bem como demonstrar a possibilidade de ampliação do estudo realizado. A Figura 12 apresenta os possíveis trabalhos que podem ser incorporados a estrutura que será apresentada na próxima seção. Destaca-se que a realização desta atualização demandará uma nova análise e conseqüentemente uma reestruturação do que está sendo proposto.

Figura 12 – Complementação da classificação dos artigos encontrados na literatura



Fonte: do Autor

Na Figura 12 observa-se em negrito os trabalhos recentes encontrados na literatura (SILVA *et al.*, 2020; MARKOVIĆ; GOSTOJIĆ, 2020; ZHANG; XIAO, 2020; MARMIER; METTLER, 2020; KRASIKOV *et al.*, 2020; ROYO-MONTANES; BENITEZ-GOMEZ, 2019; IBÁÑEZ *et al.*, 2019). Destaca-se a criação de um quarto grupo denominado **Sistemas Computacionais**, sendo que os trabalhos (WEBER *et al.*, 2020; KIRSTEIN *et al.*, 2019; DEGBELO; TEKA, 2019; CORREA *et al.*, 2019) relacionados a este grupo focam em ferramentas e métodos que contribuem na automatização do processo de coleta e avaliação de portais de dados abertos.

Diante do resumo das limitações dos estudos apresentados na seção 2.10.1, utilizaram-se os dois principais conceitos que contribuem na avaliação da estrutura do portal, de conjuntos de dados e recursos. Além dos conceitos elencados, foram consideradas as boas práticas de dados na Web, conforme seção 2.8. Deste modo, foi possível elaborar uma estrutura de referência que tem por objetivo avaliar holisticamente os portais de dados abertos.

## 2.11 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os fundamentos acerca de dados abertos e seus benefícios, sendo possível vislumbrar as diferentes áreas que utilizam dados abertos evidenciando a sua importância. Também foram apresentados conceitos sobre dados conectados, dados abertos e dados na Web, respaldando a intersecção entre estes conceitos e os definindo. Foram abordadas questões que envolvem a padronização de dados e as boas práticas de dados na Web, estas desenvolvidas a fim de proporcionar benefícios através do uso e reuso de dados abertos.

Explicitaram-se as plataformas utilizadas por organizações públicas e privadas para disponibilizar e divulgar dados abertos, sendo possível observar as diferenças entre estas plataformas.

Em relação aos modelos de avaliação apresentados, foi possível vislumbrar a classificação dos estudos, desde proposição de modelos de avaliação, revisões de literatura e a aplicação ou análise de modelos de avaliação. Outro ponto que pode ser observado foi a relação entre os estudos e o que se propõe com relação a esta tese. Por fim, a complementação dos trabalhos relacionados sobre modelos de avaliação de portais de dados abertos foi evidenciada, com o objetivo de mostrar os novos estudos, possibilitando a realização de trabalhos futuros e demonstrando a relevância do trabalho proposto.

Deste modo, os conceitos abordados nesta seção fornecem um embasamento científico para o modelo que será descrito na próxima seção.

### 3 O MODELO DE REFERÊNCIA PARA AVALIAÇÃO DE PORTAIS DE DADOS ABERTOS

Dado o contexto, percebeu-se que existem recomendações acerca de padrões de dados. No entanto, se faz necessário verificar se os padrões para dados abertos estão sendo utilizados, visto que o uso de padrões viabilizam a compreensão, facilidade de processamento, a facilidade de descoberta, a confiança, a capacidade de conexão, a facilidade de acesso e a interoperabilidade (LÓSCIO, B. F. *et al.*, 2017). Assim nota-se que ao avaliar a conformidade dos padrões de dados abertos disponibilizados potencializa-se o seu uso e reuso, dentre outros benefícios que podem ser alcançados.

Considerando a premissa de avaliar portais de dados abertos e observando as estruturas presentes na literatura que viabilizam a avaliação, nota-se que as estruturas são subjetivas e específicas. Deste modo, salienta-se a necessidade de compor um modelo de referência que proporcione um alto nível de detalhamento, independente do tamanho do portal de dados e da organização.

Sendo que os resultados obtidos por meio da avaliação devem auxiliar as organizações a melhorarem seus portais e identificarem inconsistência nos padrões adotados para dados abertos, visto que a não utilização de padrões podem até impossibilitar o uso dos dados.

Portanto, o Modelo de Referência de Avaliação de Portais dos Dados Abertos (MRAPDA) foi pensado com objetivo de possibilitar a avaliação dos recursos, conjuntos e portais de dados abertos, focando em padrões e boas práticas encontrados na literatura. Para isso, na continuação será tratado como se dá a composição do modelo.

#### 3.1 ORGANIZAÇÃO DOS CONCEITOS DA LITERATURA

Após a investigação da literatura, foram encontrados alguns conceitos nos trabalhos relacionados à avaliação. Com base nisso, foi realizado o agrupamento por conta das diferentes nomenclaturas e tratativas utilizadas nos estudos encontrados. Para isso, foram definidos dois grupos, o **conceitos frequentes** onde são apresentados os conceitos abordados em mais de um estudo; e o grupo, **conceitos eventuais**, onde são expostos os conceitos específicos de cada estudo. A seguir o Quadro 9 apresenta o grupo de conceitos que são frequentemente encontrados na literatura.

Quadro 9 – Grupo de conceitos frequentes na literatura

<b>Conceitos</b>	<b>Descrição</b>	<b>Referências</b>
Formato de dados	Verifica-se a existência dos diferentes tipos de formatos de dados.	Andreiwid Sh Corrêa <i>et al.</i> (2017), Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018), Xiaohua Zhu e Freeman (2019), ODM (2019) e Abella <i>et al.</i> (2019)
Dados legíveis por máquinas	Verifica-se a existência dos formatos de dados legíveis por máquina (XML, JSON, CSV, RDF e outros).	Andreiwid Sh Corrêa <i>et al.</i> (2017), Xiaohua Zhu e Freeman (2019), ODM (2019) e Abella <i>et al.</i> (2019)
Dados atualizados	Verifica-se se os conjuntos de dados são atualizados.	Andreiwid Sh Corrêa <i>et al.</i> (2017), Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018), Xiaohua Zhu e Freeman (2019), ODM (2019) e Abella <i>et al.</i> (2019)
Informações de publicadores	Verifica-se a existência de informações sobre os responsáveis pela publicação dos dados (e-mail, nome, organização, telefone, entre outros).	Andreiwid Sh Corrêa <i>et al.</i> (2017), Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018), Xiaohua Zhu e Freeman (2019) e ODM (2019)
Licença	Verifica-se a existência de licenças e se estas são abertas.	Andreiwid Sh Corrêa <i>et al.</i> (2017), Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018), Xiaohua Zhu e Freeman (2019), ODM (2019) e Abella <i>et al.</i> (2019)
Pesquisa	Verifica-se a existência da opção de pesquisa no portal.	Andreiwid Sh Corrêa <i>et al.</i> (2017), Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018) e Xiaohua Zhu e Freeman (2019)
Filtros	Verifica-se a existência de filtros que facilitem a busca e compreensão dos conjuntos de dados.	Máchová e Lnénicka (2017) e Sisto <i>et al.</i> (2018)

Disponibilização de API	Verifica-se a existência da disponibilização do acesso via API.	Andreiwid Sh Corrêa <i>et al.</i> (2017), Máchová e Lnénicka (2017), Xiaohua Zhu e Freeman (2019) e Abella <i>et al.</i> (2019)
Categoria de conjunto de dados	Verifica-se a existência de categorias (grupos) de modo a facilitar a busca por conjuntos de dados.	Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018) e Xiaohua Zhu e Freeman (2019)
Solicitação de conjunto de dados	Verifica-se se existem opções onde os usuários possam solicitar dados ou conjuntos de dados para os publicadores.	Máchová e Lnénicka (2017) e Sisto <i>et al.</i> (2018)
Fórum, comentários e avaliações.	Verifica-se a existência de fórum, espaços para mensagens e/ou avaliação dos conjuntos de dados.	Máchová e Lnénicka (2017) e Sisto <i>et al.</i> (2018)
URI permanente	Verifica-se se as URIs e/ou URLs são mantidas, assim garantem o acesso pelos usuários ao longo do tempo.	Máchová e Lnénicka (2017), Xiaohua Zhu e Freeman (2019) e Abella <i>et al.</i> (2019)
Visualização dos conjuntos de dados e/ou recursos interativos	Verifica-se a existência de meios que facilitem a visualização dos conjuntos de dados on-line.	Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018) e Xiaohua Zhu e Freeman (2019)
Número de visitas dos conjuntos de dados	Verifica-se a existência da quantidade total de visitas de um conjunto de dados.	Máchová e Lnénicka (2017) e Xiaohua Zhu e Freeman (2019)
Número de download	Verifica-se a existência da quantidade total de downloads de um conjunto de dados.	Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018), Xiaohua Zhu e Freeman (2019) e ODM (2019)
Documentação de apoio aos dados	Verifica-se a existência de documentação para apoio a compreensão dos conjuntos de dados e seus recursos ou mesmo questões estruturais dos dados, como exemplo, dicionário de dados.	Andreiwid Sh Corrêa <i>et al.</i> (2017) e Xiaohua Zhu e Freeman (2019)

Georreferenciamento dos conjuntos de dados	Verifica-se a existência do georreferenciamento nos conjuntos de dados.	Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018), Xiaohua Zhu e Freeman (2019) e Abella <i>et al.</i> (2019)
Acesso Gratuito	Verifica-se se as plataformas cobram algum valor para liberar o acesso aos conjuntos de dados.	Máchová e Lnénicka (2017) e Sisto <i>et al.</i> (2018)
Perguntas Frequentes (FAQ)	Verifica-se a existência de perguntas mais frequentes.	Máchová e Lnénicka (2017) e Xiaohua Zhu e Freeman (2019)
Número de aplicativos criados	Verifica-se a existência de um número total de aplicativos que foram desenvolvidos utilizando os dados abertos publicados.	Máchová e Lnénicka (2017) e Xiaohua Zhu e Freeman (2019)
Mídias Sociais	Verifica-se a existência da possibilidade de compartilhamento dos dados em redes sociais.	Máchová e Lnénicka (2017) e Xiaohua Zhu e Freeman (2019)
Multi-Idioma	Verifica-se a disponibilização de outros idiomas na plataforma.	Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018) e Xiaohua Zhu e Freeman (2019)
Apresentação do portal	Verifica-se a existência de um espaço de apresentação do portal, das publicações, de qual estrutura é adotada para a publicação dos dados por parte da organização ou das parceiras, qual a periodicidade da publicação, quem são os responsáveis pelo gerenciamento da plataforma, dentre outras informações.	Máchová e Lnénicka (2017), Sisto <i>et al.</i> (2018) e Xiaohua Zhu e Freeman (2019)

Fonte: do Autor

Constata-se que existem vários conceitos que transpassam os trabalhos, alguns com foco nos conjuntos de dados, nos recursos e outros com foco em aspectos relacionados à estrutura do portal de dados. No que lhe concerne, o Quadro 10 apresenta

os conceitos que aparecem de forma exclusiva em cada estudo, sendo este o grupo de conceitos eventuais encontrados na literatura.

Quadro 10 – Grupo de conceitos eventuais na literatura

<b>Conceitos</b>	<b>Descrição</b>	<b>Referência</b>
Acessibilidade	Avalia a disponibilização de recursos de acessibilidade para pessoas com deficiência.	Andreiwid Sh Corrêa <i>et al.</i> (2017)
Título e Descrição	Verifica-se a existência dos títulos e das descrições dos conjuntos de dados.	Máchová e Lnénicka (2017)
Palavras-Chave	Verifica-se a existência de palavras-chave nos conjuntos de dados.	Máchová e Lnénicka (2017)
Conta de usuários	Verifica-se se as plataformas possibilitam a criação de contas de usuário, possibilitando acompanhar as atualizações, notificações e outros.	Máchová e Lnénicka (2017)
Tamanho do conjunto de dados	Verifica-se a existência do tamanho total em Kilobyte, Megabyte ou Gigabyte dos recursos disponibilizados.	Máchová e Lnénicka (2017)
Consultas SPARQL	Verifica-se se o portal disponibiliza consultas aos dados em formato semântico através do SPARQL.	Abella <i>et al.</i> (2019)
Vocabulários	Verifica-se se os conjuntos de dados dispõem de vocabulários.	Abella <i>et al.</i> (2019)
Fonte Oficial	Verifica-se se os dados publicados são da fonte oficial (responsável designado, organização, dentre outros).	Sisto <i>et al.</i> (2018)
Site parceiros	Verifica-se se os dados publicados por um portal de um parceiro constam no portal de origem.	Sisto <i>et al.</i> (2018)

Painel On-line	Verifica-se a existência de um painel on-line onde são exibidos todos os conjuntos de dados.	Sisto <i>et al.</i> (2018)
Espaço para divulgação	Verifica-se a existência de um espaço no portal onde são apresentados os aplicativos desenvolvidos, projetos oriundos dos dados e eventos que visam o fomento do uso dos dados.	Sisto <i>et al.</i> (2018)
Disponibilidade dos conjuntos de dados	Verifica-se a existência do número total de conjuntos de dados disponibilizados no portal.	ODM (2019)
Descoberta	Verifica-se a facilidade de encontrar o portal, conjuntos de dados ou recursos na internet.	ODM (2019)
Pontuação geral sobre qualidade	Verifica-se a média sobre a existência dos aspectos mais importantes sobre os dados (Licença aberta, legibilidade por máquina, disponibilidade e a integridade dos metadados).	ODM (2019)
Total geral dos recursos	Verifica-se a existência do número total dos recursos presentes nos conjuntos de dados.	ODM (2019)
Número de distribuição	Verifica-se a quantidade em Megabytes distribuídos no portal.	ODM (2019)
Número total de editores	Verifica-se a existência do número total dos editores que publicam em portais de dados.	ODM (2019)
Número total de catálogos coletados	Verifica-se a existência do número total dos catálogos coletados pela plataforma.	ODM (2019)

Granularidade dos dados	Verifica-se a existência de um detalhamento maior sobre os dados publicados.	Xiaohua Zhu e Freeman (2019)
Metadados	Verifica-se se existem metadados vinculados aos conjuntos de dados.	Xiaohua Zhu e Freeman (2019)
Disponibilizar formato de citação	Verifica-se a existência da opção de citação sobre os dados publicados.	Xiaohua Zhu e Freeman (2019)
Personalizar portal	Verifica-se a existência de personalização do portal para o usuário.	Xiaohua Zhu e Freeman (2019)
Permitir comparação	Verifica-se a existência da opção de comparação entre portais de dados e conjuntos de dados.	Xiaohua Zhu e Freeman (2019)
Recomendação de conjunto de dados	Verifica-se a existência de recomendações de conjunto de dados.	Xiaohua Zhu e Freeman (2019)
Download	Verifica-se se os conjuntos de dados estão disponíveis, se é possível fazer download em lote ou se é possível efetuar a impressão.	Xiaohua Zhu e Freeman (2019)

Fonte: do Autor

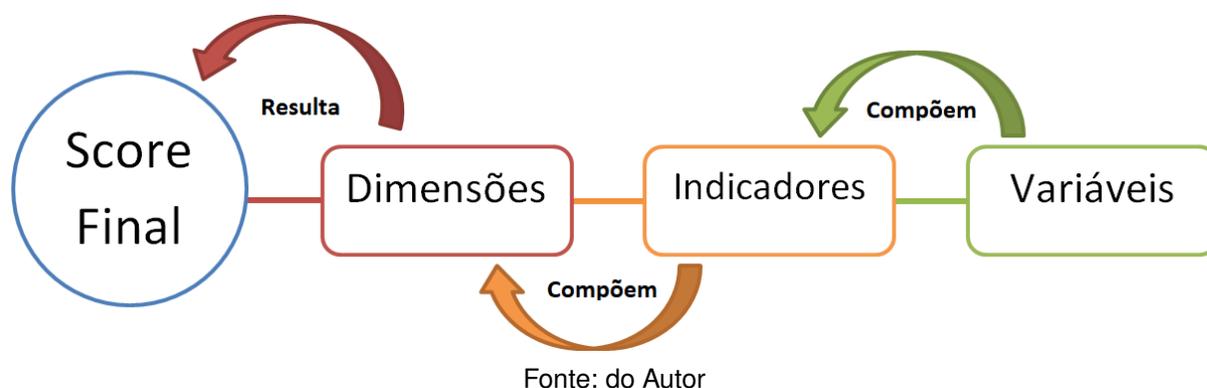
Além dos estudos identificados na literatura, também foram utilizadas as boas práticas recomendadas pela W3C conforme seção 2.8. A sintetização dos conceitos identificados na literatura com as recomendações da W3C tornam possível a composição de um modelo robusto, onde envolve aspectos apontados por estudos científicos juntamente com recomendações técnicas. As próximas seções irão apresentar quais destes conceitos da literatura e quais boas práticas são utilizadas no modelo.

### 3.2 ESTRUTURA DO MODELO

Mencionado por Corbetta (2003), a construção dos indicadores é realizada baseando-se em quatro passos, os quais foram apresentados na seção 1.8.2. No entanto, esses passos não necessariamente precisam ser realizados na sequência.

Iniciou-se a estruturação do modelo identificando os indicadores. Seguindo para a identificação das dimensões, de modo a, organizar e agrupar os indicadores e suas variáveis. A Figura 13 representa a hierarquia da estrutura do modelo.

Figura 13 – A estrutura do modelo



Vê-se na Figura 13 o score final (índice geral) que é o resultado obtido através da composição das dimensões, sendo essas compostas pelos indicadores e esses compostos pelas variáveis.

A cunho de explicação, primeiramente apresentam-se os indicadores, seguido das dimensões e por fim as variáveis. Isso ocorre, para explicitar de forma coerente as variáveis, visto que estas dependem dos indicadores e das dimensões.

### 3.2.1 Indicadores

Os indicadores foram definidos baseados nos conceitos oriundos da literatura e nas boas práticas de dados na Web. O Quadro 11 expõe os indicadores que são utilizados no modelo de referência, evidenciando sua origem.

Quadro 11 – Concepção dos indicadores do modelo de referência

<b>Indicadores</b>	<b>Conceitos frequentes na literatura</b>	<b>Conceitos eventuais na literatura</b>	<b>Presente nas Boas Práticas - W3C</b>
Atualização	Dados Atualizados	—	Dados Atualizados - (BP 21)
Download em Massa	—	Download	Fornecer download em massa (bulk download) - (BP 17); Fornecer subconjuntos para conjuntos de dados extensos - (BP 18).

Metadados	—	Metadados	Fornecer metadados – (BP 01).
Metadados Descritivos	Informação dos publicadores; Categoria de conjunto de dados	Título e Descrição; Tamanho dos dados; Palavras-Chaves	Fornecer metadados descritivos (BP 02).
Metadados Estrutural	Documentação de apoio aos dados	—	Fornecer metadados estruturais (BP 03)
Versionamento	—	—	Fornecer indicador de versão – (BP 7); Fornecer o histórico de versão (BP 08)
URI	URI	—	Usar URIs persistentes como identificadores de conjuntos de dados (BP 9); Usar URIs persistentes como identificadores dentro de conjuntos de dados (BP 10); Atribuir URIs para as versões dos conjuntos de dados e séries (BP 11).
Formatos	Formato de Dados; Dados Legíveis por Máquinas	—	Fornecer dados em formatos múltiplos - (BP 14); Formatos de dados padronizados legíveis por máquinas - (BP 12)
Licenças	Licenças	—	Fornecer informações sobre a licença de dados – (BP 04).
Proveniência	—	Fonte Oficial; Site Parceiros;	Fornecer informações de procedência dos dados – (BP 05).
Vocabulários	—	Vocabulários	Reutilizar vocabulários, dando preferência aos padronizados – (BP 15);
Pesquisa	Filtro; Pesquisa	—	—

Conta de usuário	—	Conta de usuários	—
Recomendação	—	Recomendação de conjuntos de dados	—
Acessibilidade	—	Acessibilidade	—
Multi-idioma	Multi-idioma	—	—
Visualização de dados	Visualização dos conjuntos de dados; Georreferenciamento	—	Fornecer visualizações complementares – (BP 32)
Temática	Categoria de conjunto de dados	—	Fornecer metadados descritivos (BP 02)
API   SPARQL	Disponibilidade API	Disponibilidade SPARQL	Disponibilizar dados por meio de uma API – (BP 23); Usar padrões Web como base para construção de APIs - (BP 24); Fornecer documentação completa para as APIs – (BP 25).
Interatividade	Solicitação, conjuntos de dados; Fórum; Comentários; avaliações; Mídias Sociais;	Formatos de citação	Coletar feedback de consumidores de dados – (BP 29)   Compartilhar o feedback disponível – (BP 30)
Compreensão do portal	Apresentação portal; Perguntas Frequentes; Quantidade de aplicativos criados	Espaço divulgação	—
Quantitativos	Quantidade de visitas; Quantidade de download; Quantidade de APP criados	Número total de recursos; Numero de distribuição; Número de Editores; Granularidade dos Dados	—

Fonte: do Autor

Para a definição dos indicadores foram consideradas as variáveis que permitissem a sua operacionalização, ou sejam que possibilitem fazer uso através de um sistema computacional. Por conta disto, alguns indicadores apresentados nos Quadros 9 e 10, bem como algumas das boas práticas não foram utilizados. No Quadro 12 são apresentados os indicadores não utilizados e a justificativa para estes.

Quadro 12 – Conceitos da literatura não utilizados

<b>Conceitos não utilizados</b>	<b>Referência</b>	<b>Justificativa</b>
Pontuação geral sobre a qualidade	ODM (2019)	Não utilizado por ser a média de outros indicadores relacionados ao estudo de origem. Neste sentido, o modelo proposto, utiliza a ponderação das variáveis, indicadores e dimensões para originar o score final, conseqüentemente a média não será abordada nesta estrutura.
Descoberta	ODM (2019)	Não foi utilizado, pois o foco deste elemento é a localização através de mecanismos de busca (Google, Bing, etc). Pelo fato de utilizar uma plataforma, esta já possui por padrão uma configuração que auxilia na localização por estes mecanismos de buscas a Web.
Personalização do portal	Xiaohua Zhu e Freeman (2019)	Não foi considerado pelo fato de que a personalização está focada no usuário e questões de Interação Humano-Computador (IHC).
Permitir comparação	Xiaohua Zhu e Freeman (2019)	Não foi utilizado, pois a comparação seria entre portais e seus conjuntos de dados. Sendo assim, está fora do escopo de pesquisa.
Acesso Gratuito	Máchová e Lnénicka (2017) e Sisto <i>et al.</i> (2018)	Não foi utilizado, pois o objetivo deste elemento era verificar se a plataforma realiza alguma cobrança para acesso aos dados e como este modelo foi construído com a delimitação da plataforma CKAN, esta não tem a opção de cobrança de acesso.

Painel On-line	Sisto <i>et al.</i> (2018)	Não foi utilizado, pois o objetivo deste elemento era verificar se existe uma interface (painel online) que permite aos <i>stakeholders</i> visualizarem os conjuntos de dados publicados. Como o modelo foi construído com uma perspectiva de plataforma, esta já possui uma estrutura própria que atende por padrão este elemento.
Usar representações de dados que sejam independentes de localidade ( <i>locale neutral</i> )	Boa Prática 13 - W3C	Não usado, pois para esta verificação é necessária à averiguação do conteúdo, saindo do escopo de pesquisa.
Escolher o nível de formalização adequado	Boa prática 16 - W3C	Não utilizado por ser uma recomendação de análise subjetiva, dificultando a mensuração desta boa prática.
Usar negociação de conteúdo para disponibilizar dados em formatos múltiplos	Boa prática 19 - W3C	Não foi utilizado por conta de dificuldade de mensurar esta boa prática.
Fornecer uma explicação para os dados que não estão disponíveis	Boa prática 22 - W3C	Não utilizado por conta da dificuldade para a mensuração e para avaliação desta boa prática. Entende-se que esta boa prática seria uma variável descritiva, dificultando a avaliação no portal.
Evitar alterações que afetem o funcionamento de sua API	Boa prática 26 - W3C	Não utilizado por ser uma recomendação com foco no desenvolvimento/ implementação da plataforma, não sendo aplicável ao modelo.
Preservar identificador	Boa prática 27 - W3C	Não utilizado pela complexidade dada para a avaliação dos portais de dados, visto que seria necessário um perfil técnico para compreensão dos códigos de respostas informados pelo servidor.
Avaliar a cobertura do conjunto de dados	Boa prática 28 - W3C	Não usado, pois para a verificação é necessária averiguação de conteúdo, saindo do escopo pretendido de pesquisa.

Enriquecer dados por meio da geração de novos dados	Boa prática 31 - W3C	Não usado, pois para verificação é necessária averiguação de conteúdo, saindo do escopo pretendido de pesquisa.
Fornecer feedback para o publicador original	Boa prática 33 - W3C	Não usado por ser uma recomendação para quem utilizar os conjuntos de dados.
Obedecer aos termos de licença	Boa prática 34 - W3C	Não usado, pois esta recomendação sugere a leitura da licença na íntegra e verificar se o conjunto de dados estão adequados com as normativas da licença. Desta forma, tornaria o processo de avaliação complexo, além da dificuldade de mensurar.
Citar a publicação original do conjunto de dados	Boa prática 35 - W3C	Não usado por ser uma recomendação para quem utilizar os conjuntos de dados.

Fonte: do Autor

### 3.2.2 Dimensões

Com a estruturação dos indicadores, foi possível definir as dimensões que auxiliam a segmentação do modelo e sua compreensão. Estas dimensões foram definidas com base no objetivo dos indicadores. Deste modo, foram segmentadas em seis dimensões, sendo estas: **I) Dimensão Acesso**, **II) Dimensão Formatos**, **III) Dimensão Direitos**, **IV) Dimensão Semântica**, **V) Dimensão Estrutura** e **VI) Dimensão Apresentação**. Todas essas dimensões foram criadas com base na literatura; o Quadro 13 apresenta as dimensões e os estudos que as utilizam ou fazem referência a indicadores que fazem referência às dimensões definidas.

Quadro 13 – Definição das dimensões

Dimensão	Referência
Acesso	Bernadette Farias Lóscio <i>et al.</i> (2017), Andreiwid Sh Corrêa <i>et al.</i> (2017), Máchová e Lnénicka (2017), Xiaohua Zhu e Freeman (2019) e Abella <i>et al.</i> (2019)
Formatos	Bernadette Farias Lóscio <i>et al.</i> (2017), Andreiwid Sh Corrêa <i>et al.</i> (2017), Abella <i>et al.</i> (2019) e ODM (2019)

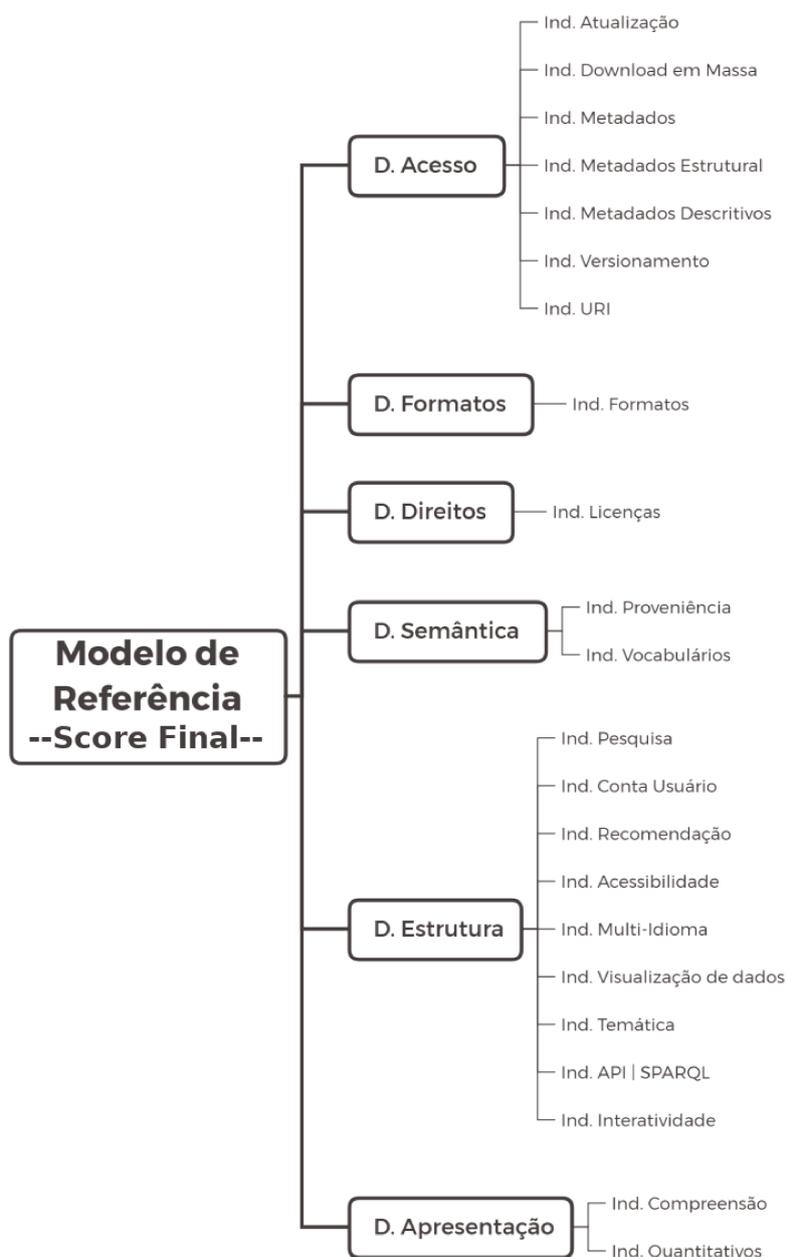
Direitos	Bernadette Farias Lóscio <i>et al.</i> (2017), Máchová e Lnénicka (2017), Xiaohua Zhu e Freeman (2019), Abella <i>et al.</i> (2019) e ODM (2019)
Semântica	Bernadette Farias Lóscio <i>et al.</i> (2017) e Abella <i>et al.</i> (2019)
Estrutura	Sisto <i>et al.</i> (2018), Máchová e Lnénicka (2017) e Xiaohua Zhu e Freeman (2019)
Apresentação	Sisto <i>et al.</i> (2018), Máchová e Lnénicka (2017) e Xiaohua Zhu e Freeman (2019)

Fonte: do Autor

Com a definição das dimensões e dos indicadores, têm-se parte do modelo de referência estruturado, a Figura 14 ilustra a estrutura de dimensão e indicadores.

Percebe-se que o modelo conta com 6 dimensões e 22 indicadores. Seguindo a metodologia de construção de indicadores, foram criadas as variáveis que possibilitam a composição dos indicadores. Na continuidade apresenta-se a síntese das variáveis com seus respectivos indicadores.

Figura 14 – Estrutura de Dimensão e indicadores do modelo



Fonte: do Autor

### 3.2.3 Variáveis

Esta seção mostra as variáveis e seus tipos. Para mais detalhes sobre as definições e recomendações que originou cada variável pode-se consultar o (Apêndice B).

Os Quadros 14, 15, 16, 17, 18 e 19 apresentam em síntese as dimensões, seus respectivos indicadores e suas variáveis, sendo possível observar a estrutura completa de cada dimensão. Algumas dimensões possuem um único indicador pois estes não

foram possíveis de serem agrupados em outra dimensão sendo necessário ter uma própria.

Quadro 14 – Dimensão Acesso com seus indicadores e variáveis

<b>Dimensão Acesso</b>			
<b>Indicador</b>	<b>V nº</b>	<b>Variável</b>	<b>Tipo Variável</b>
<b>Atualização</b>	V01	Atualização nos últimos 12 meses - Conjuntos de Dados	Inteiro
	V02	Atualização nos últimos 12 meses - Recursos	Inteiro
	V03	Adicionados nos últimos 12 meses - Conjuntos de dados	Inteiro
	V04	Adicionados nos últimos 12 meses - Recursos	Inteiro
<b>Download em Massa</b>	V05	Recursos disponíveis - API ou FTP	Inteiro
	V06	Recursos compactados disponibilizados no portal	Inteiro
<b>Metadados</b>	V07	Conjuntos de Dados - Metadados (Formato HTML)	Inteiro
	V08	Recursos - Metadados (Formato HTML)	Inteiro
	V09	Conjuntos de Dados - Metadados (Formato serializado)	Inteiro
	V10	Recursos - Metadados (Formato serializado)	Inteiro
<b>Metadados Descritivos</b>	V11	Conjuntos de Dados - Metadados (Nome)	Inteiro
	V12	Recursos - Metadados (Nome)	Inteiro
	V13	Conjuntos de Dados - Metadados (Descrição)	Inteiro
	V14	Recursos - Metadados (Descrição)	Inteiro
	V15	Conjuntos de Dados - Metadados (Data Criação)	Inteiro
	V16	Recursos - Metadados (Data Criação)	Inteiro
	V17	Metadados (Mantenedor)	Inteiro
	V18	Metadados (Mantenedor e-mail)	Inteiro
	V19	Metadados (Autor)	Inteiro
	V20	Metadados (Autor e-mail)	Inteiro
	V21	Metadados (Palavras-Chave)	Inteiro
	V22	Metadados - Formato de arquivo	Inteiro
	V23	Metadados - Tamanho de arquivo	Inteiro
<b>Metadados Estruturais</b>	V24	Metadados estruturais (Formato HTML)	Inteiro
	V25	Metadados estruturais (Formato serializado)	Inteiro
	V26	Metadados estruturais (Documentos separados)	Inteiro

<b>Versionamento</b>	V27	Conjuntos de dados com controle de versão ou com data na URI	Inteiro
	V28	Conjuntos de dados com histórico de versão	Binária
<b>URI</b>	V29	Conjunto de Dados - URI encontrado ou redirecionado	Inteiro
	V30	Recursos - URI encontrado ou redirecionado	Inteiro

Fonte: do Autor

Quadro 15 – Dimensão Formato com seus indicadores e variáveis

<b>Dimensão Formato</b>			
<b>Indicador</b>	<b>V nº</b>	<b>Variável</b>	<b>Tipo Variável</b>
<b>Formatos</b>	V31	Cobertura de formatos de dados	Float (núm. Reais)
	V32	Arquivos (Recursos) - Formato aberto	Inteiro
	V33	Arquivos (Recursos) - Formato legíveis por máquina	Inteiro
	V34	Formatos com URIs condizente ao formato cadastrado	Inteiro

Fonte: do Autor

Quadro 16 – Dimensão Direitos com seus indicadores e variáveis

<b>Dimensão Direitos</b>			
<b>Indicador</b>	<b>V nº</b>	<b>Variável</b>	<b>Tipo Variável</b>
<b>Licenças</b>	V35	Licença Aberta	Inteiro
	V36	Licença C/ Restrição	Inteiro
	V37	Lic. Específica	Inteiro
	V38	Licença Fechada	Inteiro

Fonte: do Autor

Quadro 17 – Dimensão Semântica com seus indicadores e variáveis

<b>Dimensão Semântica</b>			
<b>Indicador</b>	<b>V nº</b>	<b>Variável</b>	<b>Tipo Variável</b>
<b>Proveniência</b>	V39	Conjuntos de dados que possuem vínculo com uma organização	Inteiro
	V40	Conjuntos de dados que possuem informações da fonte oficial dos dados	Inteiro
<b>Vocabulários</b>	V41	Vocabulários próprios (SIM/NÃO)	Binária
	V42	Vocabulários nos conjuntos de dados	Inteiro
	V43	Vocabulário Licenças	Inteiro
	V44	Vocabulário Proveniência	Inteiro
	V45	Vocabulários Qualidade	Inteiro
	V46	Vocabulário Versionamento	Inteiro
	V47	Vocabulário Histórico de Versão	Inteiro
	V48	Vocabulário Atualização	Inteiro
	V49	Vocabulário Feedback Usuário	Inteiro

Fonte: do Autor

Quadro 18 – Dimensão Estrutura com seus indicadores e variáveis

<b>Dimensão Estrutura</b>			
<b>Indicador</b>	<b>V nº</b>	<b>Variável</b>	<b>Tipo Variável</b>
<b>Pesquisa</b>	V50	Barra de pesquisa disponível	Binária
	V51	Refinamento disponível	Binária
<b>Conta Usuário</b>	V52	Disponibilidade de cadastro de usuário	Binária
	V53	Disponibilidade de acesso de usuário	Binária
<b>Recomendação</b>	V54	Sugestões na barra de pesquisa	Binária
	V55	Recomendação de conjunto de dados semelhantes	Binária
<b>Acessibilidade</b>	V56	Nota WCAG 2.1	Float (núm. Reais)
<b>Multi-Idioma</b>	V57	Disponibilidade em outros idiomas além do nativo	Binária
<b>Visualização</b>	V58	Visualizador de conteúdo HTML	Binária
	V59	Visualizador de imagens	Binária
	V60	Visualizador de texto	Binária
	V61	Visualizador com uso de mapas	Binária
	V62	Visualizador de gráficos	Binária
	V63	Visualizador de PDF	Binária
	V64	Visualizador de arquivos JSON	Binária

<b>Temática</b>	V65	Conjunto de dados com pelo menos um grupo/tema vinculado	Inteiro
<b>API SPARQL</b>	V66	Acesso via API	Binária
	V67	Acesso via SPARQL	Binária
	V68	Informações sobre API	Binária
<b>Interatividade</b>	V69	Formulários ou informações sobre a solicitação de conjuntos de dados	Binária
	V70	Disponibilidade de citações	Binária
	V71	Disponibilidade de fórum	Binária
	V72	Disponibilidade de comentário	Binária
	V73	Disponibilidade de avaliação	Binária
	V74	Compartilhamento no Facebook	Binária
	V75	Compartilhamento no Google	Binária
	V76	Compartilhamento no Twitter	Binária

Fonte: do Autor

Quadro 19 – Dimensão Apresentação com seus indicadores e variáveis

<b>Dimensão Apresentação</b>			
<b>Indicador</b>	<b>V nº</b>	<b>Variável</b>	<b>Tipo Variável</b>
<b>Compreensão</b>	V77	Introdução sobre o portal de dados	Binária
	V78	Disponibilidade de perguntas mais frequentes (FAQ)	Binária
	V79	Disponibilização de projetos	Binária
<b>Quantitativos</b>	V80	Quadro com o resumo das atividades do portal	Binária
	V81	Disponibiliza número de downloads	Binária
	V82	Disponibiliza total de conjunto de dados	Binária
	V83	Disponibiliza número de projetos desenvolvidos	Binária
	V84	Disponibiliza total de recursos	Binária
	V85	Disponibiliza total em MB de dados	Binária
	V86	Disponibiliza total de editores	Binária
	V87	Disponibiliza total de catálogos coletados	Binária
	V88	Disponibiliza número de visitas	Binária

Fonte: do Autor

Na sequência são descritas as etapas do processo de avaliação. Assim, mostra-se e discute-se como é realizada a coleta dos dados, a lógica de avaliação, os cálculos utilizados e como é dada a interpretação da análise dos resultados.

### 3.3 PROCESSO DE AVALIAÇÃO UTILIZANDO DO MODELO

Esta seção tem por objetivo esclarecer como serão realizados os cálculos para obter o score final e como será o processo de avaliação utilizando o modelo. Para iniciar a avaliação é necessário realizar a coleta de dados dos portais, podendo ser realizada de duas maneiras, a **i)** de forma manual verificando o que cada variável solicita e ir preenchendo a planilha<sup>1</sup> conforme cada tipo de variável; e **ii)** de modo automatizado, neste caso poderá ser utilizado SiAPDA, este sistema foi desenvolvido para auxiliar na coleta, na aplicação do modelo proposto, bem como para o auxílio na avaliação e análise dos resultados, sendo este apresentado na seção 4.1.

Devido ao modelo poder ser aplicado de modo manual, serão esclarecidos os cálculos para a realização da avaliação. Destaca-se que o SiAPDA realiza estes mesmos cálculos de maneira automática. A Figura 15 apresenta o processo de avaliação e a seguir serão abordadas cada uma das etapas.

Figura 15 – Processo de avaliação utilizando do modelo



Fonte: do Autor

#### 3.3.1 Coletar

Nesta etapa é realizada a coleta para o preenchimento das variáveis, podendo ser feita de modo manual ou utilizando sistema. Vale salientar que caso seja realizada a coleta manual, é necessário que todas as variáveis sejam preenchidas. Caso não sejam encontradas respostas em relação às variáveis não binárias estas devem ter seu valor zerado (0), por sua vez, as variáveis binárias caso não sejam encontradas respostas deverá ter seu valor zerado (0), equivalente a resposta NÃO.

O processo de coleta utilizando o sistema é realizado de maneira simples, basta acessar o menu "Portais->Coletas" será exibida a tela de **coletas**, onde são listados os portais de dados previamente coletados (se houver). Deste forma, é possível refazer uma coleta de modo rápido com apenas um clique, possibilitando manter um histórico de coletas realizadas. Nesta tela também são listadas informações, como o nome do portal, esfera a qual pertence, a URL e a data e hora que foi iniciado o processo de coleta. A Figura 16 ilustra a tela com a listagem de coletas já realizadas.

<sup>1</sup> Planilha disponível em: <<https://1drv.ms/x/s!AgUHQrKlywKtiR7DkjrV8xyoF8ir?e=0ccESf>>

Figura 16 – Tela coleta - SiAPDA

**Coleta do portal de dados**

**SiAPDA informa:**

Esta etapa possibilita coletar as informações sobre os conjuntos de dados através dos metadados cadastrados nos portais e elementos do HTML, ou seja, da interface gráfica dos portais. Este processo é realizado através de robôs ou consultas através de APIs. Por tanto, visando aspectos de segurança, somente o administrador pode inicializar a coleta dos portais de dados abertos.

Para que uma coleta seja realizada, o portal de dados pretendido deve usar a plataforma **CKAN**. Esta restrição é necessária, pois a plataforma fornece meios padronizados para a coleta dos metadados, bem como o uso de TAGs HTML padronizadas. É importante salientar que, dependendo do tamanho do portal, ou seja, quantidade de conjuntos de dados e recursos, este processo de coleta poderá demorar várias horas.

Os usuários classificados como colaborador poderão enviar uma solicitação de coleta aos administradores. Por sua vez, os administradores irão receber um e-mail com a solicitação do usuário, podendo iniciar a nova coleta. Ao final de toda coleta, tanto administrador e colaborador irão receber um e-mail informando o término do processo de coleta.

**+ Nova Coleta**

**Resultados** 10 **Buscar:**

Portal	Esfera	URL Portal	Data de Coleta	Ação
Bem vindo - Alagoas em Dados e Informações	Estadual	https://dados.al.gov.br/catalogo	29/01/2021 - 12:01:15	<b>Coletar Novamente</b>

Fonte: do Autor

Ao iniciar uma nova coleta é solicitada a URL do portal de dados e a qual esfera o portal pertence (Municipal, Estadual ou Federal). A Figura 17 ilustra a tela para realizar a nova coleta.

Figura 17 – Tela nova coleta - SiAPDA

**Coleta do portal de dados**

**SiAPDA informa:**

Esta etapa possibilita coletar as informações sobre os conjuntos de dados através dos metadados cadastrados nos portais e elementos do HTML, ou seja, da interface gráfica dos portais. Este processo é realizado através de robôs ou consultas através de APIs. Por tanto, visando aspectos de segurança, somente o administrador pode inicializar a coleta dos portais de dados abertos.

Para que uma coleta seja realizada, o portal de dados pretendido deve usar a plataforma **CKAN**. Esta restrição é necessária, pois a plataforma fornece meios padronizados para a coleta dos metadados, bem como o uso de TAGs HTML padronizadas. É importante salientar que, dependendo do tamanho do portal, ou seja, quantidade de conjuntos de dados e recursos, este processo de coleta poderá demorar várias horas.

Os usuários classificados como colaborador poderão enviar uma solicitação de coleta aos administradores. Por sua vez, os administradores irão receber um e-mail com a solicitação do usuário, podendo iniciar a nova coleta. Ao final de toda coleta, tanto administrador e colaborador irão receber um e-mail informando o término do processo de coleta.

**Endereço URL do portal**

URL

**Esfera na qual o portal pertence**

Selecione uma esfera

**Coletar** **Cancelar**

Copyright © 2014-2019 AdminLTE.io. All rights reserved. Anything you want

Fonte: do Autor

Após a finalização da coleta é necessário aplicar o modelo, para isso basta acessar o menu "Portais->Aplicar modelo". Na tela aplicar modelo, será exibida a listagem dos portais de dados coletados para que seja selecionado o portal que será aplicado o modelo, a Figura 18 ilustra esta tela.

Figura 18 – Tela aplicar modelo - SiAPDA

SIAPDA - Sistema de Avaliação de Portais de Dados Abertos

Início Portais ▾ Padrões de análise ▾ Configurações Sair

### Aplicar Modelo de Referência

**SiAPDA informa:**

Após a conclusão da etapa de coleta, será possível aplicar o modelo proposto sobre os portais coletados com sucesso. Nesta etapa faz-se uso de robôs. Isso ocorre, pois, algumas variáveis necessitam de novas verificações e coletas, implicando também em uma possível demora (várias horas) para o término do processo. Importante salientar que ao término do processo, todos os administradores serão notificados.

O modelo é composto por seis dimensões (Acesso, Formatos, Direitos, Semântica, Estrutura do portal e Compreensão do portal) e cada dimensão no que lhe concerne possui um conjunto de indicadores e variáveis. Com isso, este modelo possibilitará uma análise tanto dos conjuntos de dados e seus recursos, bem como aspectos relacionados a interface do portal. Deste modo, possibilita identificar se foram seguidas e aplicadas todas as boas práticas durante o processo de publicação dos dados.

Importante salientar que a opção "Aplicar Modelo" está disponível apenas para o administrador.

Resultados: 10

Buscar:

Portal	Esfera	URL Portal	Data de Coleta	Ação
Bem vindo - Alagoas em Dados e Informações	Estadual	<a href="https://dados.al.gov.br/catalogo">https://dados.al.gov.br/catalogo</a>	29/01/2021 - 12:01:15	

Fonte: do Autor

Como pode ser observado o software contribui significativamente com a coleta de dados. A seguir é apresentado como são atribuídos os padrões para as avaliações.

### 3.3.2 Definir Padrões

A definição dos padrões é dada pela aplicação de pesos que variam de acordo com a necessidade, desta forma é possível criar diferentes cenários de avaliação os pesos são atribuídos para cada dimensão, indicador e variável. Além da atribuição dos pesos, é atribuído um **SCORE META**, ou seja, a nota que se almeja ao final da avaliação que pode variar entre zero(0) e dez(10).

Esta etapa é importante, pois possibilita flexibilizar a avaliação, visto que podem ser atribuídos peso zero(0) para dimensões, indicadores ou variáveis que não devem ser consideradas em um determinado cenário.

Em caso de optar em avaliar o portal de dados de modo manual, é salientado que na planilha possui a coluna "PESOS" que deverá ser preenchida.

É importante salientar que existe um padrão definido como **Default**, sendo este a atribuição do peso 100 para todas as variáveis, indicadores e dimensões e neste caso para o score meta é atribuído o valor 5, sendo este a metade do total. A Figura 19 apresentará a lista de padrões criados no sistema.

Figura 19 – Tela lista padrões - SiAPDA

 SiAPDA - Sistema de Avaliação de Portais de Dados Abertos

Início Portais ▾ Padrões de análise ▾ Configurações Sair

## Definir novo padrão de análise

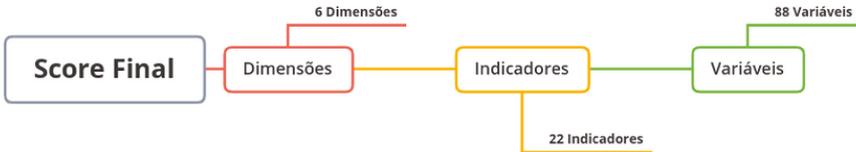
**i** SiAPDA informa:

Os padrões são um conjunto de pesos que são criados por colaboradores. Estes padrões serão utilizados para a avaliação dos portais de dados. É importante salientar que serão criados pesos para dimensões, indicadores e variáveis, possuindo valores de 0 a 100.

A criação de diferentes padrões possibilita a avaliação sobre diferentes perspectivas. Contudo, existe um padrão criado chamado "Default", que possui pesos distribuídos igualmente de modo ao somatório do total ser 100, esta distribuição igual de pesos é feita sobre as dimensões, indicadores e variáveis. O padrão "Default" não pode ser editado nem excluído.

A imagem a seguir apresenta como é a estrutura do modelo, sendo o padrão criado começando a definir o score meta que se almeja (esta definição ajuda a visualizar o quão distante está o score final da meta, auxiliando desenvolver ações evolutivas a fim de atingir uma excelência nos resultados), seguido da definição dos pesos das dimensões e assim por diante até as variáveis. É possível navegar entre os níveis da estrutura para alterar os pesos.

Estrutura dos padrões usando como representação, o modelo de árvore



**+ Novo Padrão**

**Resultados** 10 **Buscar:**

Nome Padrão	Descrição	Data de Criação	Colaborador	Ações
Default	Todos os pesos são distribuídos igualmente de modo ao somatório do total ser 100.	26/11/2020 - 15:48:40	Administrador	Padrão Default não pode ser alterado ou apagado!

Fonte: do Autor

Salienta-se que os pesos podem ter qualquer valor, pois os cálculos são realizados com base no valor máximo obtido pela soma dos pesos, como exemplo, uma dimensão possui 6 indicadores, sendo atribuído os pesos (40, 20, 50, 10, 10, 5) a soma desses valores corresponde ao valor máximo (135) que será utilizado para a normalização, como demonstrado na seção 3.3.3. A Figura 20 ilustra a atribuição de pesos em relação às dimensões.

Figura 20 – Tela definir padrões - SiAPDA

A composição das dimensões irá resultar no Score final da avaliação.

Dimensão	Peso
Dimensão Acesso	40
Dimensão Formatos de dados	20
Dimensão Direitos	50
Dimensão Semântica	10
Dimensão Estrutura Portal	10
Dimensão Apresentação Portal	5
Total	135

Próximo

Copyright © 2014-2019 AdminLTE.io. All rights reserved. Anything you want

Fonte: do Autor

### 3.3.3 Efetuar Cálculos

Por conta das diferentes grandezas utilizadas pelas variáveis foi necessário realizar a transformação dos valores para que haja a sua harmonização e possibilite a realização dos cálculos referentes ao score meta. Para isso, realizou-se a normalização dos valores para que estes fiquem entre 0 e 1; este processo é aplicado para as variáveis e para os pesos definidos. O calculo para a normalização é dado pela fórmula 1.

$$f(v_i) = v_i \div \max v_i \quad (1)$$

Tem-se um conjunto de  $n$  variáveis representadas por  $(v_1, v_2, v_3, \dots, v_n)$ , com  $i$  variando de 1 até  $n$ , onde  $v_i$  representa o valor correspondentes a cada variável e o  $\max v_i$  representa o valor total de ocorrências no universo de cada variável durante a existência do portal. Sendo considerado  $\max v_i = 1$  quando  $v_i$  for uma variável binária. Assim é possível realizar a normalização das  $n$  variáveis, visto que quando  $v_i = \max v_i$ , tem-se o maior valor admitido, ou seja  $f(v_i) = 1$ . Em relação à normalização dos pesos é utilizada a formula 2.

$$f(p_i) = p_i \div \max p_i \quad (2)$$

Tem-se um conjunto de  $n$  pesos representado por  $(p_1, p_2, p_3, \dots, p_n)$ , onde  $p_i$  representa o valor correspondentes ao peso atribuídos a cada variável e o  $\max p_i$

representa a soma de todos os pesos atribuído para as variáveis de um indicador. Assim é possível realizar a normalização dos  $n$  pesos, visto que quando  $p_i = \max p_i$ , tem-se o maior valor admitido, ou seja  $f(p_i) = 1$ .

Portanto o valor ponderado de cada variável será dado pela formula 3.

$$VIP_i = f(v_i) \times f(p_i) \quad (3)$$

Para se obter o valor ponderado referente a cada variável  $VIP_i$  é feita a multiplicação do valor normalizado  $f(v_i)$  pelo peso normalizado  $f(p_i)$ . A Tabela 2 exemplifica os cálculos aplicando as fórmulas apresentadas no indicador **Atualização** utilizando dados fictícios.

Tabela 2 – Exemplificação do cálculo da avaliação

		Dimensão Acesso					
Indicador Atualização		$v_i$	$p_i$	$f(p_i)$	$\max v_i$	$f(v_i)$	$VIP_i$
V01	Atualização nos últimos 12 meses - Conjuntos de Dados	1725	25	0,25	9853	0,175	0,044
V02	Atualização nos últimos 12 meses - Recursos	1371	25	0,25	47766	0,029	0,007
V03	Adicionados nos últimos 12 meses - Conjuntos de dados	5605	25	0,25	9853	0,569	0,142
V04	Adicionados nos últimos 12 meses - Recursos	9593	25	0,25	47766	0,201	0,05
		$\max p_i$	100			$\sum VIP_i$	0,243

Fonte: do Autor

Ao final de todo este processo teremos o score correspondente ao indicador atualização, sendo estes obtidos através de  $\sum VIP_i$ . Esta fórmula será aplicada em outros dois momentos, sendo estes: relacionado as dimensões e ao score final. A Tabela 3 exemplifica os cálculos utilizados em relação a **Dimensão Acesso** utilizando dados fictícios.

Tabela 3 – Exemplificação dos cálculos de indicadores

<b>Dimensão Acesso</b>	$i_j$	$p_j$	$f(p_j)$	$VIP_j$
Ind. Atualização	0,243	14,29	0,143	0,035
Ind. Download em Massa	0,039	14,29	0,143	0,006
Ind. Metadados	1	14,29	0,143	0,143
Ind. Metadados Descritivos	0,858	14,29	0,143	0,123
Ind. Metadados Estrutural	0,012	14,29	0,143	0,002
Ind. Versionamento	0,173	14,29	0,143	0,025
Ind. URI	0,945	14,29	0,143	0,135
	$max p_j$	100	$\sum VIP_j$	0,467

Fonte: do Autor

Como pode ser observado na tabela 3 é utilizada a mesma lógica de cálculo apresentada previamente, sendo que o foco agora são os indicadores. Deste modo é realizada a normalização dos pesos através da fórmula 2 e como os valores dos indicadores  $i_j$  estão normalizados não é necessário aplicar a normalização. Desta forma, é necessário apenas utilizar a formula 3 e por fim realizar o somatório da coluna  $VIP_j$ . deste modo, obtêm-se também o score referente à dimensão.

Por fim, após a realização dos cálculos para todas as dimensões, será possível calcular o score de cada dimensão. A Tabela 4 exemplifica todas as seis dimensões que resultará no score final gerado pelo modelo utilizando dados fictícios.

Tabela 4 – Exemplificação cálculo dimensões

<b>Dimensões</b>	$d_j$	$p_j$	$f(p_j)$	$VIP_j$
Dim. Acesso	0,467	80	0,64	0,299
Dim. Formatos	0,357	10	0,08	0,029
Dim. Direitos	0,199	20	0,16	0,032
Dim. Semântica	0,44	5	0,04	0,018
Dim. Estrutura Portal	0,269	5	0,04	0,011
Dim. Apresentação Portal	0,667	5	0,04	0,027
	$max\_i$	125	$\sum VIP_j$	0,41446

Fonte: do Autor

A Tabela 4 segue a mesma lógica de cálculos apresentados na Tabela3, sendo este o último conjunto de cálculos realizados. É importante salientar que a partir do resultado  $\sum VIP_j$  é determinado o score final do portal de dados avaliado. Como pode

ser observado, o valor obtido na Tabela 4 apresenta um valor normalizado; para que este valor seja comparado ao score meta, faz-se necessária a conversão do valor 0,41446 para a escala de 10, resultando no valor 4,1446, ou seja, se o score meta fosse 5, significaria que o portal está próximo do desejado. Em relação ao SiAPDA, não são exibidas figuras para ilustrar os cálculos, pois estes são realizados em *background*.

### 3.3.4 Analisar

A aplicação do modelo de referência possibilita visualizar uma análise detalhada acerca do portal de dados abertos, nesse sentido devem ser utilizadas as recomendações descritas no apêndice B.

Ao fazer uso do SiAPDA obtém-se uma coleta automatizada proporcionando que os resultados obtidos sejam íntegros. Evitando inconsistências devido ao processo de coleta se tornar massivo, demorado e conter um volume de dados expressivo a serem manipulados.

Neste sentido, o SiAPDA auxilia na realização deste processo de maneira acessível permitindo a disponibilização de informações que viabilizam a compreensão e a interpretação dos resultados, com gráficos, com nuvem de palavras-chaves, entre outros.

O sistema disponibiliza dois tipos de análise, sendo: **i)** visualizar os dados coletados que passaram pelo modelo de forma limpa, ou seja, sem o uso dos padrões. Os cálculos aqui realizados utilizam a "regra de três simples" com o objetivo de determinar o percentual dos valores, como pode ser observado na Figura 21; e **ii)** visualizar os dados coletados utilizando aplicação dos padrões, possibilitando a avaliação do portal sobre os diferentes tipos de padrões disponibilizados, como pode ser observado na Figura 22.

Figura 21 – Tela resultados sem uso de padrões - SiAPDA

**Portais coletados**

http://dados.sc.gov.br --> 15/12/2020 - 02:29:42

Selecionar

**Totais Coletados**

Total de Conjuntos de Dados Coletados: 100  
Total de Recursos Coletados: 826

Dimensão Acesso   **Atualização**   Download em Massa   Metadados

Metadados Descritivos   Metadados Estrutural   Versionamento   URIs

Esta indicador tem por objetivo verificar os conjuntos de dados e os recursos se estão atualizados. Desta forma, conjuntos de dados atualizados são mais atrativos e confiáveis, proporcionando o seu uso ou reuso. A atualização é dada pelo ato de alterar, apagar, adicionar informações ou recursos nos conjuntos de dados. A tabela 1 a seguir apresenta os dados coletados oriundos do portal selecionado

**Recomendações**

- Disponibilizar de dados atualizados regularmente com pelo menos o mesmo nível de qualidade e quantidade.
- Manter os dados atualizados sempre que possível;

Tabela 1 - Operacionalização das variáveis, indicador atualização

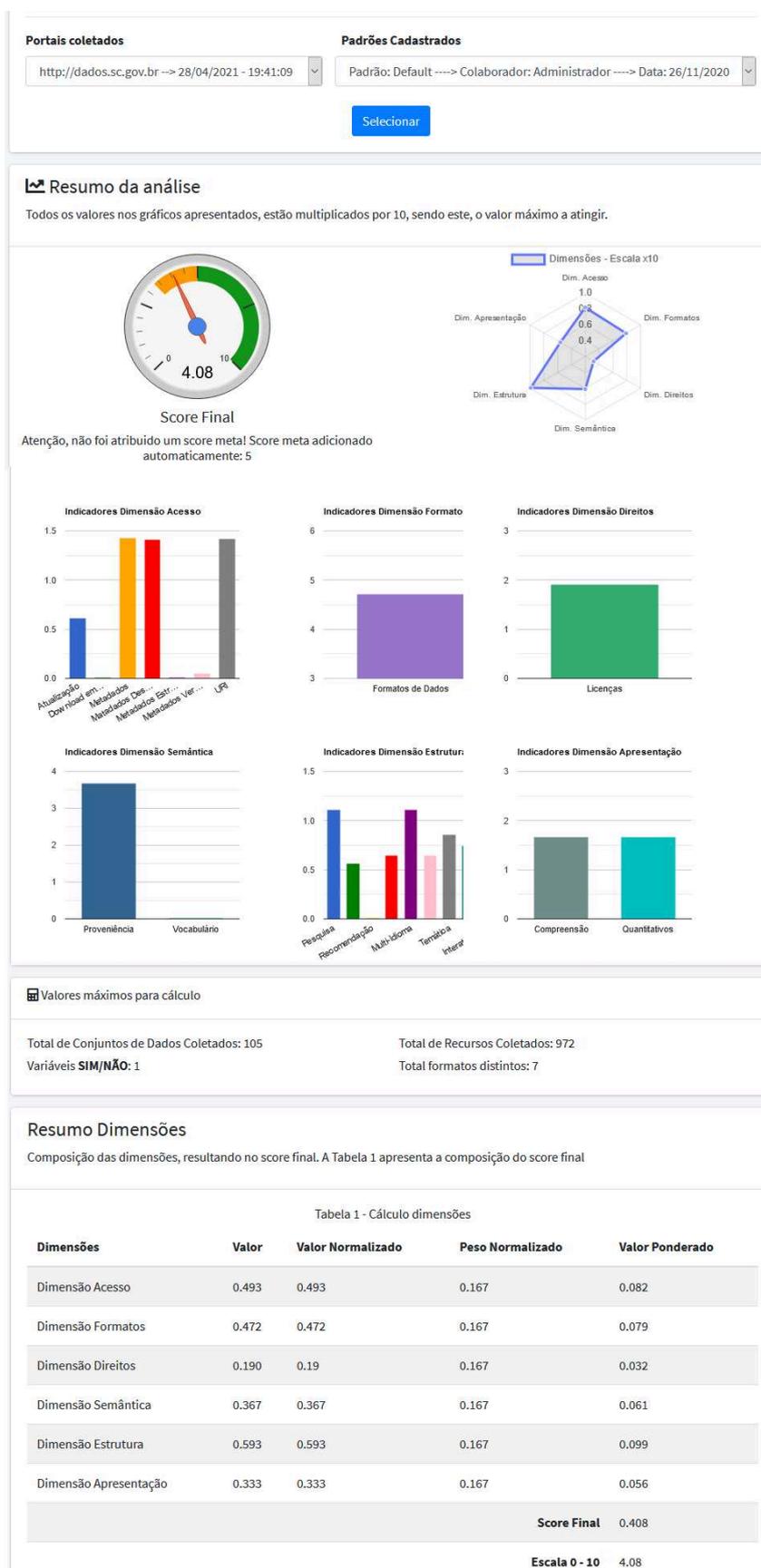
ID	Variáveis	Valores	Percentual
V01	Conjuntos de Dados - Atualização nos últimos 12 meses	70	70.00%
V02	Recursos - Atualização nos últimos 12 meses	27	3.27%
V03	Conjuntos de Dados - Adicionados nos últimos 12 meses	80	80.00%
V04	Recursos - Adicionados nos últimos 12 meses	789	95.52%

Informações complementares  
Gráfico 1 - Série histórica - Atualizações

Conjuntos de dados - Atualizados

Fonte: do Autor

Figura 22 – Tela resultado da avaliação - SiAPDA



A visualização dos resultados conforme Figura 22 possui maiores detalhes em relação a Figura 21, a fim de possibilitar uma análise minuciosa sobre todo o portal. É importante salientar que com relação ao **Resumo da análise** aprestada na Figura 22 os resultados estão na escala de 10.

### 3.4 CONSIDERAÇÕES FINAIS

Neste capítulo foi explicitada a concepção do modelo de referência. Sendo realizado primeiramente o agrupamento dos conceitos encontrados na literatura em dois grupos (conceitos frequentes na literatura e conceitos eventuais na literatura) de modo a harmonizar as diferentes nomenclaturas existentes e, desta forma, identificar os conceitos e iniciar a composição do modelo.

A composição do modelo se deu através da identificação dos indicadores, relacionando os conceitos agrupados, seguindo da definição das dimensões. É importante salientar que as variáveis foram definidas juntamente com os indicadores, verificando a viabilidade da sua operacionalização, mas a apresentação é dada após a exibição das dimensões. Isso ocorreu porque as variáveis apresentadas estão vinculadas aos indicadores e suas dimensões, sendo difícil apresentação somente das variáveis sem este vínculo.

Também foi apresentado o processo de avaliação utilizado no modelo isso porque, quando utilizado o modelo para avaliar um portal de dados, é necessário compreender as etapas envolvidas, os padrões que devem ser criados conforme cada contexto, o entendimento sobre os cálculos que serão executados e as orientações sobre a análise dos resultados.

## 4 PROVA DE CONCEITO

Neste capítulo é exposta a prova de conceito, também conhecida com a nomenclatura inglesa *Proof of Concept* (POC), tendo por objetivo confirmar um modelo prático que possa provar o conceito (teórico) estabelecido por uma pesquisa ou estudo técnico. Sendo este utilizado de várias maneiras, como: *frameworks*, padrões, arquiteturas, modelos, simulação através de software ou ainda um protótipo executável, entre outros (SILVA, M. A. da, 2014).

Desta forma, é apresentada a arquitetura utilizada pelo sistema computacional desenvolvido, além dos módulos criados que são responsáveis pela coleta automatizada dos dados e aplicação do modelo. Também serão apresentados os resultados oriundos do sistema e o resultado do questionário aplicado junto aos especialistas convidados, a fim de avaliar o modelo e o sistema desenvolvido.

### 4.1 SISTEMA DESENVOLVIDO - SIAPDA

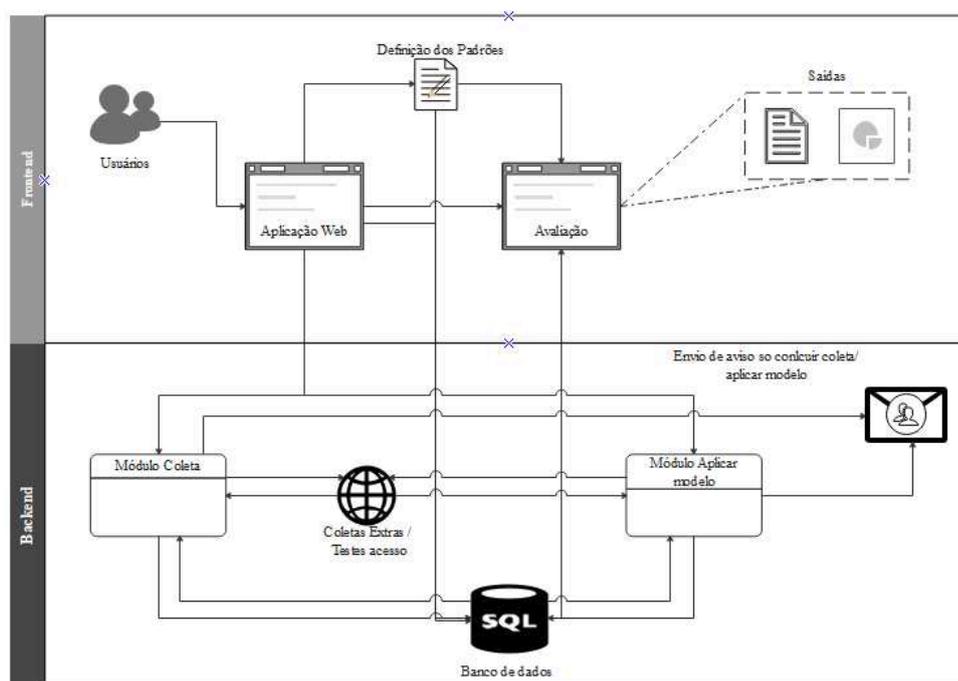
Nesta seção será apresentada a arquitetura do sistema desenvolvido, aspectos relacionados aos tipos de acesso e o fluxo de atividades de cada módulo. Questões sobre o uso do sistema foram abordadas na seção 3.3.

O sistema computacional SiAPDA foi desenvolvido para auxiliar na aplicação do modelo e na avaliação automatizada dos portais de dados abertos, sendo que este sistema possui um registro junto ao Instituto Nacional da Propriedade Industrial (INPI).

Esta aplicação foi desenvolvida utilizando algumas tecnologias, sendo estas, a linguagem de programação Python, banco de dados MySQL, Java Script/Jquery e diversas bibliotecas.

O sistema é dividido em dois grupos, sendo **I**- relacionado ao *back-end* do sistema, onde se encontram os códigos responsáveis pela coleta e aplicação do modelo de referência; e **II**- relacionado ao *front-end* do sistema, onde possibilita toda a interação com os usuários, solicitação de coleta, aplicação do modelo, definição e aplicação dos padrões, bem como a execução da avaliação automatizada e a disponibilização dos resultados para análise. A Figura 23 apresenta a arquitetura do software, sendo esta detalhada a seguir.

Figura 23 – Arquitetura do software - SiAPDA



Fonte: do Autor

Como pode ser observado na Figura 23, a raia que representa o *front-end* é responsável pela interface do sistema. Para acesso ao sistema existem três perfis de usuários, sendo o **i) Administrador** que tem permissão máxima; o **ii) Colaborador** que pode solicitar coletas, a aplicação do modelo e pode criar os padrões; e o **iii) Visitante** que pode visualizar os resultados, sendo estes com e sem a atribuição dos padrões criados. Mais informações sobre as permissões de acesso dos perfis podem ser observadas no Quadro 20.

Quadro 20 – Perfis de usuários e permissões

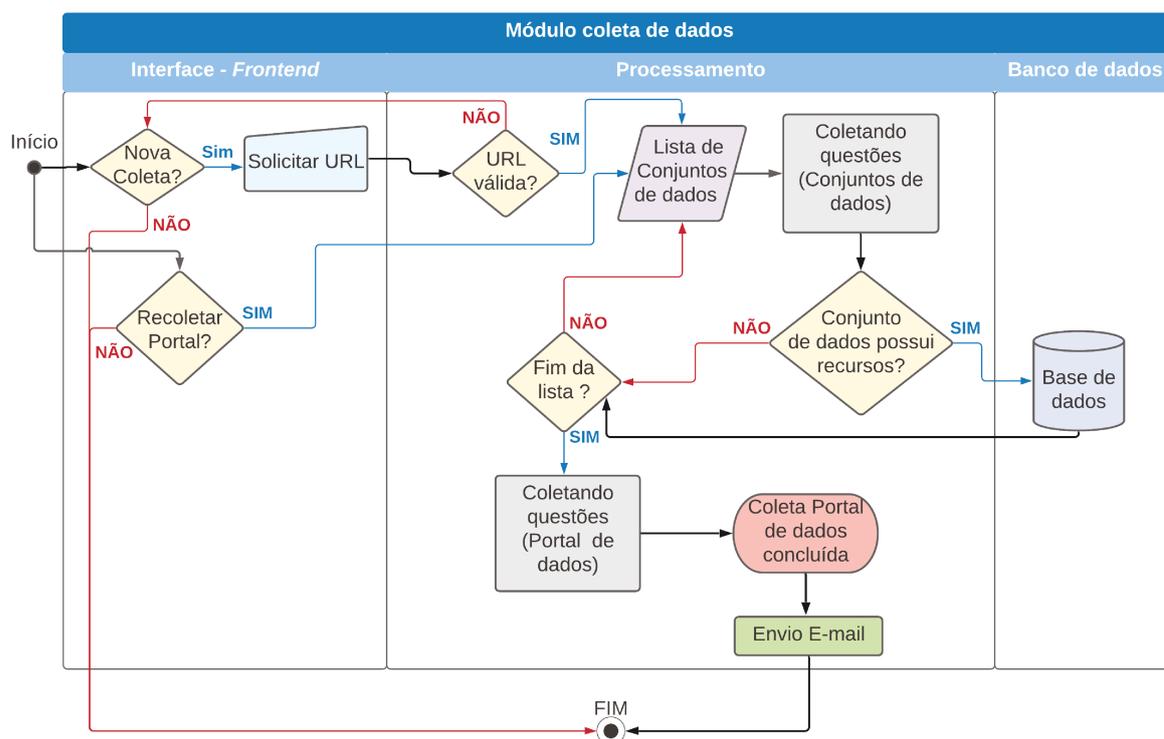
Permissões	Administrador	Colaborador	Visitante
Definir padrões	✓	✓	
Visualizar, editar e excluir os padrões	✓	✓ <sup>1</sup>	
Aplicar os padrões	✓	✓	✓
Visualizar quem definiu os padrões	✓		

Fonte: do Autor

Por sua vez, o *back-end* é o responsável por toda a coleta e aplicação do modelo, a Figura 24 apresenta o módulo **Coleta** sendo este descrito a seguir.

<sup>1</sup> Só o que ele criou

Figura 24 – Modulo Coleta - SiAPDA



Fonte: do Autor

Na Figura 24 o diagrama é separado em três raias, sendo: a **i) Interface Front-End** que está relacionada às atividades da interação, solicitando ao usuário a URL do portal para a coleta ou iniciando um nova coleta de um portal previamente coletado; a **ii) Coleta** relacionada às atividades de coleta dos dados; e o **iii) Banco de dados** responsável pelo arquivamento dos dados.

É possível iniciar a coleta de duas maneiras, sendo a primeira utilizando o URL do portal previamente coletado e realizando uma nova coleta da listagem de conjuntos de dados cadastrados. Utilizando da listagem obtida é iniciado a coleta individual, sendo verificado se existem recursos vinculados a cada conjunto de dados. Caso o conjunto de dados possua algum recurso este será armazenado no banco de dados. Será repetido este procedimento até o último conjunto de dados listado.

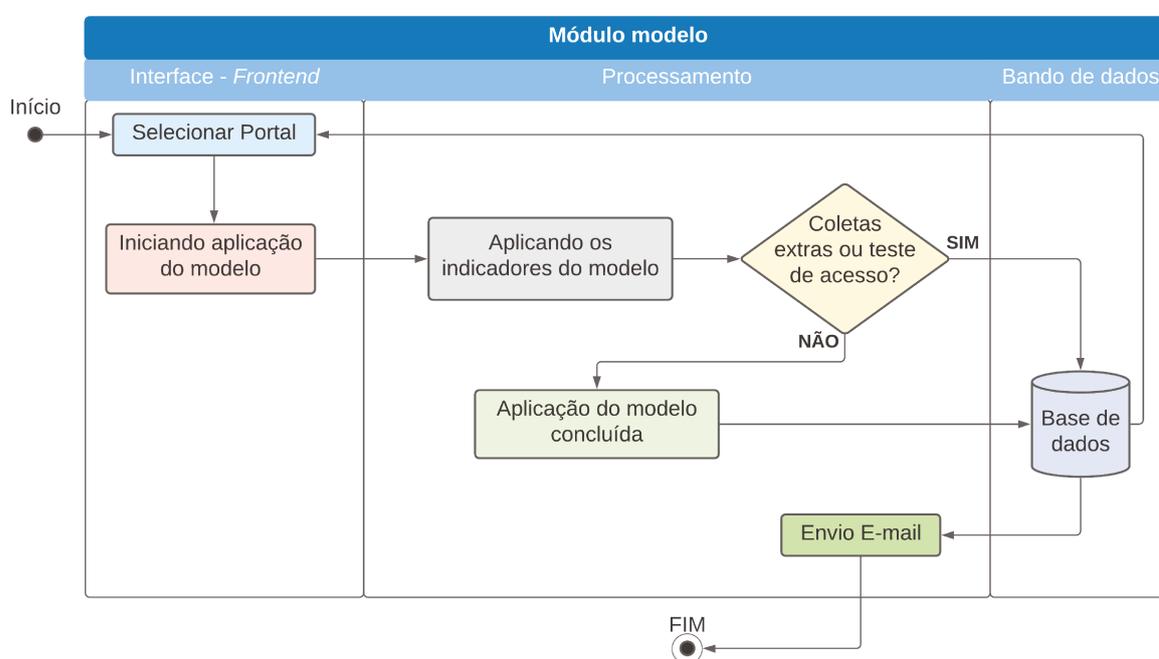
A segunda maneira possível de coleta é inserindo a nova URL para que seja coletada a listagem de conjuntos de dados e posteriormente iniciado a coleta. Destaca-se que quando iniciado uma nova coleta é realizada a validação automática da URL, verificando se o portal de dados é baseado na plataforma CKAN (limitação do SiAPDA-v1.0). Caso seja obtida a resposta verdadeira, inicia-se a coleta dos conjuntos de dados. No caso da validação retornar resposta negativa, será informado ao usuário que o portal de dados não é uma plataforma CKAN e será solicitado novamente se deseja iniciar uma nova coleta.

Ao finalizar a coleta de todos os conjuntos, será iniciado o processo de coleta sobre os aspectos relacionados ao portal de dados, sendo este realizado através de robôs e ao final do processo será informado o solicitante por e-mail do fim desta etapa.

A verificação de conjuntos de dados que possuem recursos é necessária, pois o CKAN permite cadastrar diferentes tipos de dados e alguns portais além de cadastrar os conjuntos de dados, cadastram *log de dados*<sup>2</sup> de coleta sobre plataformas de terceiros, por exemplo.

Assim o módulo **modelo** tem por objetivo complementar a coleta atribuindo valores agregados à variáveis dependentes da primeira coleta e concluir a aplicação do modelo. A seguir a Figura 25 apresenta o fluxo de atividades do módulo modelo.

Figura 25 – Módulo Modelo - SiAPDA



Fonte: do Autor

Na Figura 25 o diagrama é separado em três raias, sendo: a **i) Interface Front-End** relacionada às atividades de interação, onde é selecionado o portal para aplicação do modelo; a raia de **ii) Processamento** relacionada às atividades de processamento dos indicadores sobre os dados coletados; e a raia de **iii) Banco de Dados** relacionada a recuperação e o armazenamento dos dados.

Este módulo inicia trazendo a listagem dos portais de dados coletados para que seja selecionado e aplicado o modelo, como foi apresentado na seção 3.3.1 na Figura 18. Após a seleção do portal é iniciada a aplicação do modelo e a comple-

<sup>2</sup> Na computação, log de dados é utilizado para registrar os processos ou eventos relevantes num sistema computacional, de modo a ser utilizado para auditorias e diagnósticos de possíveis problemas relacionados a um sistema ou rotina computacional.

mentação da etapa de coleta (coleta extras e teste de acesso), onde faz-se o uso de dados da primeira coleta para atribuição de valores compilados para algumas variáveis. Caso haja a complementação da coleta será realizado o armazenamento no banco de dados. Após a aplicação do modelo o resultado também irá ser armazenado no banco de dados e após o termino será notificado por e-mail o administrador e o solicitante, caso seja realizado por usuários do tipo colaborador.

Na sequência serão apresentados os resultados obtidos com a aplicação do modelo utilizando do SiAPDA.

#### 4.2 APLICAÇÃO DO MODELO EM TRÊS CENÁRIOS DE USO

Nesta seção serão apresentados os resultados obtidos através do SiAPDA. Primeiramente foram coletados os dados. Deste modo, foi possível aplicar o modelo. A seguir o Quadro 21 apresenta quais foram os portais de dados coletados.

Quadro 21 – Portais coletados para avaliação e análise

Portal	Esfera Portal	Qtd. Conjuntos de dados coletados	Qtd. Recursos Coletados	Data de coleta
Prefeitura Municipal de São Paulo	Municipal	195	3232	06/02/2021
Portal de dados do estado do Rio Grande do Sul	Estadual	1151	16039	01/02/2021
Portal do Governo Federal - Brasil	Federal	10851	51100	05/02/2021

Fonte: do Autor

Todos os resultados apresentados podem ser visualizados utilizando o sistema<sup>3</sup>, sendo necessário apenas o cadastro. Em relação à definição dos padrões, utilizou-se padrão **Default**, sendo os pesos atribuídos igualmente para todas as dimensões, indicadores e variáveis.

Os portais foram selecionados para a aplicação do modelo por estes conterem um grande volume de dados em relação aos três níveis do governo brasileiro e por utilizarem plataforma CKAN.

<sup>3</sup> <https://murilogomes.studiokem.ufsc.br/acesso>

A seguir serão apresentados os resultados com foco nos indicadores e nas dimensões, contudo as dimensões formatos e direitos serão apresentadas também as variáveis para melhor explicitar a análise dos portais. As variáveis não serão expostas nesta seção, por conta do grande volume de dados a ser analisado em cada portal. No entanto, o resultado detalhado estará disponível online e no(Apêndice C).

Em relação aos valores de cada indicador apresentado, estes por sua vez estão normalizados. Os valores relacionados aos indicadores são resultado do somatório das variáveis, conforme apresentado na seção 3.3.3.

A literatura não apresenta valores de referência para que possam ser comparados com os valores aqui apresentados, no entanto, o score final será dado em relação ao score meta.

#### 4.2.1 Análise dos resultados - Dimensão Acesso

Esta dimensão remete-se ao acesso aos dados disponibilizados no portal. Sendo possível identificar se os dados estão acessíveis e localizáveis, para isso esta dimensão conta com sete indicadores, sendo todos abordados na Tabela 5.

Tabela 5 – Aplicação modelo - Dimensão Acesso

Dimensão Acesso				
Portal	Indicadores	Valor	Peso Normalizado	Valor Ponderado
Federal	Atualização	0,258	0,143	0,037
	Download em Massa	0,039	0,143	0,006
	Metadados	1,000	0,143	0,143
	Metadados Descritivos	0,854	0,143	0,122
	Metadados Estrutural	0,006	0,143	0,001
	Versionamento	0,179	0,143	0,026
	URI	0,887	0,143	0,127
<b>Valor Dimensão Acesso – Gov. Federal</b>				<b>0,461</b>
Estado do RS	Atualização	0,008	0,143	0,001
	Download em Massa	0,005	0,143	0,001
	Metadados	1,000	0,143	0,143
	Metadados Descritivos	0,773	0,143	0,110
	Metadados Estrutural	0,000	0,143	0,000

	Versionamento	0,020	0,143	0,003
	URI	0,968	0,143	0,138
	<b>Valor Dimensão Acesso – Gov. Estadual</b>			<b>0,396</b>
Município de SP	Atualização	0,189	0,143	0,027
	Download em Massa	0,107	0,143	0,015
	Metadados	0,997	0,143	0,142
	Metadados Descritivos	0,886	0,143	0,124
	Metadados Estrutural	0,009	0,143	0,001
	Versionamento	0,087	0,143	0,012
	URI	0,989	0,143	0,141
	<b>Valor Dimensão Acesso – Gov. Municipal</b>			<b>0,464</b>

Fonte: do Autor

Na Tabela 5 em relação ao portal de dados do **governo federal**, temos o indicador atualização que apresenta o valor ponderado de 0,037, sendo este, superior em relação aos outros portais, indicando uma maior periodicidade nas atualizações. Este destaque é natural, considerando ser um portal federal e por englobar diferentes órgãos da federação que publicam e atualizam seus dados. Por sua vez, o **estado do Rio Grande do Sul** possui valor 0,001 de atualização, sendo um valor baixo, comparando ao do governo federal e levando em conta o valor 0,027 do **município de São Paulo**.

Desta forma, destaca-se o estado do Rio Grande do Sul que nos últimos doze meses teve atualizações esporádicas de seus dados em relação aos outros dois portais, sendo importante salientar que foram publicados e atualizados dados sobre a pandemia da COVID-19, como percebido nos outros portais.

O indicador atualização proporciona a verificação da periodicidade de atualização e isso indica que este está operacional, podendo ser observado que os três portais dispõem de alguma atualização (conjunto de dados e recursos), atendendo a recomendação de manter os dados atualizados sempre que possível.

Por sua vez, o indicador **Download em Massa** se destacou no **município de São Paulo** obtendo valor ponderado 0,015 maior que o portal do **governo federal** com 0,006 e do estado do Rio Grande do Sul. O estado do **Rio Grande do Sul** novamente possui o menor valor ponderado sendo 0,001.

Este indicador proporciona perceber a preocupação das organizações em disponibilizar grandes volumes de dados organizados e consistentes, mesmo quando

acessados ou coletados de uma única vez, independente da quantidade disponibilizada no portal. Por mais que os portais analisados tenham tido um valor baixo neste indicador, é importante destacar que a literatura não determina um quantitativo ideal, sendo assim, ao disponibilizar dados que se enquadram neste indicador, os portais estão em consonância a esta recomendação.

O indicador **Metadados** é um dos indicadores que possui o valor máximo nos três portais analisados. Isso ocorre, porque é através deste indicador que viabiliza-se a localização dos conjuntos de dados e seus recursos, além de que a literatura aponta como um requisito fundamental para a publicação de dado na Web. Deste modo, temos o **governo federal** e **estado do Rio Grande do Sul** com o valor ponderado de 0,143 sendo este o máximo considerando. O **município de São Paulo** com o valor 0,142. Destaca-se que ao utilizar da plataforma CKAN o portal de dados possui todo o suporte para disponibilizar metadados, desta forma, naturalmente os portais terão valores próximos ao máximo.

Sobre o indicador **Metadados Descritivos** pode ser observado que o **município de São Paulo** tem o maior valor ponderado 0,124, seguindo do **governo federal** com valor 0,122 e o menor com o **estado do Rio Grande do Sul** com valor 0,110. Este indicador é uma extensão do indicador metadados, pois por meio deste são analisadas as informações que descrevem os metadados, como por exemplo, quem é o autor, quando foi criado, dentre outros. Neste sentido, o ideal seria possuir o valor máximo nos três portais, pois segundo as recomendações, os metadados devem estar presentes em todos os conjuntos de dados e o mais detalhado possível.

Referente ao indicador **Metadado Estrutural** pode-se observar que o **estado do Rio Grande do Sul** possui valor ponderado 0, por sua vez, os portais do **município de São Paulo** e do **governo federal** possuem o mesmo valor ponderado 0,001. Contudo, destaca-se que o portal do **município de São Paulo** possui maior valor em relação ao portal do governo federal.

Este indicador permite verificar se o portal de dados disponibiliza informações sobre a estrutura dos dados disponibilizados, ou seja, informações sobre o conteúdo dos recursos viabilizando a sua compreensão. Este tipo de metadados é comumente disponibilizado através de recursos e informam detalhes sobre as colunas de uma planilha ou arquivo CSV, por exemplo. Observam-se valores baixos sobre os metadados estruturais, além da ausência no portal do estado do Rio Grande do Sul. Isso implica na dificuldade do entendimento sobre os dados disponibilizados.

No que concerne ao indicador **Versionamento** se destaca no portal do **governo federal** tendo como valor ponderado 0,026, seguindo do **município de São Paulo** com 0,012 e o portal do **estado do Rio Grande do Sul** com o menor valor ponderado de 0,003. A literatura não impõe nenhum quantitativo ideal, apenas recomenda a enumeração dos conjuntos de dados ou mesmo o uso de datas nas URIs. Deste modo,

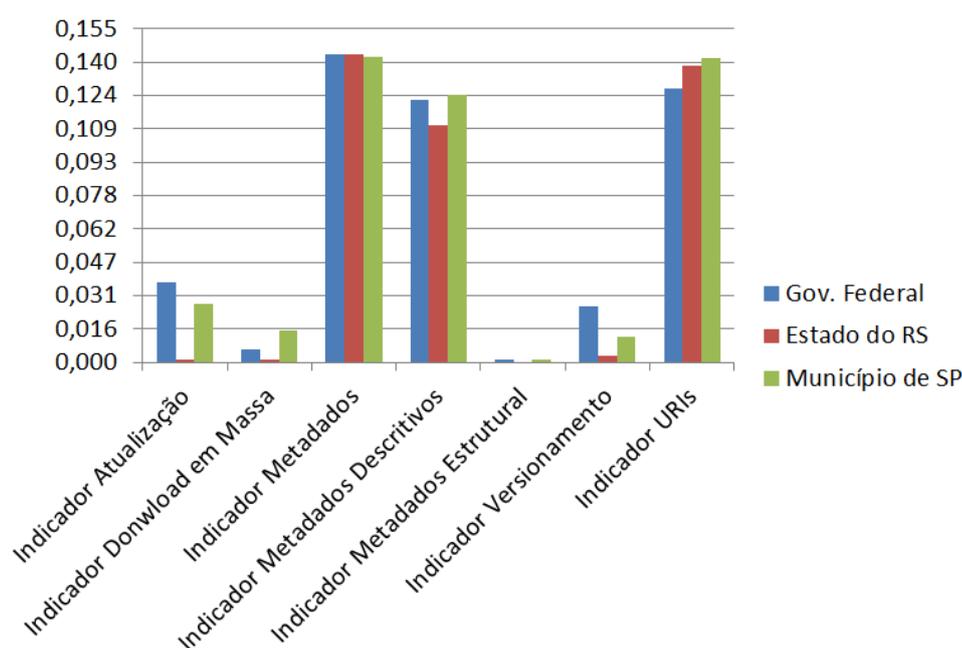
permite aos *stakeholders* identificarem as versões publicadas e acessarem as versões que desejam, possibilitando a identificação com facilidade ou mesmo proporcionar a continuidade no uso de uma versão antiga. Destaca-se que ao utilizar plataforma, esta também possibilita o suporte para atribuir o controle de versão, sendo este mais efetivo e evidente no uso de versionamento em relação à atribuição de datas nas URIs.

Por fim, o **indicador URI** expõem os valores consideráveis em relação aos três portais, contudo o portal do **município de São Paulo** obteve o maior valor ponderado de 0,141, seguindo do **estado do Rio Grande do Sul** com 0,138 e com menor valor ponderado está o portal do **governo federal** com o valor 0,127. Este indicador possibilita identificar se os endereços URI estão acessíveis, ou seja, caso um *stakeholder* selecione um determinado conjunto de dados, este será encontrado, tornando-o acessível. O recomendado é que todos os endereços URI estejam acessíveis.

Neste sentido, percebe-se que o portal do governo federal possui um maior quantitativo de dados que não estão acessíveis através das URIs disponibilizadas. Por sua vez, observa-se que os outros portais também possuem URIs inacessíveis, apenas em menor quantidade. Vale ressaltar que o portal do governo federal possui outros órgãos e parceiros que publicam junto ao portal, sendo difícil o monitoramento sobre a sua totalidade. Contudo, as ações de monitoramento possibilitam a identificação destas inconsistências dentre outras que podem surgir.

O Gráfico 1 apresenta os valores referentes aos indicadores de cada portal, sendo possível observar quais portais se sobressaem em relação a um determinado indicador.

Gráfico 1 – Resultado - Dimensão Acesso

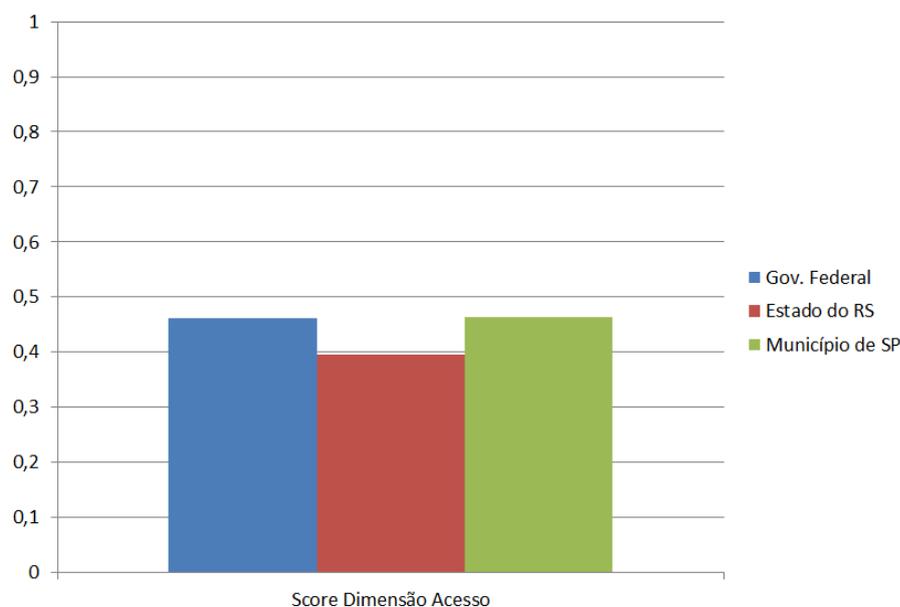


Fonte: do Autor

Observa-se no Gráfico 1 que os três portais possuem similaridade nos valores obtidos em três indicadores, sendo: metadados, metadados descritivos e URIs. Estes indicadores possuem valores próximos ao máximo, indicando que os conjuntos de dados dos portais poderão ser localizados tanto no portal como em buscadores(Google, Bind, etc) e possuem o acesso garantido por conta dos resultados apresentados pelo indicador URI. No entanto, em relação aos demais indicadores, os portais podem realizar ações para melhorar.

Em relação aos pesos, em todos os indicadores foram utilizados o peso 0,143 para que houvesse a composição da dimensão acesso. O Gráfico 2 apresenta o score obtido em relação à dimensão acesso por cada portal, sendo que o portal do município de São Paulo ficou melhor avaliado, seguindo o portal do governo federal e em último o portal do estado do Rio Grande do Sul.

Gráfico 2 – Comparação do Score - Dimensão Acesso



Fonte: do Autor

Ao analisar toda a dimensão, percebe-se que os portais atendem parcialmente às recomendações oriundas da literatura e das boas práticas. Contudo, é importante destacar que é possível melhorar os indicadores que apresentam deficiências com ações de curto, médio e longo prazo.

#### 4.2.2 Análise dos resultados - Dimensão Formatos

Esta dimensão refere-se aos formatos de arquivos disponibilizados, ou seja, os diferentes recursos disponibilizados nos conjuntos de dados, sendo estes legíveis por máquinas ou humanos. A Tabela 6 apresenta o indicador empregado nos três portais.

Tabela 6 – Aplicação modelo - Dimensão Formatos de Dados

Dimensão Formatos de Dados				
Portal	Indicadores	Valor	Peso Normalizado	Valor Ponderado
Federal	Formatos de Dados	0,398	1,000	0,398
	Valor Dimensão Formatos – Gov. Federal			0,398
Estado do RS	Formatos de Dados	0,498	1,000	0,498
	Valor Dimensão Formatos – Gov. Estadual			0,498
Município de SP	Formatos de Dados	0,394	1,000	0,394
	Valor Dimensão Formatos – Gov. Municipal			0,394

Fonte: do Autor

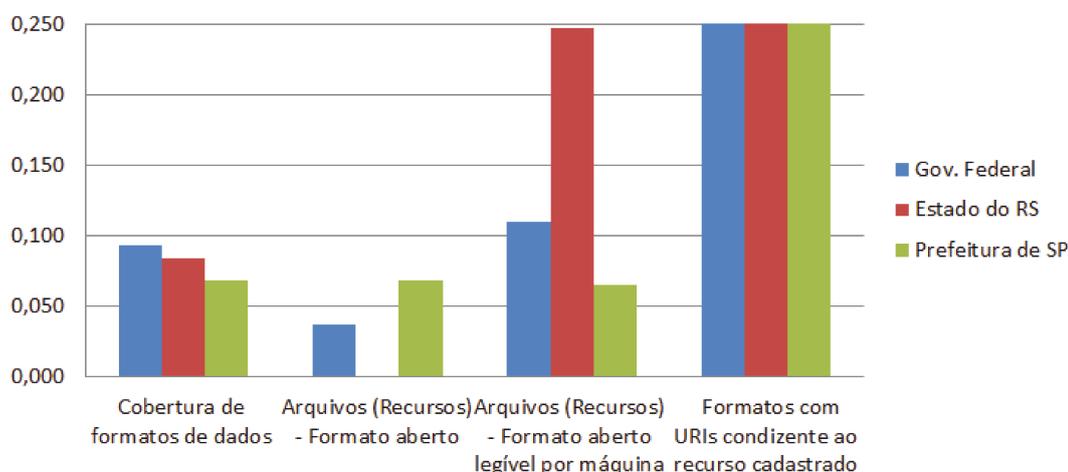
Em relação ao indicador **Formatos de dados** pode ser observado na Tabela 6 que o portal do **estado do Rio Grande do Sul** possui o maior valor ponderado 0,498, seguindo pelo portal do **governo federal** com valor de 0,398 e o portal do **município de São Paulo** com o valor ponderado 0,394.

Com o objetivo de analisar a dimensão formatos, o Gráfico 3 ilustra os valores ponderados correspondentes às variáveis de cada portal analisado.

Observando o Gráfico 3 tem-se quatro variáveis para este indicador, sendo atribuído o peso de 0,25 a cada variável. Assim, a variável **formatos com URI condizente ao recurso cadastrado** está com o valor máximo 0,250 para os três portais. Isso é excelente, pois, indica que os formatos informados nos metadados são os mesmos formatos disponibilizados para acesso. Esta variável permite identificar se existem discrepâncias entre os formatos de arquivos cadastrados nos metadados e o real formato disponibilizado, sendo importante haver a homogeneidade entre estes dois aspectos. Visto que em caso de discrepância pode haver confusão, frustração e até mesmo inviabilizar o uso dos recursos por parte dos *stakeholders*.

Em relação às variáveis **arquivos(recursos) - formato aberto e formato aberto legível por máquinas** ambas são complementares em relação ao quesito formato aberto. Contudo, a variável formato aberto legível por máquina é o tipo de formato

Gráfico 3 – Resultados Variáveis - Dimensão Formatos



Fonte: do Autor

que ganha destaque na literatura por contribuir com a redução de custos (tempo e processamento) ao fazer uso de agentes computacionais.

Neste sentido, o portal do estado do Rio Grande do Sul se destaca por ter o valor próximo ao máximo, seguido do portal do governo federal e por último o portal da município de São Paulo. É importante ressaltar que a literatura salienta que deve-se disponibilizar preferencialmente formatos legíveis por máquina, mas também podem ser disponibilizados formatos abertos.

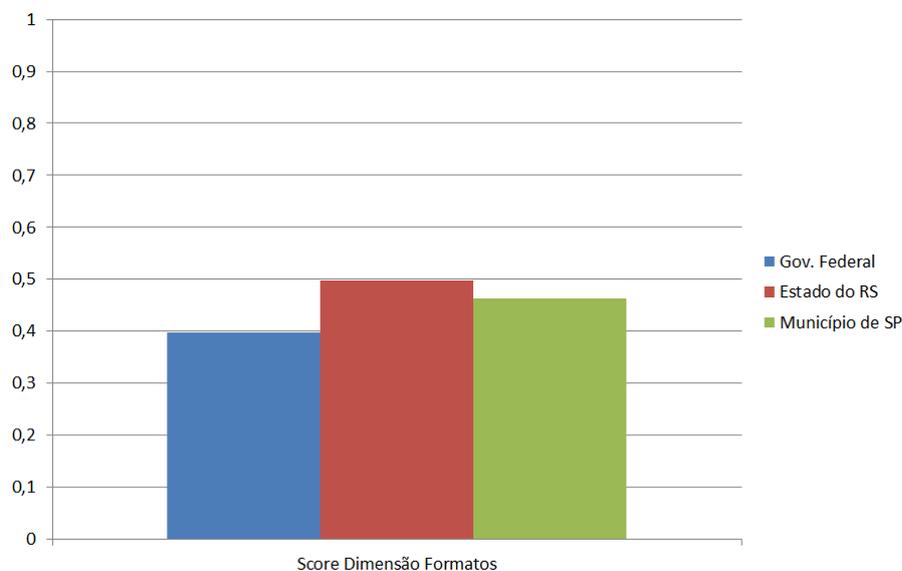
Ao disponibilizar diferentes formatos, o portal também atinge a recomendação acerca da variedade de formatos disponibilizados. Desta forma, o portal do Rio Grande do Sul está alinhado com relação à disponibilização de dados legíveis por máquinas, contudo, seu foco está apenas neste aspecto deixando a desejar na questão da variedade. Neste caso, o portal pode disponibilizar além dos recursos legíveis por máquinas outros formatos abertos, como *Open Document (ODT)*, *Portable Document Format (PDF)*, *Open Document Spreadsheets (ODS)* e etc, disponibilizando uma variedade de formatos.

A variável **cobertura de formatos de dados** permite vislumbrar a variedade de formatos de dados não repetidos por conjuntos de dados. Deste modo, é possível identificar o quão diversificado são os conjuntos de dados disponibilizados nos portais, esta diversificação permite alcançar um maior número de *stakeholders* e no aumento da probabilidade do uso destes dados.

A literatura não define valores quantitativos relacionados aos formatos, mas informa que quanto maior for a variedade de formatos de dados melhor será para o portal e para os *stakeholders*, por este motivo destaca-se a importância das variáveis que compõem este indicador.

Esta dimensão possui apenas um indicador e para este foi atribuído o peso 1. O Gráfico 4 aborda o comparativo entre os scores obtidos pelos portais, nesta dimensão.

Gráfico 4 – Comparação do Score - Dimensão Formatos



Fonte: do Autor

Ao analisar esta dimensão, compreende-se que os portais disponibilizam diferentes formatos de arquivos. Deste modo percebe-se que como na maioria dos estudos enfatizam o uso de formatos legíveis por máquina este é o que mais se destaca nesta dimensão. No entanto nota-se que isso pode restringir o perfil de *stakeholders*. O ideal é manter uma variedade e quando disponibilizar um conjunto de dados com formato legível por máquina, também disponibilizar outros formatos complementares.

Em geral, os portais estão publicando diferentes tipos de formatos e se preocupando em disponibilizar os formatos legíveis por máquina, ou seja, percebe-se que os portais estão em conformidade com as recomendações.

#### 4.2.3 Análise dos resultados - Dimensão Direitos

Esta dimensão refere-se aos direitos sobre os dados abertos. Estes direitos são atribuídos através das licenças aos conjuntos de dados. Vale ressaltar que ao empregar a licença a um conjunto de dados, normalmente esta licença se estende aos recursos. As licenças são apontadas como crucial na publicação dos dados, pois é através desta que os *stakeholders* poderão fazer uso de forma livre e sem impedimentos legais. A Tabela 7 apresenta o indicador responsável por identificar os direitos sobre os dados em relação aos portais.

Tabela 7 – Aplicação modelo - Dimensão Direitos

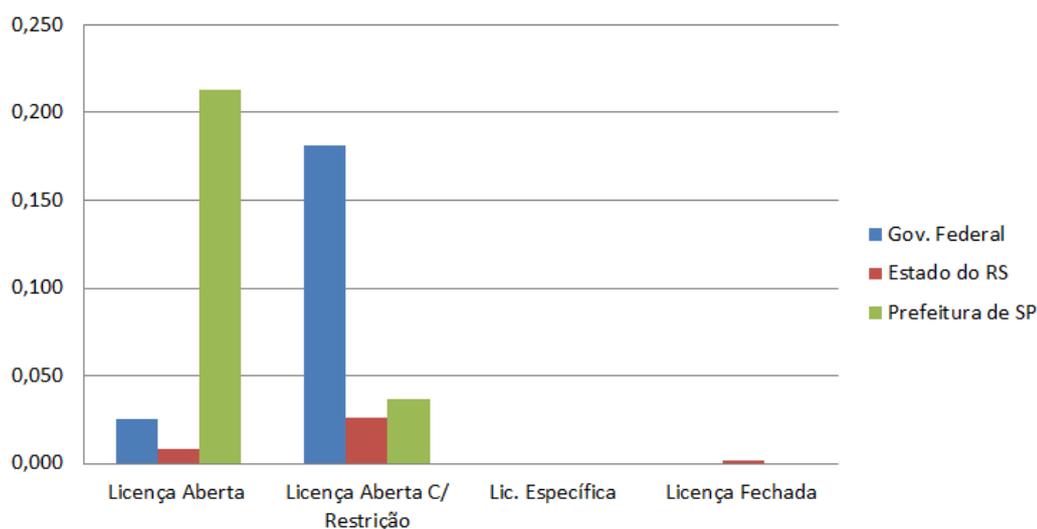
Dimensão Direitos				
Portal	Indicadores	Valor	Peso Normalizado	Valor Ponderado
Federal	Licença	0,206	1,000	0,206
	Valor Dimensão Direitos – Gov. Federal			0,206
Estado do RS	Licença	0,035	1,000	0,035
	Valor Dimensão Direitos – Gov. Estadual			0,035
Município de SP	Licença	0,250	1,000	0,250
	Valor Dimensão Direitos – Gov. Municipal			0,250

Fonte: do Autor

Como observado na Tabela 7 o portal do **município de São Paulo** possui o valor 0,250, seguido pelo **governo federal** com o valor 0,206 e o valor 0,035 para o portal do **estado do Rio Grande do Sul**. A literatura não determina um valor mínimo ou máximo de licenças, mas levando como premissa que a licença é crucial em dados abertos, é importante que todos os conjuntos de dados possuam alguma licença atribuída.

Para apresentar a dimensão direitos que também possui apenas um indicador, o Gráfico 5 mostra os valores correspondentes às variáveis.

Gráfico 5 – Resultados Variáveis – Dimensão Direitos



Fonte: do Autor

Percebe-se no Gráfico 5 a variável **licença específica**, que nenhum portal faz uso isso indica que os portais utilizam somente de licenças definidas por organizações

regulamentadoras. Isso facilita a compreensão no uso dos dados visto que as licenças definidas por estas organizações são utilizadas globalmente. Este tipo de licença específica normalmente é definida por governos a fim de atender normas locais e específicas, como exemplo, a Licencia de DAG de Uruguay v0.1<sup>4</sup>, que é usada no portal de dados do Uruguai<sup>5</sup> na maioria dos dados disponibilizados.

Em relação à variável **licença fechada** apenas o estado do Rio Grande do Sul utiliza licenças fechadas atribuídas aos conjuntos de dados. O portal do governo Federal possui apenas 10 conjuntos de dados que fazem uso desta licença; por ser uma pequena quantidade de licenças em relação ao total de conjuntos de dados disponibilizados, não aparece no gráfico. Esta licença é a menos indicada, pois inviabiliza o uso dos dados, sendo contra os princípios de dados abertos.

As variáveis **licença aberta e licença aberta com restrição**<sup>6</sup> permitem o uso dos dados disponibilizados. A única diferença é que ao utilizar a **licenças com restrições** deve ser atribuída a fonte dos dados.

Ambas as licenças são recomendadas, contudo destaca-se que ao atribuir uma **licença aberta** os dados tornam-se públicos para qualquer finalidade, dando maior autonomia e liberdade no uso destes. Neste sentido, o portal de dados do município de São Paulo se destaca por utilizar na maioria dos seus conjuntos de dados a licença aberta e em relação ao uso de licença aberta com restrição o portal do governo Federal se sobressai.

O portal do estado do Rio Grande do Sul possui poucas licenças em uso, sendo que cerca de 84% dos seus conjuntos de dados não possuem licenças atribuídas ou não especificadas, dificultando o seu uso e falhando em um dos princípios básicos sobre dados abertos, onde para ser considerado aberto este deve ter vínculo com uma licença aberta.

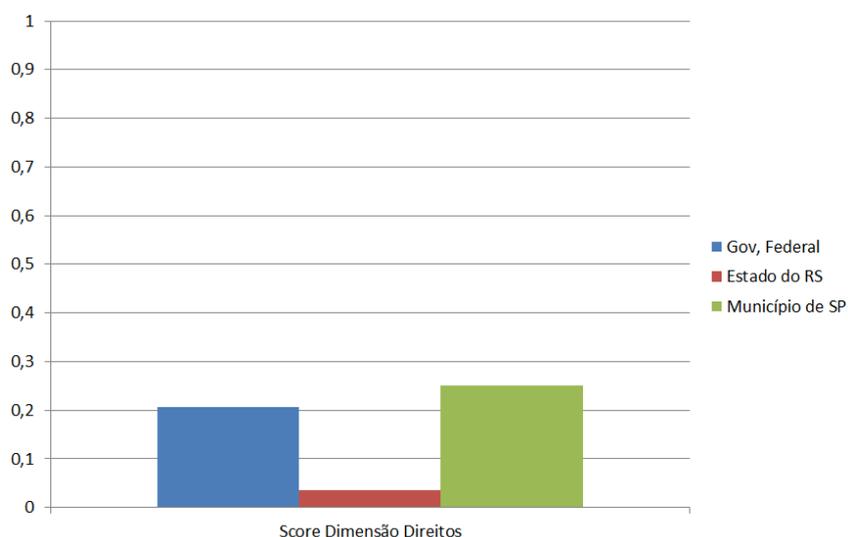
Neste caso, percebe-se que os portais possuem algumas licenças (aberta, com restrição ou fechada), contudo é importante salientar que o tipo de licença aberta (CC-ZERO e ODC-PDDL) são as mais indicadas. No geral, os portais fazem uso dessas licenças, estando de acordo com as recomendações, mas destaca-se que devem ser definidas as licenças em todos os conjuntos de dados para que os *stakeholders* saibam como utilizar e qual o limite do uso dos dados disponibilizados. O Gráfico 6 apresenta o comparativo do score desta dimensão entre os portais.

<sup>4</sup> Mais informações em <[https://www.gub.uy/agencia-gobierno-electronico-sociedad\informacion-conocimiento/sites/agencia-gobierno-electronico-sociedad\informacion-conocimiento/files/documentos/publicaciones/licencia\\_de\\_datos\\_abiertos\\_0.pdf](https://www.gub.uy/agencia-gobierno-electronico-sociedad\informacion-conocimiento/sites/agencia-gobierno-electronico-sociedad\informacion-conocimiento/files/documentos/publicaciones/licencia_de_datos_abiertos_0.pdf)>

<sup>5</sup> Acesso para o portal de dados abertos do Uruguai <<https://catalogodatos.gub.uy>>

<sup>6</sup> Esta licença é chamada de Licença por Atribuição, neste trabalho foi denominada de Licença aberta com restrição para destacar as diferenças entre as licenças analisadas e por englobar diferentes categorias de licenças.

Gráfico 6 – Comparação do Score – Dimensão Direitos



Fonte: do Autor

#### 4.2.4 Análise dos resultados - Dimensão Semântica

Esta dimensão possui um vínculo direto com a Web semântica, visto que tem como objetivo a criação da Web de dados, conforme apresentado na seção 2.3. A Web de dados possui padrões e formatos para integrar dados oriundos de diferentes fontes, para que haja uma compreensão dos diversos domínios, sendo os vocabulários que contribuem para este entendimento. Assim, pelo fato da Web de dados possuir várias fontes publicadoras, faz-se necessário o uso da proveniência, de modo a contribuir com a confiabilidade sobre as fontes que publicam dados. Deste modo, a Tabela 8 apresenta os dois indicadores que compõem a dimensão semântica.

Tabela 8 – Aplicação modelo - Dimensão Semântica

Dimensão Semântica				
Portal	Indicadores	Valor	Peso Normalizado	Valor Ponderado
Federal	Proveniência	0,879	0,5	0,440
	Vocabulários	0,000	0,5	0,000
	Valor Dimensão Semântica – Gov. Federal			0,440
Estado do RS	Proveniência	0,512	0,5	0,256
	Vocabulários	0,000	0,5	0,000
	Valor Dimensão Semântica – Gov. Estadual			0,256
Município de SP	Proveniência	0,774	0,5	0,387
	Vocabulários	0,000	0,5	0,000
	Valor Dimensão Semântica – Gov. Municipal			0,387

Fonte: do Autor

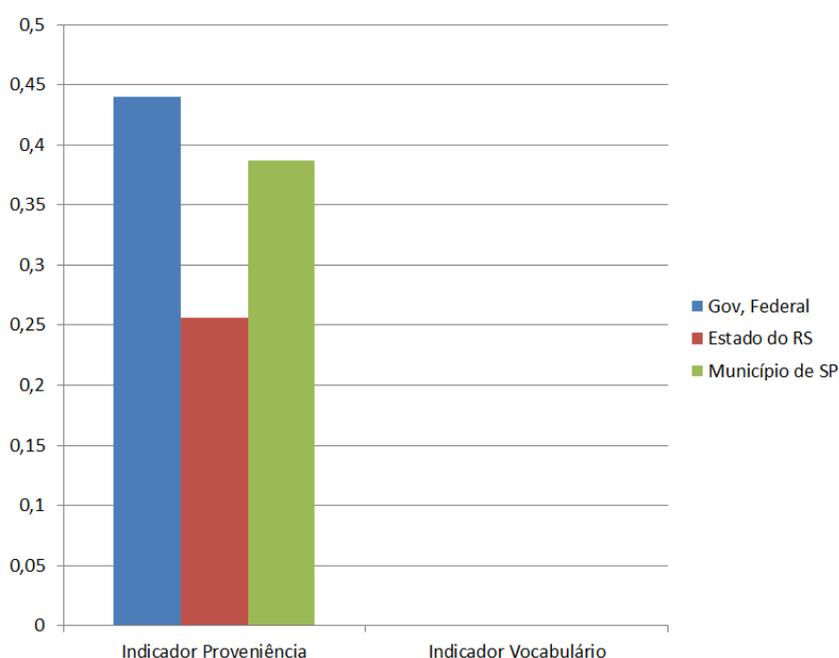
Ao observar a Tabela 8 o indicador vocabulário esta ausente nos três portais. Este indicador proporciona a compreensão do significado, a interoperabilidade e flexibiliza o uso por agentes computacionais, sendo fortemente recomendado pela literatura e importante para o desenvolvimento da Web de dados.

Entende-se a complexidade e tempo demandado para que organizações disponibilizem conjuntos de dados com vocabulários. No entanto, as plataformas disponibilizam uma estrutura de modo a facilitar o emprego de vocabulários. O auxílio oferecido pelas plataformas, em específico o CKAN proporciona o emprego do vocabulário DCAT salienta-se que a organização pode aprimorar este vocabulário pré-definido a fim de complementar a compreensão dos significados dos dados.

Por sua vez, o **indicador proveniência** está presente nos três portais, onde estes disponibilizam informações sobre a fonte publicadora dos conjuntos de dados. O portal do **governo federal** possui o valor 0,440, seguido do portal **município de São Paulo** com 0,387 e o portal do **estado do Rio Grande do Sul** com o valor 0,256.

Isso indica que estão disponibilizando informações sobre as organizações que originam os dados, passando confiança aos *stakeholders* em relação aos dados disponibilizados, estando de acordo com a recomendação de proveniência. Deste modo, é importante salientar que os portais possuem conjuntos de dados que não disponibilizam informações sobre proveniência, sendo necessárias ações para complementar este indicador isso é perceptível visto que os valores apresentados poderiam estar no máximo 0,5. O Gráfico 7 apresenta os valores referentes aos indicadores de cada portal, sendo possível identificar quais portais se sobressaem.

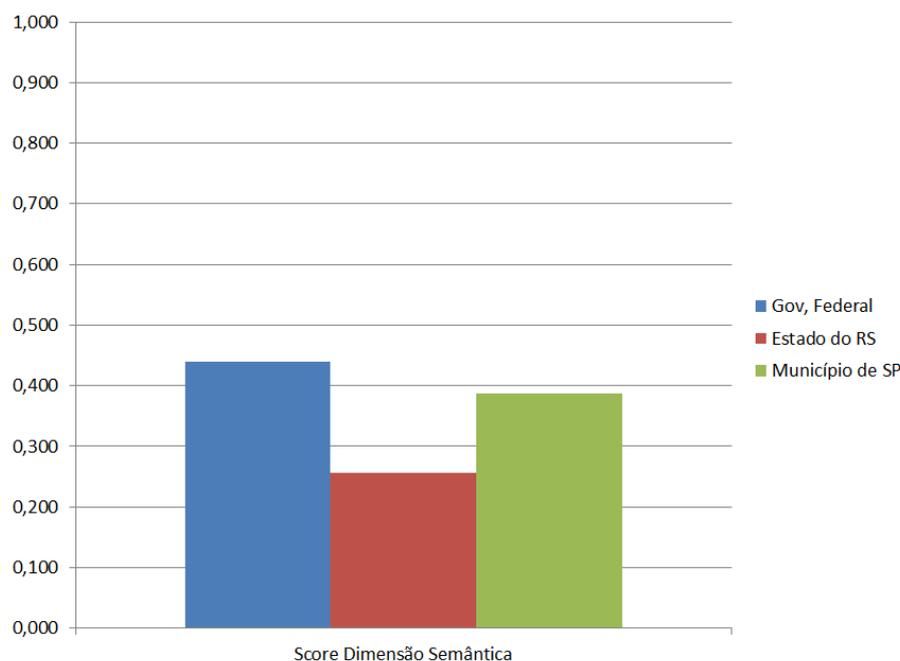
Gráfico 7 – Resultado indicadores – Dimensão Semântica



Fonte: do Autor

Em relação aos pesos, em todos os indicadores foi utilizado o peso 0,5 para apresentar os resultados, viabilizando a composição da dimensão semântica. O Gráfico 8 exibe o score obtido por cada portal em relação a dimensão semântica.

Gráfico 8 – Comparação do Score – Dimensão Semântica



Fonte: do Autor

Nesta dimensão percebe-se que os portais atendem parcialmente as recomendações, focando apenas em questões que envolvem a proveniência. No entanto, o uso dos vocabulários não são priorizados, mesmo quando a plataforma disponibiliza uma estrutura que auxilia no seu uso, como é o caso da extensão DCAT<sup>7</sup>. Ao utilizar vocabulários aumentam-se as possibilidades para o uso dos dados, visto que proporcionam a interoperabilidade com outros dados e o cruzamento dos dados.

#### 4.2.5 Análise dos resultados - Dimensão Estrutura

Esta dimensão tem como objetivo verificar aspectos relacionados à construção do portal de dados. A forma que apresenta-se o conteúdo pode influenciar na experiência dos *stakeholder*. A Tabela 9 apresenta os indicadores que compõem esta dimensão sobre os três portais de dados utilizados na aplicação do modelo.

<sup>7</sup> Mais informações disponível no endereço <<https://extensions.ckan.org/extension/dcat/>>

Tabela 9 – Aplicação modelo - Dimensão Estrutura

Dimensão Estrutura				
Portal	Indicadores	Valor	Peso Normalizado	Valor Ponderado
Federal	Pesquisa	1,000	0,111	0,111
	Conta Usuário	0,000	0,111	0,000
	Recomendação	0,000	0,111	0,000
	Acessibilidade	0,550	0,111	0,061
	Multi-Idioma	0,000	0,111	0,000
	Visualização	0,000	0,111	0,000
	Temática	0,016	0,111	0,002
	API   SPARQL	0,333	0,111	0,037
	Interatividade	0,625	0,111	0,069
	Valor Dimensão Estrutura – Gov. Federal			
Estado do RS	Pesquisa	1,000	0,111	0,111
	Conta Usuário	0,500	0,111	0,056
	Recomendação	0,000	0,111	0,000
	Acessibilidade	0,360	0,111	0,040
	Multi-Idioma	0,000	0,111	0,000
	Visualização	0,429	0,111	0,048
	Temática	0,951	0,111	0,106
	API   SPARQL	0,333	0,111	0,037
	Interatividade	0,500	0,111	0,056
	Valor Dimensão Estrutura – Gov. Estadual			
Município de SP	Pesquisa	1,000	0,111	0,111
	Conta Usuário	0,000	0,111	0,000
	Recomendação	0,000	0,111	0,000
	Acessibilidade	0,440	0,111	0,049
	Multi-Idioma	1,000	0,111	0,111
	Visualização	0,857	0,111	0,095
	Temática	0,779	0,111	0,087
	API   SPARQL	0,667	0,111	0,074
	Interatividade	0,375	0,111	0,042
	Valor Dimensão Estrutura – Gov. Municipal			

Fonte: do Autor

Em relação ao **Indicador Pesquisa** todos os portais obtiveram o valor máximo 0,111, que indica a presença de mecanismos de busca e refinamento.

Por sua vez, o **Indicador Conta Usuário** está presente parcialmente apenas no portal do **estado do Rio Grande do Sul** com o valor de 0,056 que indica a possi-

bilidade apenas de acesso (*login*), com ausência de cadastro de novos usuários. Os portais do **governo federal e do município de São Paulo** não possibilitam o *login* e cadastro de usuários.

Deste modo, percebe-se que neste indicador apenas um portal atende parcialmente as recomendações. Portanto, mesmo atendendo parcialmente as recomendações, só terá o acesso quem tiver o cadastro realizado pelo administrador do portal, dificultando o cadastro de novos *stakeholders*. A possibilidade de registro de novos usuários e o acesso, proporciona o acompanhamento das atividades do portal, notificação de novos conjuntos de dados disponibilizados entre outros benefícios.

Sobre o **Indicador Recomendação** nenhum portal de dados possui sistemas para recomendação de conjuntos de dados ou mesmo a sugestão de pesquisa junto aos mecanismos de busca. A literatura coloca que ao disponibilizar recomendações para os *stakeholders* instiga-se a exploração e o interesse pelo portal. Neste caso, é interessante ressaltar que a plataforma CKAN disponibiliza a extensão *discovery*<sup>8</sup> que contribui com este indicador, trazendo um conjunto de opções, por exemplo, sugestão de busca, conjuntos de dados similares, dentre outros.

No que tange o **Indicador Acessibilidade** o valor 0,061 corresponde ao portal do **governo federal**, aparece em segundo o portal do **município de São Paulo** com 0,049 e o valor 0,040 é para o portal do **estado do Rio Grande do Sul**. Deste modo, percebe-se que os portais atendem parcialmente os aspectos de acessibilidades recomendados pela *Web Content Accessibility Guidelines (WCAG) 2.1*<sup>9</sup>, visto que quanto mais próximo de 1 melhor será o portal em relação aos requisitos de acessibilidade<sup>10</sup>. Esta questão é importante pois possibilita a inclusão de pessoas com deficiência, fomentando a participação de qualquer indivíduo da sociedade a se beneficiar dos dados abertos conforme abordado na seção 2.2.

Em relação ao **Indicador Multi-Idioma** apenas o portal do **município de São Paulo** disponibiliza o acesso ao portal de dados em outros idiomas, além do idioma nativo, sendo atribuído o valor ponderado máximo de 0,111. Ao disponibilizar o portal de dados em outros idiomas, aumenta-se a abrangência do público que acessa, não se limitando a áreas locais e abrangendo para escala global. É importante salientar que neste indicador é verificado se o portal disponibiliza a sua estrutura em outro idioma, não é verificado se os conjuntos de dados ou os dados se encontram também em outro idioma, visto que a dimensão em análise é a de Estrutura.

A respeito do **Indicador Visualização** verifica-se quais os meios de visualização estão disponíveis, o portal do **município de São Paulo** possui o valor 0,095, seguido do portal do **estado do Rio Grande do Sul** com 0,048. O portal do **governo federal**

<sup>8</sup> Mais informações em <<https://extensions.ckan.org/extension/discovery/>>

<sup>9</sup> <https://www.w3.org/TR/WCAG21/>

<sup>10</sup> A WCAG 2.1 coloca uma nota de 0 a 10, contudo, por conta da normalização utilizada no modelo, o valor ficou entre 0 e 1.

não disponibiliza a visualização de dados *on-line* através da plataforma. Deste modo, quanto mais próximo do valor ponderado máximo 0,111, mais opções de visualização o portal possui.

Ao disponibilizar a visualização dos dados pela plataforma, permite-se aos *stakeholders* explorar e envolver-se com o conteúdo dos recursos, além de contribuir com aspectos de transparência e responsabilidade. Compreende-se que em portais que publicam dados de terceiros, como é o caso do governo federal, pode-se ter dificuldades na visualização *on-line* dos dados, por conta de que os dados estão alocados em outros portais. No entanto, salienta-se que sempre que possível os dados devem estar disponíveis para visualização conforme é mencionado na literatura.

O **Indicador Temática** do portal do **estado do Rio Grande do Sul** possui o valor 0,106, seguido do **município de São Paulo** com o valor 0,087 e com o valor 0,002 aparece o portal do **governo federal**. Baseando-se nos valores apresentados, pode-se entender que os portais possuem conjuntos de dados disponibilizados com ao menos um tema vinculado, mesmo que a maioria dos conjuntos de dados estejam sem temas, salvo apenas o portal do estado do Rio Grande do Sul que está próximo ao valor máximo.

O indicador temática proporciona o entendimento sobre a atribuição de temas (categorias ou grupos) aos conjuntos de dados. Entende-se que um conjunto de dados pode possuir um ou mais grupos vinculados, permitindo uma categorização e organização, bem como pode facilitar a exploração e a navegação por parte dos *stakeholders*. Desta forma, a literatura recomenda que o ideal é a atribuição de ao menos um grupo em cada conjunto de dados, deste modo, todos os conjuntos de dados devem possuir um grupo vinculado. Sendo assim, apenas o portal do Rio Grande do Sul está próximo do ideal, os demais possuem espaço para melhorias.

Referente ao **Indicador API|SPARQL** o portal do **município de São Paulo** possui o valor 0,074, os portais do **governo federal e do estado do Rio Grande do Sul** possuem o mesmo valor 0,037. Este indicador possibilita vislumbrar se o portal de dados disponibiliza o acesso aos dados publicados através de requisições feitas por agentes computacionais. Como pode ser observado, o portal do município de São Paulo se destaca, pois além de disponibilizar o acesso via API, também disponibiliza o acesso da documentação para o uso desta funcionalidade, diferentemente dos outros dois portais que disponibilizam apenas o acesso via API.

A disponibilização destes tipos de acesso é importante, e por isso salienta-se que os meios de acesso disponibilizados não são excludentes, ou seja, ao disponibilizar o acesso via API o portal não está isento de disponibilizar o acesso via SPARQL. O acesso via SPARQL é importante quando empregado o uso dos vocabulários, pois é o meio para executar consultas semânticas e usar interoperabilidade dos dados. No entanto, percebe-se que ao disponibilizar ao menos ao acesso via API todos os portais

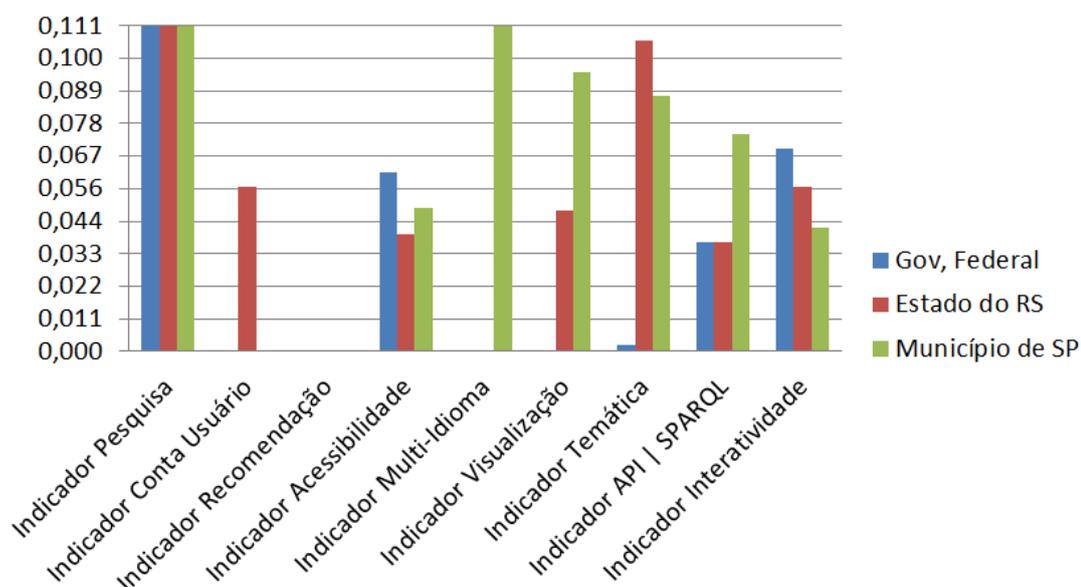
atendem parcialmente a este indicador.

Por fim, o **Indicador Interatividade** se destaca no portal do **governo federal** com o valor de 0,069, seguido do portal do **estado do Rio Grande do Sul** com 0,056 e com valor 0,042 aparece o portal do **município de São Paulo**. Este indicador propicia identificar quais portais disponibilizam meios para a interação entre *stakeholders* e o portal.

Desta forma, contribui-se com discussões sobre os temas, troca de ideias, podendo refletir em potenciais novos negócios ou mesmo um canal direto com a organização publicadora. Neste sentido, pode-se observar que os portais estão parcialmente alinhados com este indicador e salienta-se que as questões de interatividade são pouco exploradas pelas organizações.

O Gráfico 9 apresenta os valores referentes aos indicadores de cada portal, sendo possível identificar quais portais se sobressaem em relação a cada um dos indicadores.

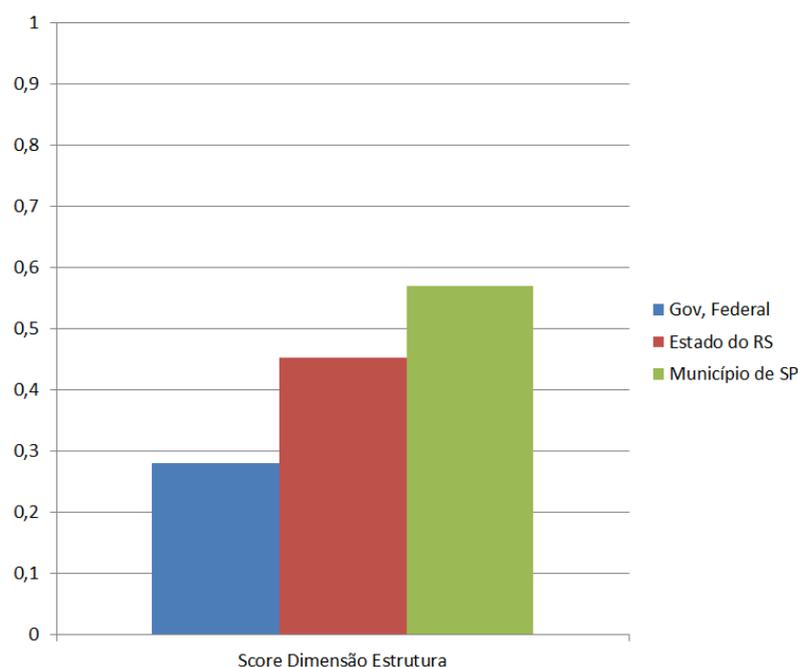
Gráfico 9 – Resultado indicadores – Dimensão Estrutura



Fonte: do Autor

Em relação aos pesos, para todos os indicadores foi utilizado o peso 0,111 de modo a possibilitar a composição da dimensão estrutura. O Gráfico 10 apresenta o comparativo entre os portais em relação a dimensão estrutura.

Gráfico 10 – Comparação do Score – Dimensão Estrutura



Fonte: do Autor

Nesta dimensão é possível observar que os portais do estado do Rio Grande do Sul e do município de São Paulo satisfazem alguns requisitos quanto a estrutura obtendo os maiores valores em relação aos indicadores. Por sua vez, o portal do governo federal é o que obteve maior valor em relação ao indicador acessibilidade e interatividade. Observa-se ainda a linearidade apenas no indicador pesquisa. No entanto, o indicador recomendação está ausente em todos os portais, visto que a plataforma CKAN dispõe de suporte através de extensões para esta funcionalidade.

Apesar das limitações encontradas, pode-se dizer que os portais atendem parcialmente as recomendações relacionadas a esta dimensão. No entanto, os portais podem realizar ações a fim de melhorar os resultados oriundos desta dimensão.

#### 4.2.6 Análise dos resultados - Dimensão Apresentação

Esta dimensão refere-se a questões que contribuem com o entendimento do portal, de modo a, explicitar informações simples e objetivas. É importante destacar que não foram analisados os textos disponibilizados pelos portais a fim de verificar se realmente as informações são relevantes sobre sua apresentação, contudo, foi apenas verificada a existência destes espaços. Para isso, a Tabela 10 apresenta os indicadores que compõem a dimensão.

Tabela 10 – Aplicação modelo - Dimensão Apresentação

Dimensão Apresentação				
Portal	Indicadores	Valor	Peso Normalizado	Valor Ponderado
Federal	Compreensão	1	0,500	0,500
	Quantitativos	0,444	0,500	0,222
	Valor Dimensão Apresentação - Gov. Federal			0,722
Estado do RS	Compreensão	0,667	0,500	0,333
	Quantitativos	0,333	0,500	0,167
	Valor Dimensão Apresentação - Gov. Estadual			0,500
Município de SP	Compreensão	0,667	0,500	0,333
	Quantitativos	0,444	0,500	0,222
	Valor Dimensão Apresentação - Gov. Municipal			0,556

Fonte: do Autor

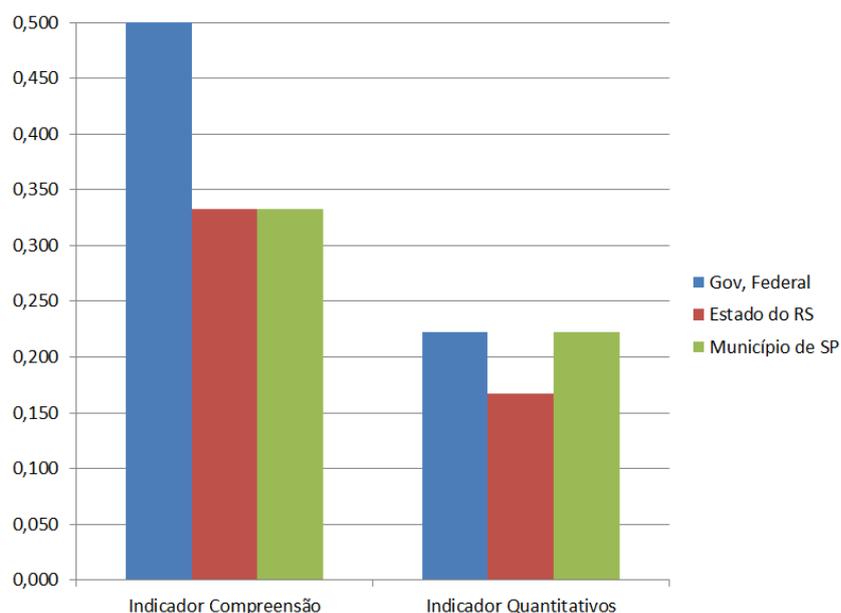
Para o **Indicador Compreensão** os portais do estado do Rio Grande do Sul, município de São Paulo e do governo federal obtiveram os respectivos valores 0,333, 0,333 e 0,500, isso indica que os portais divulgam informações que contribuem com o entendimento dos *stakeholders* sobre objetivos do portal, sua periodicidade em relação às publicações, os projetos que fizeram uso dos dados publicados, são algumas das informações que podem ser divulgadas.

Por sua vez, o **Indicador Quantitativo** em relação aos portais do **governo federal e do município de São Paulo** aparecem com o valor 0,222, seguido do portal do **estado de Rio Grande do Sul** que possui o valor 0,167. Este indicador refere-se às informações quantitativas divulgadas pelo portal de dados, como a quantidade de conjuntos de dados e recursos disponibilizados ou se essas informações estão em algum quadro visível, são alguns dos itens que compõem o indicador.

A literatura não apresenta um mínimo ou máximo sobre este tipo de divulgação, apenas informa que é recomendado disponibilizar ao menos uma destas informações quantitativas.

O Gráfico 11 apresenta os valores referentes aos indicadores de cada portal, sendo possível identificar quais portais se sobressaem em relação a um indicador.

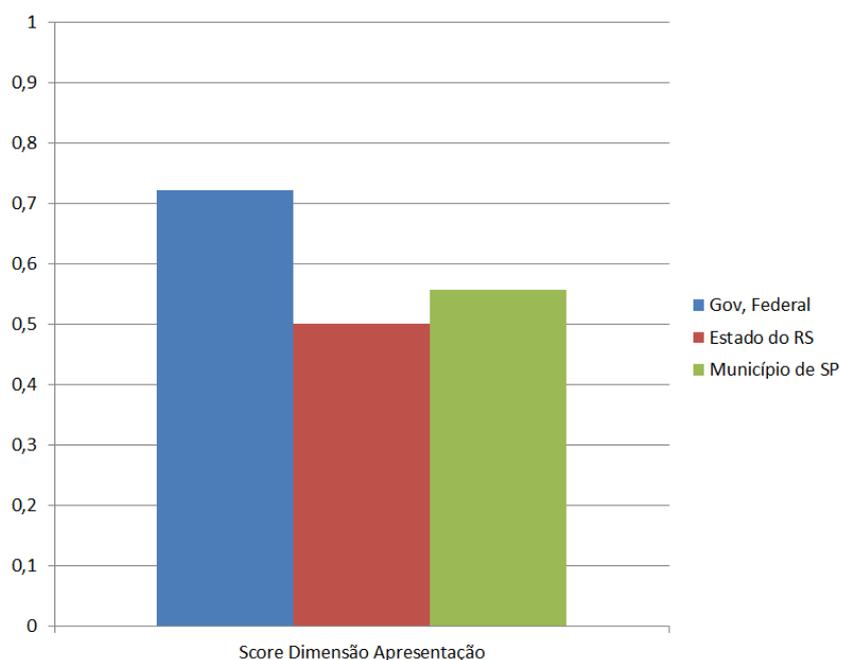
Gráfico 11 – Resultado indicadores – Dimensão Apresentação



Fonte: do Autor

Em relação aos pesos, em todos os portais foi utilizado o peso 0,500 e assim possibilitando a composição da dimensão estrutura, sendo apresentados pelo Gráfico 12 o score obtido por cada um dos portais referentes a esta dimensão.

Gráfico 12 – Comparação do Score - Dimensão Apresentação



Fonte: do Autor

Nesta dimensão, pode-se perceber que os portais atendem parcialmente as recomendações, visto que a literatura não define mínimos ou máximos. Contudo, ressalta-

se que tanto o indicador compreensão e quantitativos contribuem com o contexto de dados abertos.

#### 4.2.7 Análise dos resultados - Dimensões e Score Final

Nesta seção aborda-se o score de cada dimensão, de modo a compor o score final sobre os portais. Apresenta-se o valor do score final na escala de 0 a 1 e de 0 a 10, além da diferença entre o valor do score final e do score meta definido. O score meta utilizado para a aplicação do modelo foi o valor 5. A Tabela 11 apresenta todas as seis dimensões correspondentes aos três portais de dados.

Tabela 11 – Aplicação modelo - Dimensões

Dimensões				
Portal	Dimensão	Valor	Peso Normalizado	Valor Ponderado
Federal	Acesso	0,461	0,167	0,077
	Formatos	0,398	0,167	0,066
	Direitos	0,206	0,167	0,034
	Semântica	0,440	0,167	0,073
	Estrutura	0,281	0,167	0,047
	Apresentação	0,722	0,167	0,120
	Score Final (Escala 0 - 1)			
Score Final (Escala 0 - 10)				4,18
Score Meta				5
Diferença Score Final X Score Meta				0,82
Estado do RS	Acesso	0,396	0,167	0,066
	Formatos	0,498	0,167	0,083
	Direitos	0,035	0,167	0,006
	Semântica	0,256	0,167	0,043
	Estrutura	0,453	0,167	0,075
	Apresentação	0,500	0,167	0,083
	Score Final (Escala 0 - 1)			
Score Final (Escala 0 - 10)				3,56
Score Meta				5
Diferença Score Final X Score Meta				1,44
	Acesso	0,464	0,167	0,077
	Formatos	0,394	0,167	0,066
	Direitos	0,250	0,167	0,042
	Semântica	0,387	0,167	0,065
	Estrutura	0,569	0,167	0,095
	Apresentação	0,556	0,167	0,093

Município de SP	Score Final (Escala 0 - 1)	0,436
	Score Final (Escala 0 - 10)	4,36
	Score Meta	5
	Diferença Score Final X Score Meta	0,64

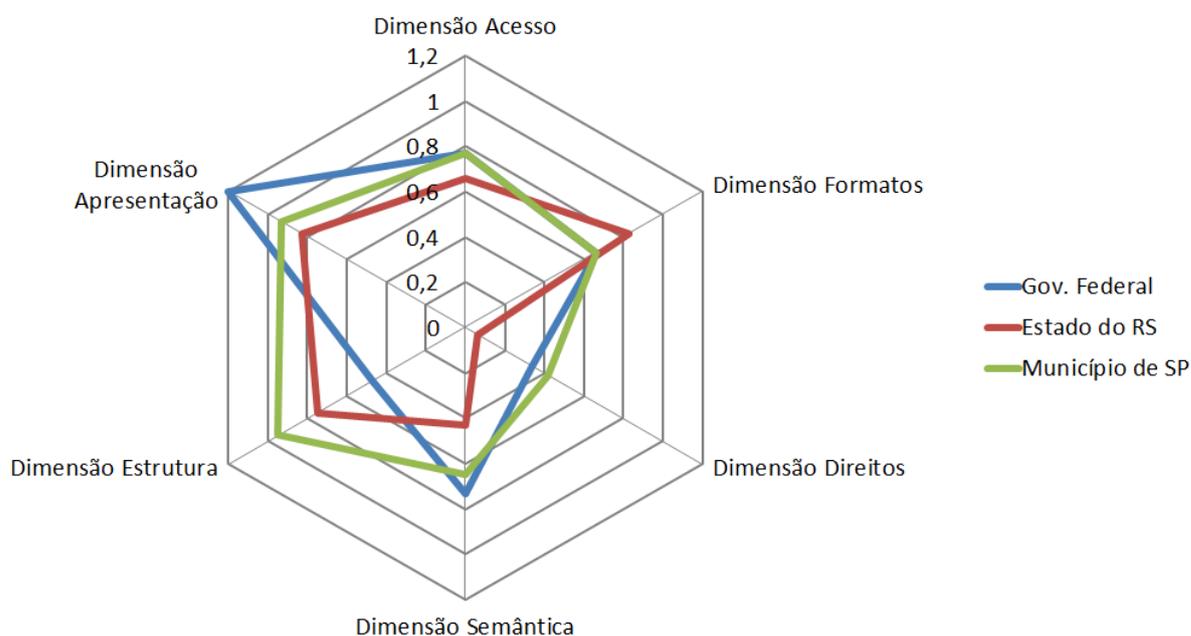
Fonte: do Autor

Observa-se na Tabela 11 que dentre os três portais, o portal do **município de São Paulo** é o que mais se destaca em relação ao score final 4,36, seguido do portal do **governo federal** com o score final 4,18 e com o menor score final 3,56 o portal do **estado do Rio Grande do Sul**.

Percebe-se que o portal do município de São Paulo por mais que seja pequeno em termos de conjuntos de dados comparado ao portal do governo federal e do estado do Rio Grande do Sul, se destaca conforme cenário apresentado previamente.

Isso pode ocorrer por alguns motivos, por exemplo: por ser um município (nível de governo menor) este pode ter maior controle no processo de publicação e disponibilização de dados, mais pessoas com capacidade técnicas envolvidas com a disponibilização dos dados, dentre outros. O Gráfico radar 13 permite visualizar as dimensões que se destacam em relação a cada portal. Salienta-se que os valores apresentados estão em escala de 0 à 10. Todavia o gráfico está representando a escala 0 à 1,2, isso porque os valores obtidos em cada dimensão são baixos e ao representar o gráfico em maior escala, seria difícil perceber as diferenças entre os portais.

Gráfico 13 – Resultado Dimensões



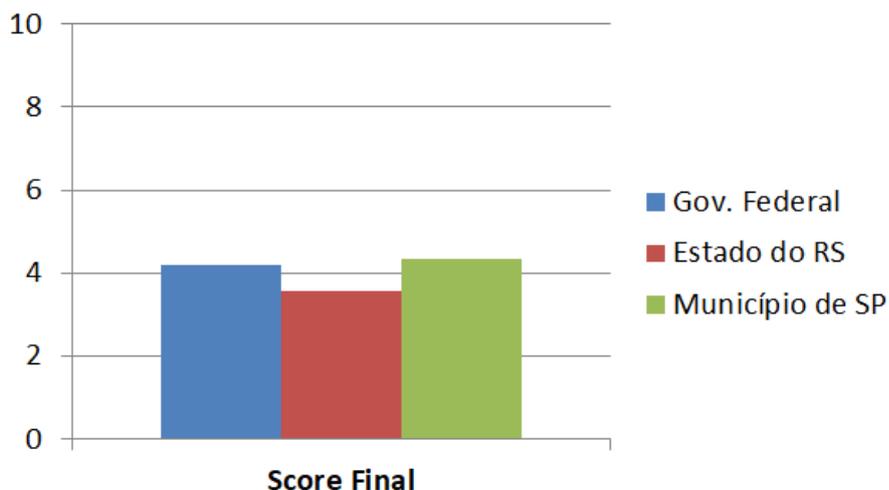
Fonte: do Autor

Nota-se no Gráfico 13 que cada portal se destaca em alguma dimensão em específico, como exemplo, o portal do município de São Paulo tem um bom score nas dimensões direitos e estrutura, além de estar empatado praticamente com o portal do governo federal nas dimensões acesso e formatos.

Por sua vez, o portal do governo federal se destaca nas dimensões semântica e de apresentação. Vale ressaltar que o portal do governo federal mesmo não fazendo uso dos vocabulários se destacou por conta do volume de conjuntos de dados disponibilizados que possuem informações de proveniência. Em relação ao portal do estado do Rio Grande do sul, este destaca-se na dimensão formatos, no entanto, possui score baixo quase zero em relação à dimensão direitos, isso porque cerca de 85% dos conjuntos de dados cadastrados não possuem licença atribuída ou definida, como pode ser observado no Gráfico 32.

O Gráfico 14 permite comparar o score final dos portais. Observa-se o portal do município de São Paulo com o score final 4,36, seguido do portal do governo federal com score 4,18 e por último o portal do estado do Rio Grande do Sul com score 3,56.

Gráfico 14 – Resultado Score Final – (Escala de 0 – 10)



Fonte: do Autor

Assim constata-se que as organizações responsáveis pelos portais de dados realizarem algumas adequações, poderão atingir a meta ou mesmo ultrapassar. O emprego da meta, possibilita que as organizações possam ir definindo pequenos objetivos a fim de manter uma constante evolução ou mesmo possibilitar que determinem um valor de referência para que possam manter um nível de qualidade no portal de dados.

#### 4.3 AVALIAÇÃO COM ESPECIALISTAS

Com o intuito de avaliar o modelo e o sistema computacional foi realizada uma pesquisa de campo envolvendo um grupo de especialistas que atuam no âmbito de dados abertos, sendo descrito o perfil destes especialistas na seção 1.8.3.5. Para mais informações verificar a seção 1.8.

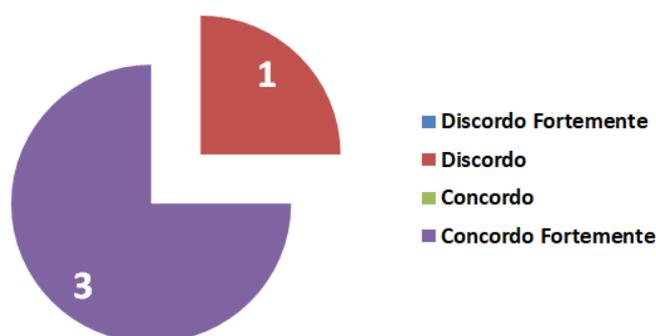
Com o objetivo de realizar a avaliação foram convidados 30 especialistas, mas somente 4 se dispuseram a avaliar por intermédio do questionário. É importante destacar que acredita-se que por conta da pandemia de Covid-19 não obteve-se maior número de participantes, por reflexo da sobrecarga de trabalho neste período. Este fato foi informado por participantes que justificaram sua não participação. Para auxiliar na avaliação foram elaborados vídeos para apresentar a proposta, o sistema computacional e também foram elaborados materiais extras e enviados por e-mail.

O questionário (Apêndice D) foi definido em dois critérios: Critério 1- Modelo de Referência e Critério 2- Sistema Web, além de uma pergunta descritiva aberta para que os participantes pudessem expor sua percepção sobre o trabalho ou mesmo deixar sugestões, elogios ou críticas.

### 4.3.1 Análise das Respostas às Perguntas do Critério 1

Neste primeiro critério são apresentadas as seis questões sobre o modelo de referência. Na questão um os respondentes foram questionados se o modelo proposto é adequado para avaliar a aplicação das boas práticas de disponibilização em portais de dados abertos. O Gráfico 15 apresenta as respostas obtidas através do formulário.

Gráfico 15 – Respostas critério 1 – Questão 1

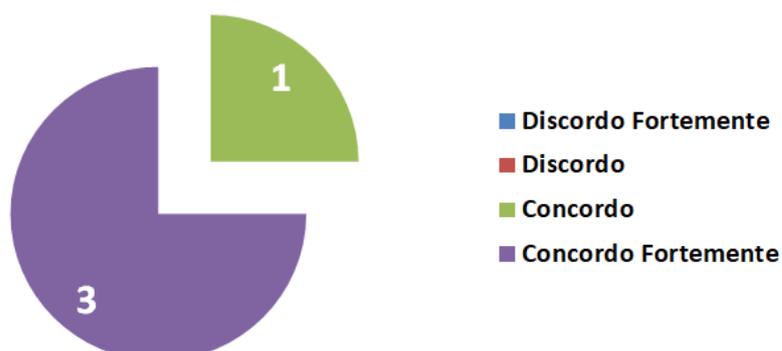


Fonte: do Autor

Dos especialistas que responderam o questionário apenas um profissional discordou deste questionamento, sendo que os demais concordaram fortemente que o modelo é adequado para avaliar a aplicação das boas práticas de disponibilização de dados em portais. Ao discordar o profissional destacou que são alguns pontos que merecem uma reflexão devido a complexidade do tema estudado. Ao final desta seção estes pontos serão evidenciados.

Para a questão dois os respondentes foram questionados se o modelo proposto possibilita identificar deficiências relacionadas às boas práticas de disponibilização em portais de dados abertos. O Gráfico 16 apresenta as respostas obtidas através do formulário.

Gráfico 16 – Respostas critério 1 – Questão 2

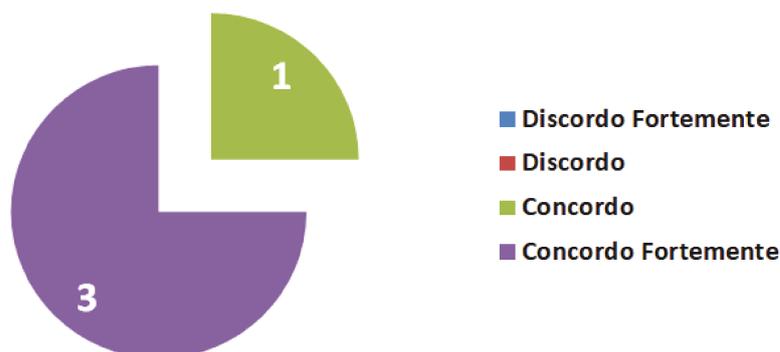


Fonte: do Autor

Nesta questão um participante concorda e os demais concordam fortemente que o modelo possibilita identificar deficiências no portal de dados em relação as boas praticas.

Na questão três os respondentes foram questionados se o modelo proposto possui uma estrutura (dimensões, indicadores e variáveis) consistente e compreensível. O Gráfico 17 apresenta as respostas obtidas através do formulário.

Gráfico 17 – Respostas Critério 1 – Questão 3

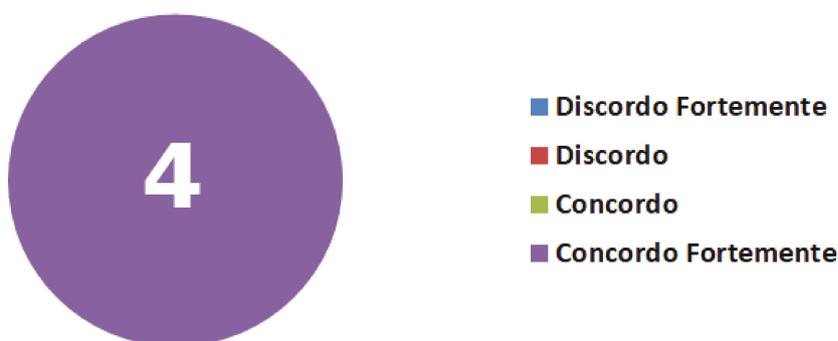


Fonte: do Autor

Observa-se que no Gráfico 17 que apenas um participante concorda e os demais concordam fortemente que o modelo possui uma estrutura compreensível.

Na questão 4 os respondentes foram questionados se o modelo proposto pode ser aplicado em portais de dados abertos no âmbito federal, estadual, municipal e outras organizações. O Gráfico 18 apresenta as respostas obtidas através do questionário.

Gráfico 18 – Respostas Critério 1 – Questão 4

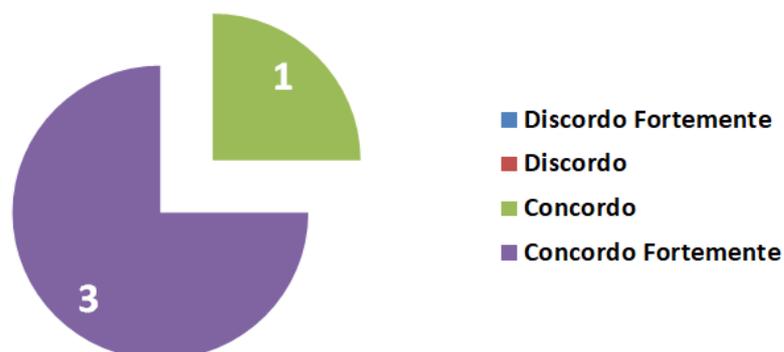


Fonte: do Autor

Percebe-se que todos os participantes concordam fortemente que o modelo pode ser aplicado independente do tamanho do portal e do nível de governo.

Para a questão cinco os respondentes foram questionados se o modelo proposto auxilia na qualidade dos portais de dados abertos. O Gráfico 19 apresenta as respostas obtidas através do questionário.

Gráfico 19 – Respostas Critério 1 - Questão 5

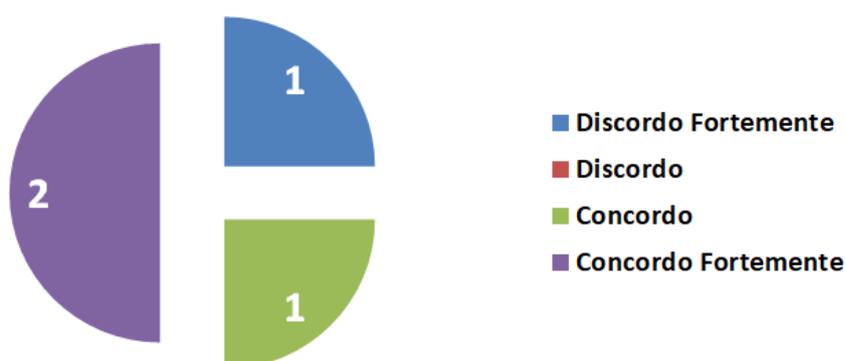


Fonte: do Autor

Na questão cinco apenas um participante concorda e os demais concordam fortemente que o modelo pode auxiliar na melhora da qualidade dos portais de dados.

Referente à questão seis os respondentes foram questionados se o modelo proposto requer conhecimento especializado para sua aplicação. O Gráfico 20 apresenta as respostas obtidas através do questionário.

Gráfico 20 – Respostas Critério 1 - Questão 6



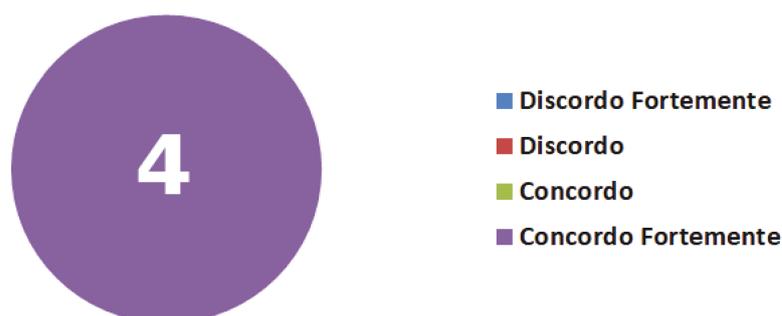
Fonte: do Autor

No Gráfico 20 um respondente discordou fortemente em relação à necessidade de possuir o conhecimento especializado para aplicar o modelo, outro participante apenas concorda e os outros dois concordam fortemente. Pode ser observada uma discrepância sobre a percepção dos especialistas, onde na percepção do analista de TI o modelo pode ser aplicado por qualquer pessoa, por sua vez o administrador concorda que deve se ter um conhecimento mínimo sobre o assunto e por fim o professor pesquisador e o auditor acreditam que um conhecimento especializado é necessário para se ter uma melhor percepção dos resultados.

### 4.3.2 Análise das Respostas às Perguntas do Critério 2

Neste segundo critério são apresentadas as quatro questões sobre o sistema computacional desenvolvido. Na questão um os respondentes foram questionados se a criação de um sistema computacional para automatização da coleta, avaliação e análise contribui para o uso do modelo proposto. O Gráfico 21 apresenta o as respostas obtidas através do questionário.

Gráfico 21 – Respostas Critério 2 - Questão 1

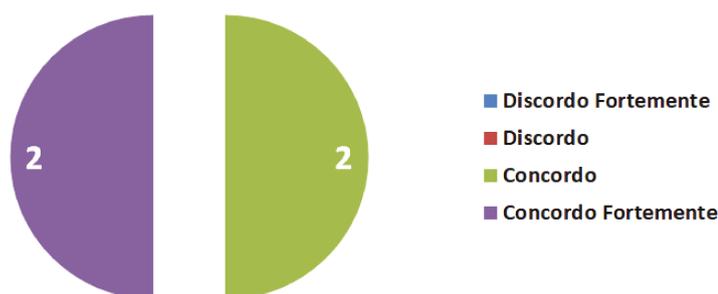


Fonte: do Autor

Percebe-se que todos os participantes concordam fortemente que o uso de um software para automatizar a coleta e avaliação contribuiu significativamente na aplicação e na utilização do modelo.

Na questão dois os respondentes foram questionados se o uso de padrões de ponderação de avaliação dos portais atende às particularidades dos diferentes portais de dados. O Gráfico 22 apresenta as respostas obtidas através do questionário.

Gráfico 22 – Respostas Critério 2 - Questão 2



Fonte: do Autor

Como observado no Gráfico 22 dois especialistas concordam e dois concordam fortemente que ao utilizar os padrões é possível atender as particularidades dos portais de dados, ou seja, possibilita avaliar diferentes cenários.

A questão três aborda o questionamento se o sistema computacional apresenta interação adequada com o usuário para avaliar portais de dados. O Gráfico 23 apresenta o as respostas obtidas através do questionário.

Gráfico 23 – Respostas Critério 2 - Questão 3

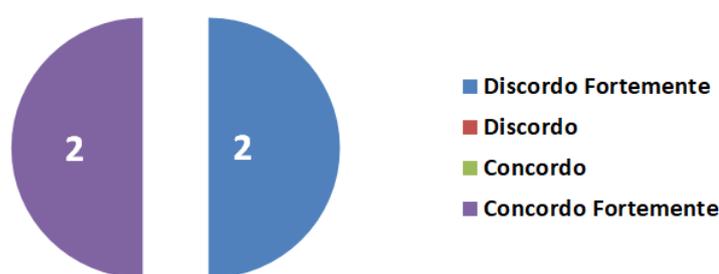


Fonte: do Autor

É possível observar que dois especialistas concordam e dois concordam fortemente que o sistema computacional desenvolvido possibilita uma interação adequada com o usuário favorecendo o seu uso.

Na questão quatro os respondentes foram questionados se concordam em utilizar o sistema computacional para avaliar o portal de dados de sua organização. O Gráfico 24 apresenta as respostas obtidas através do questionário.

Gráfico 24 – Respostas Critério 2 - Questão 4



Fonte: do Autor

Nesta questão dois especialistas discordam fortemente e dois concordam fortemente em relação a utilizar o sistema desenvolvido. É importante salientar que os especialistas que discordaram fortemente optaram por esta resposta, pois a sua organização não possui um portal de dados para utilizar aplicação desenvolvida. O portal que estes especialistas têm contato é do seu nível de governo não do setor ao qual estes estão vinculados.

Em relação a pergunta descritiva "Deixe sua sugestão, crítica ou elogio para o autor", o auditor disse:

"Muito interessante o trabalho. A gestão por indicadores é fundamental para a melhoria do serviço público. Também é um diferencial importante poder comparar portais públicos e privados, nacionais ou estrangeiros, usando a mesma ferramenta."

Por sua vez o analista de TI comentou:

"A questão que está sendo pesquisada é muito importante e os resultados podem ser bastante úteis para propor melhorias aos portais. Embora também seja um enorme desafio. O software de coleta viabiliza que análises e avaliações sejam feitas. Parabéns pelo trabalho! Ainda é necessário, entretanto, melhorar as definições e ter atenção às formas de coleta de algumas das variáveis, pois podem estar medindo outra coisa senão a que se propõem."

O administrador disse:

"Acredito que o modelo proposto vem a trazer uma importante contribuição no campo dos estudos relacionados à transparência pública e em especial aos estudos relativos à disponibilização de dados em formato aberto no âmbito das instituições brasileiras. As dimensões, indicadores e variáveis dão conta de abranger os diferentes aspectos à disponibilização e qualidade dos dados das instituições.

Ainda como contribuição relevante, a implementação do modelo via aplicação, em especial a automação da verificação dos critérios de análise é bastante relevante, tanto no que tange à possibilitar a avaliação (uma vez que em termos de recursos, os portais podem ter milhares deles o que inviabiliza a análise de forma manual), quanto no que tange a ter um processo de avaliação padronizado, não influenciado pela visão de mundo do avaliador (risco que existe no caso de a avaliação ser feita manualmente).

Parabenizo o doutorando pelo trabalho, muito bem construído, tanto do ponto de vista de critérios de avaliação quanto do ponto de vista de forma de implementação via aplicação. Em especial, para viabilizar a utilização deste tipo de modelo de análise, sua implementação em formato de aplicação é de suma importância para que seja possível o arquivamento de dados históricos e facilitar o processo de avaliação e análise dos resultados por parte dos usuários."

O professor pesquisador comentou:

"Parabenizo o doutorando pelo trabalho, pois vejo um modelo que pode contribuir de maneira muito significativa na promoção da qualidade dados abertos, além de fornecer um ferramental com muito potencial para a pronta utilização pelas organizações públicas. Considero um trabalho muito relevante, e que merece ganhar visibilidade, após sua validação e defesa."

Pode-se observar pelas respostas coletadas que o modelo, o sistema foram bem vistos pelos especialistas, sendo que em apenas três questões foram obtidas respostas do tipo: discordo ou discordo totalmente.

Vale salientar que a resposta discordo totalmente em relação a questão quatro do segundo critério (Gráfico 24) pode ser reflexo da pergunta mal formulada, visto as justificativas informadas pelos respondente.

É importante destacar que em relação à questão um do primeiro critério abordado no Gráfico 15, um profissional questionou por e-mail informando sobre um ponto que mais lhe chamou a atenção. Neste caso deu-se ênfase ao questionamento, visto que este profissional atua como analista de TI na esfera federal e foi um dos envolvidos na iniciativa de abertura dos dados em 2009 no Brasil, além de atuar na estruturação do portal de dados do governo federal.

O especialista por conhecer a trajetória do portal, informou que a variável "Conjuntos de dados adicionados nos últimos doze meses" estaria trazendo os dados incorretos, visto que dos cerca de 10 mil conjuntos de dados coletados pelo sistema, destes 6421 teriam sido adicionados nos últimos 12 meses. No entanto o questionamento surgiu por conta de uma publicação feita por Infraestrutura Nacional de Dados Abertos (INDA) em 2018<sup>11</sup>, que informa que haviam cerca de 6278 conjuntos de dados cadastrados na época.

Neste sentido, iniciou-se uma investigação sobre a questão, observando que o banco de dados do sistema desenvolvido possui dados coletados do portal do governo federal desde Janeiro de 2019, quando teve início o desenvolvimento do sistema e a última coleta efetuada em Abril de 2021. Deste modo, foi possível realizar uma comparação da adição de novos conjuntos de dados nesta faixa temporal. A Tabela 12 apresenta os conjuntos de dados adicionados com base nas coletas realizadas durante o período de 2019 a 2021.

Tabela 12 – Quantitativo de conjuntos de dados adicionados, obtidos através das coletas realizadas no portal do governo federal

Ano de Referência	Data Coleta			
	Jan de 2019	Dez de 2020	Fev de 2021	Abr de 2021
2021	-	-	333	703
<b>2020</b>	-	6065	<b>6135</b>	<b>5600</b>
<b>2019</b>	<b>1184</b>	<b>1203</b>	<b>1189</b>	<b>904</b>
<b>2018</b>	<b>5078</b>	<b>1952</b>	<b>1948</b>	<b>1883</b>
<b>2017</b>	<b>269</b>	<b>297</b>	<b>267</b>	267
2016	123	123	123	133
<b>2015</b>	<b>635</b>	<b>635</b>	<b>635</b>	<b>626</b>
2014	135	134	134	133
2013	38	38	38	36
2012	44	39	39	41
2011	12	10	10	12

Fonte: do Autor

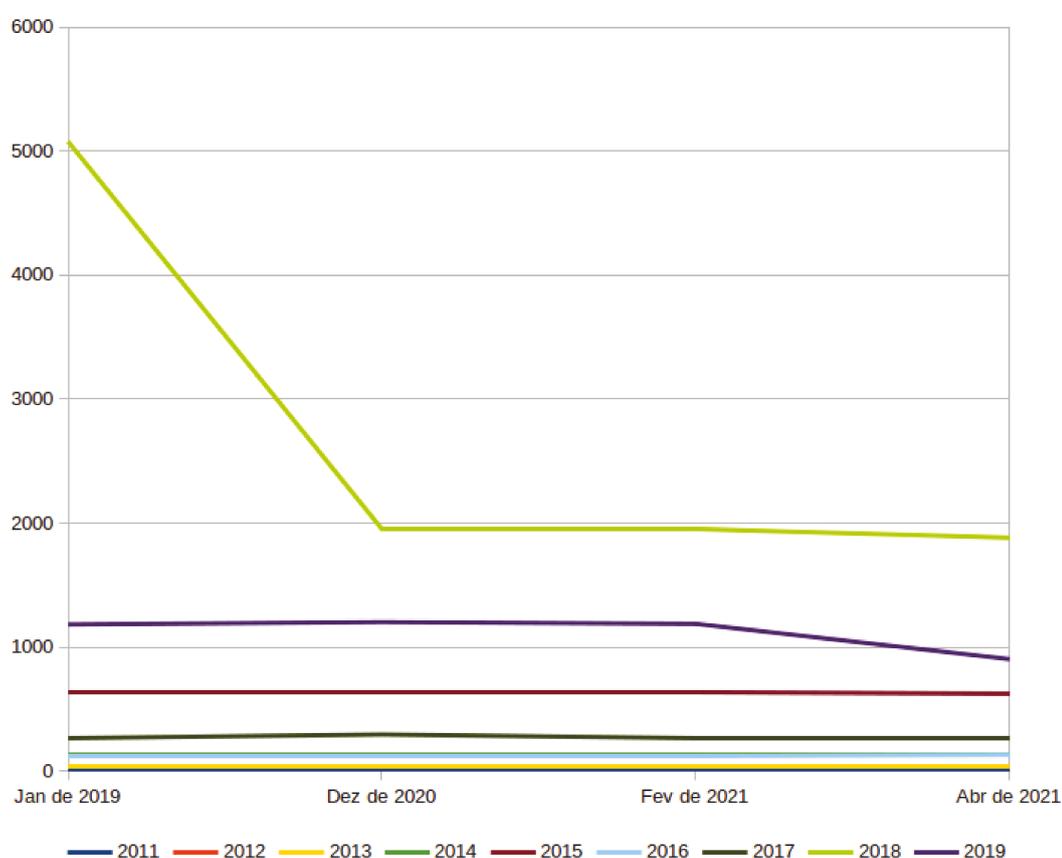
Observa-se que nos anos 2015, 2017, 2018, 2019 e 2020 (destacados em negrito) o portal apresenta oscilações na quantidade de conjuntos de dados adicionados, possibilitando o entendimento que alguns conjuntos de dados foram apagados. Ao apagar um conjunto de dados do portal, também são apagados os recursos a ele

<sup>11</sup> Publicado em <[https://wiki.dados.gov.br/Balanco-da-INDA-em-2018.ashx#\\_\\_Métricas\\_do\\_portal\\_1](https://wiki.dados.gov.br/Balanco-da-INDA-em-2018.ashx#__Métricas_do_portal_1)>

vinculados, deste modo, o portal do governo federal infringe uma das recomendações, a recomendação acerca da preservação dos dados.

A exclusão dos dados definitiva ou mesmo a realocação dos dados, devem ser informadas para os *stakeholders* através dos código de resposta HTTP conforme boa prática 27 apresentada na seção 2.8. Contudo, deve-se preservar a URI que foi criada para estes conjuntos de dados, visto que podem haver *stakeholder* que fazem uso destes conjuntos de dados. O Gráfico 25 evidencia a queda dos quantitativos dos conjuntos de dados por anos analisados.

Gráfico 25 – Queda no quantitativo de conjuntos de dados disponibilizados ao longo dos anos



Fonte: do Autor

Deste modo, foi informado ao profissional o ocorrido e constatado que tanto o modelo como o sistema desenvolvido estão corretos, sendo que foi verificada esta situação em outros portais e observado que esta inconsistência apenas ocorre no portal do governo federal. Este caso relatado pelo especialista traz uma reflexão sobre a relação tamanho do portal e quantidade de pessoas envolvidas no processo de disponibilização dos dados, emergindo algumas questões de pesquisa que serão abordadas na seção 5.2 sobre trabalhos futuros.

#### 4.4 APLICAÇÃO DO MODELO EM DIVERSOS PORTAIS

Esta seção tem como objetivo apresentar outros portais que também foram coletados e aplicados o modelo a fim de examinar o comportamento de toda a estrutura elaborada. Deste modo, estes não serão abordados em detalhes, mas apenas evidenciados.

A seguir são apresentadas as informações sobre os portais, sendo possível comparar a diferença entre eles, perceber quais organizações obtiveram score final próximo ao score meta e conseqüentemente apresentam maior conformidade com relação às recomendações, abordadas nesta pesquisa. Destaca-se que foi utilizado padrão **DEFAULT** para a realização destas avaliações.

A Tabela 13 apresenta os 25 portais nacionais e internacionais que foram coletados e aplicados o modelo, sendo apresentados em ordem decrescente de acordo com o score final.

Tabela 13 – Demais Portais Coletados

Portal	Nível de Governo	Qtd. Conjuntos de Dados Coletados	Qtd. Recursos Coletados	Data de Coleta	Score Final
Governo Uruguai	Federal	2258	10303	15/12/2020	4,91
Governo Equador	Federal	128	2583	21/12/2020	4,63
Governo da Romênia	Federal	2334	26434	22/5/2021	4,57
Município de Recife - PE	Municipal	119	851	26/1/2021	4,55
Estado de São Paulo	Estadual	292	568	29/1/2021	4,31
Estado de Alagoas	Estadual	300	3376	29/1/2021	4,28
Município de Boston - EUA	Municipal	303	1918	20/5/2021	4,25
Município de Natal - RN	Municipal	33	101	26/1/2021	4,21
Tribunal de Contas do Estado - RS	Estadual	51519	133470	20/5/2021	4,21
Município de Belo Horizonte - BH	Municipal	61	170	5/2/2021	4,17

Governo da Argentina	Federal	1019	4746	22/5/2021	4,09
Estado do Mato Grosso do Sul	Estadual	28	437	6/2/2021	4,08
Governo Irlanda	Federal	10396	29676	25/1/2021	4,07
Banco Central Brasil	Federal	3237	13049	21/4/2021	4,02
Município de Goiânia - GO	Municipal	29	150	6/2/2021	3,97
Estado de Minas Gerais	Estadual	7	117	26/1/2021	3,94
Estado do Espírito Santo	Estadual	48	694	29/1/2021	3,91
Distrito Federal	Estadual	162	801	21/4/2021	3,84
Município de Porto Alegre - RS	Municipal	53	137	19/12/2020	3,67
Governo Chile	Federal	3961	10871	18/12/2020	3,66
Estado do Illinois - EUA	Estadual	599	898	22/5/2021	3,22
Instituto Federal Catarinense	Federal	38	71	28/4/2021	3,09
Governo Islandia	Federal	112	664	25/1/2021	2,69
Município de Fortaleza - CE	Municipal	281	1252	20/5/2021	2,44

Fonte: do Autor

Como pode ser percebido o *score final* na Tabela 13 independe da quantidade de conjunto de dados existentes nos portais que foram apresentados, por exemplo, têm-se o portal Uruguai com o score 4,91 e 2558 conjuntos de dados, enquanto o portal do Tribunal de Contas do Estado do Rio Grande do Sul, possui um score 4,21, sendo que este possui 51519 conjuntos de dados. Deste modo, é notório que o modelo pode ser aplicados em diferentes cenários, mas quando considerandos os mesmos padrões, podem se ter comparativos entre os portais, independentemente do seu tempo de existência.

Vale ressaltar que cada portal deve ser avaliado considerando o objetivo da organização e o que se almeja com a disponibilização dos dados, a fim de possibilitar uma avaliação precisa. Para isso, a estrutura elaborada permite através do uso dos padrões flexibilizar o modelo de referência de modo que a organização priorize as di-

mensões, indicadores e variáveis de acordo com a sua necessidade. Todas as medidas utilizadas por esta pesquisa serviram de base para compreender e analisar a estrutura do modelo e do sistema computacional, mas observa-se que as possibilidades de avaliação e análise podem ser exploradas em pesquisas futuras.

Destaca-se ainda, que foi possível coletar e avaliar portais de dados internacionais. Contudo, deparou-se com algumas dificuldades como: de idiomas, da dinamicidade que os portais em relação a sua interface e sobre questões de organização dos dados específicas de cada país. Neste sentido, o sistema foi adaptado para atender o maior número de países possíveis, mas percebe a necessidade de futuras atualizações que serão expostas na seção 5.2 de trabalhos futuros.

#### 4.5 MONITORAMENTO DO PORTAL DE DADOS DO ESTADO DE SANTA CATARINA

Nesta seção serão abordados os dados referentes ao monitoramento realizado no período de dezembro de 2020 até junho de 2021 no portal do estado de Santa Catarina, pois o objetivo desta seção é apresentar outro potencial de uso da pesquisa desenvolvida.

Denominou-se monitoramento pois este termo é amplamente utilizado em diversos segmentos de TI com o objetivo de analisar e atentar-se a possíveis sinais de anormalidades em um determinado cenário. Quando identificadas anormalidades, isso pode indicar que estão ocorrendo problemas.

Neste sentido, ao monitorar portais de dados pode-se identificar se estão ocorrendo publicações constantes, se os dados disponibilizados estão seguindo as recomendações ou mesmo se o portal está ativo. Deste modo, ao perceber que não há atividades no portal, os responsáveis podem averiguar o que está ocorrendo, sendo útil o monitoramento principalmente em grandes organizações que possuem órgãos parceiros que publicam em seu portal ou mesmo um grande volume de pessoas envolvidas no processo de publicação, onde controlar o processo de publicação é difícil.

A Tabela 14 apresenta os dados coletados em diferentes datas para a realização do monitoramento de um determinado período do portal do estado de Santa Catarina. O Quadro 14 é composto pelas colunas **D1)** Dimensão Acesso, **D2)** Dimensão Formatos, **D3)** Dimensão Direitos, **D4)** Dimensão Semântica, **D5)** Dimensão Estrutura, **D6)** Dimensão Apresentação e o **SF)** Score Final. Salienta-se que para todas as avaliações foi utilizado do padrão **DEFAULT**.

Tabela 14 – Monitoramento Portal de Santa Catarina

Data	Qtd. Conjuntos de Dados	Qtd. Recursos	D1	D2	D3	D4	D5	D6	SF
15/12/20	100	826	0,087	0,078	0,031	0,062	0,107	0,056	4,20
29/01/21	103	874	0,086	0,078	0,032	0,061	0,107	0,056	4,19
28/04/21	105	972	0,082	0,079	0,032	0,061	0,099	0,056	4,08
10/06/21	106	1007	0,081	0,079	0,032	0,061	0,100	0,056	4,08

Fonte: do Autor

Observa-se na Tabela 14 que na faixa temporal que ocorreram as coletas, não houve a disponibilização de um número expressivo de conjuntos de dados e recursos. Deste modo, observar-se que as colunas **D1**, **D4** e **D5** baixaram o seu valor, contudo houve um aumento de 0.001 na **D2** e **D3** que corresponde aos novos recursos e sobre as licenças utilizadas em novos conjuntos de dados. A coluna **D6** se manteve com os mesmos valores, até a data 10/06/2021. Observou-se por sua vez que na coleta realizada em 10/06/2021, a coluna **D1** baixou 0,001, mas a coluna **D5** aumentou 0,001, as demais colunas não tiveram alterações.

Assim nota-se na coluna **SF** que os valores também estão baixando, demonstrando que o portal não possui uma periodicidade na disponibilização dos dados, entre outros fatores que impactaram no **SF**. Salienta-se que na última coleta realizada em 10/06/2021 o valor de **SF** permaneceu o mesmo, visto que não houve mudanças significativas. Destaca-se ainda, que se o portal continuar neste ritmo, provavelmente o valor de **SF** a longo prazo irá diminuir.

Ao utilizar do modelo e do SiAPDA percebe-se que além da avaliação é possível realizar o monitoramento dos portais. O sistema armazena todas as coletas com datas, proporcionando a criação de um histórico de coletas e avaliações, de modo a viabilizar ações para manter e acompanhar a evolução ou o retrocesso do portal.

É importante salientar que a atual versão do sistema não possui uma estrutura para a finalidade do monitoramento, mas disponibiliza os dados para que seja possível realizar esta análise.

#### 4.6 CONSIDERAÇÕES FINAIS

Neste capítulo foi apresentada a estrutura e os módulos do sistema desenvolvido a fim de auxiliar na aplicação do modelo. Vislumbrou-se também o *back-end* da aplicação que é composto pelos módulos e o *front-end* que proporciona toda a interação com os usuários, possibilitando iniciar ou solicitar as coletas, aplicar o modelo,

criar e aplicar os padrões, além da realização da avaliação automatizada com base nos padrões criados e a visualização interativa dos resultados obtidos.

Observa-se que a utilização do sistema possibilita avaliar um grande volume de conjuntos de dados e recursos, visto que em um processo de coleta manual seria trabalhoso e desgastante, além da possibilidade de ocasionar ruídos que poderiam trazer inconsistências aos resultados.

Ao apresentar os dados oriundos da aplicação do modelo com apoio do SiAPDA, fica perceptível a importância do instrumento desenvolvido, ressaltando a relevância do histórico de coletas e avaliações anteriores, possibilitando esclarecer qualquer dúvida sobre os resultados, além de evidenciar inconsistências sobre os portais de dados.

Observa-se também as limitações sobre os portais de dados acerca das recomendações, visto que cada portal se destaca em diferentes dimensão, contudo estes portais precisam promover ações de melhorias para que a médio e longo prazo possam estar alinhados com as recomendações e com isso evoluírem seus processos para a disponibilização dos dados no portal.

Destaca-se ainda que os resultados obtidos convergem em relação a problemática apontada pela literatura, onde os estudos de Lourenço (2015) e Machado *et al.* (2018) explicitam, que os portais publicam com a finalidade de cumprir uma legislação e que estes são apenas repositórios. Todavia, não pode ser generalizado esta constatação, onde percebe-se o esforço de algumas organizações em se alinhar com as recomendações.

Outro ponto de destaque, é a utilização da estrutura para o monitoramento da evolução ou regressão dos portais. Deste modo é possível comparar as diferentes coletas e com isso analisar em detalhes se as dimensões, indicadores e variáveis estão evoluindo com o passar do tempo. Neste sentido, ao fazer uso do score meta é possível ter-se pequenas metas a fim de proporcionar a evolução do portal como um todo.

Por sua vez, as repostas atribuídas por especialistas destacam a relevância e a importância desta pesquisa. Os especialistas responderam na maioria dos questionamento afirmando que tanto o modelo quando o sistema são importantes e relevantes.

## 5 CONSIDERAÇÕES FINAIS DA PESQUISA

Neste capítulo é apresentada a conclusão sobre o respectivo estudo, bem como o vislumbamento dos trabalhos futuros que foram percebidos ao decorrer da pesquisa.

### 5.1 CONCLUSÕES

Publicar dados abertos é extremamente importante tanto para a sociedade quanto para as organizações que publicam, sendo que estes dados devem estar acessíveis por todos.

Neste sentido, por conta das facilidades disponibilizadas pelos sistemas computacionais, os dados disponibilizados em sua maioria fazem uso de plataformas de catalogação de dados, denominados de portais de dados abertos. Estas plataformas proporcionam uma estrutura que auxilia na localização, acesso e exploração dos dados. No entanto, devido a existência de leis ou mesmo a não percepção das organizações sobre as vantagens de dados abertos alguns portais são considerados meros repositórios de dados. Deste modo, o uso de dados abertos ainda não atingiu todo o seu potencial e por consequência a sociedade não pode usufruir completamente de seus benefícios.

Ao utilizar das recomendações, observa-se uma série de benefícios para as organizaçõesadoras e também para a sociedade que fará uso destes dados. A partir desta premissa, objetivou-se o modelo de referência que sintetiza às recomendações e conceitos divulgados por diversos estudos oriundos da literatura e das boas práticas, além de proporcionar a utilização independentemente do nível de governo.

Para realizar a estruturação do modelo foi realizada uma revisão da literatura que através da identificação dos conceitos encontrados foi possível iniciar a composição do modelo e fez-se o uso de uma metodologia de apoio a construção de indicadores. O uso desta metodologia permitiu realizar a definição das dimensões, indicadores e suas variáveis, bem como a proposição do score final.

Para que o modelo pudesse ser utilizado, foi definido um processo de avaliação que tem por objetivo esclarecer cada etapa. Através do uso de um sistema computacional, estas etapas de avaliação serão realizadas pelo próprio sistema, contudo o seu entendimento é fundamental. Destaca-se ainda que é possível realizar a avaliação de maneira manual.

Os portais de dados possuem milhares de conjuntos de dados e recursos publicados, sendo que ao realizar a aplicação do modelo e a coleta dos dados manualmente seria uma tarefa desgastante e onerosa. Desta forma, percebeu-se a necessidade de fazer uso de um sistema computacional que auxiliasse na aplicação do modelo e na coleta dos dados, para isso foi desenvolvido o SiAPDA. Com o desenvolvimento deste sistema, foi possível realizar a coleta, a aplicação do modelo e a avaliação dos portais

de dados.

Desta forma, com o auxílio do sistema foram coletados portais de dados referentes aos três níveis de governo brasileiro (municipal, estadual e federal) com objetivo de apresentar que tanto o modelo quanto o sistema podem ser empregados nestes diferentes cenários. Neste sentido, foram coletados dos três portais cerca de 12.190 conjuntos de dados e cerca de 70.371 recursos, sendo aplicado o modelo e atribuído o padrão **DEFAULT**.

A definição dos padrões possibilitaram a criação de um conjunto de pesos a fim de ponderar a estrutura do portal avaliado, de modo, a priorizar os diferentes aspectos dos portais. Destaca-se que o padrão **DEFAULT** foi criado com o objetivo de distribuir igualmente todos os pesos entre as dimensões, indicadores e variáveis, com intuito de verificar o comportamento de toda a estrutura apresentada.

Como resultado, observou-se que todos os três portais não atendem a todas as recomendações apresentadas pela literatura. Sendo que cada portal se destaca em diferentes aspectos. Neste sentido, observou-se que o portal do governo federal por possuir um grande volume de conjuntos de dados se destacou no indicador proveniência em relação aos outros portais. Todavia, não foi o suficiente para se sobressair em relação ao portal do município de São Paulo. Por sua vez, o portal do estado do Rio Grande do Sul foi o que obteve o score final mais baixo dentre os três. Assim, observou-se que os três cenários analisados possuem espaço para melhorias. Deste modo, percebeu-se ainda que a quantidade de dados e o tempo de existência do portal não impactam na avaliação.

Observou-se também, que a literatura traz como problema a publicação de dados de forma obrigatória, refletindo em portais que servem apenas como meros repositórios de dados e isso acarreta uma série de problemas. Ao analisar os portais, percebeu-se que existem iniciativas das organizações em publicar dados e tentar alinhar-se com as recomendações. No entanto, pode ser percebido que se não houver uma sensibilização nas organizações em relação às publicações e não apenas iniciativas isoladas, estes problemas ainda podem persistir. Então, espera-se que este modelo possa contribuir indicando quais as limitações e em quais pontos os portais podem melhorar quando consideradas as recomendações.

Por sua vez, após a aplicação do modelo foram convidados especialistas que atuam na área de dados abertos, sendo estes pesquisadores, analistas de TI, administradores públicos e auditores, com objetivo de avaliar a composição do modelo de referência e o SiAPDA. As respostas obtidas foram favoráveis sobre a relevância e a importância deste trabalho.

Um fato interessante que ocorreu foi que ao aplicar o questionário, um dos respondentes informou que o sistema desenvolvido poderia estar com uma inconsistência. A inconsistência se dava pelo fato de que um quantitativo exposto pelo convidado em

relação ao portal do governo federal divergia do valor obtidos pelo SiAPDA . A partir disso, investigou-se minuciosamente o sistema e o modelo, sendo constatado através de coletas existentes desde 2019 que o portal de dados do governo federal está excluindo conjuntos de dados e que por sua vez tanto a aplicação quanto o modelo estavam corretos.

Foram efetuadas análises de outros portais de dados com o mesmo objetivo, identificar se o portal estava excluindo conjuntos de dados e concluiu-se que este fato era isolado e constatado apenas no porta do governo federal. Entende-se que por ser um portal de dados em nível federal e por existir vários envolvidos no processo de publicação, podem haver falhas em reflexo da incompreensão do tema ou mesmo a falta de alinhamento em relação as recomendações com os demais órgãos que publicam junto ao portal federal.

Por fim, com base nos resultados obtidos nos cenários e em decorrência as respostas do questionário aplicado junto aos especialistas concluiu-se que o modelo proposto tem a capacidade de avaliar a conformidade dos portais de dados em relação as recomendações elencadas da literatura. Deste modo, é possibilitado aos responsáveis pelos portais visualizarem os resultados que estão divergentes das recomendações apontadas pela literatura e assim possibilitar a melhoria dos portais de dados abertos. Portanto, o objetivo geral foi atingido e a pergunta de pesquisa respondida através da estrutura desenvolvida, apresentada e avaliada neste trabalho.

## 5.2 TRABALHOS FUTUROS

Na realização deste trabalho, diversos pontos de evolução e novas possibilidades de pesquisa e desenvolvimento foram percebidos. Dentre estes, vislumbra-se a realização de um estudo relacionando ao tamanho do portal (quantidade de conjuntos de dados e recursos) relacionando com a arrecadação de níveis de governo específicos, a fim de propor valores iniciais de referência para portais de dado abertos.

Outro ponto, a ser explorado é verificar se existe algum impacto em relação a existência de ecossistemas de dados abertos, visto que os ecossistemas podem contribuir significativamente com a disponibilização dos dados, sendo neste caso uma ação consciente e não uma obrigatoriedade da organização.

Outra possibilidade de pesquisa é a proposição de um modelo para avaliar o impacto dos dados abertos em determinada localidade, com objetivo de mensurar se a publicação de dados refletiu em novos negócios, se os cidadãos estão mais participativo se houve melhora nos serviços públicos, dentre outros.

Cabe sugerir ainda o emprego de técnicas de Inteligência Artificial (IA) com objetivo de aprimorar os algoritmos de coleta e avaliação. Outro ponto seria o uso de aprendizagem de máquina, com objetivo de facilitar aspectos de coletas relacionados a interface do portal, pois a variedade e a dinamicidade de interfaces existentes propor-

cionaria a evolução da coleta sem necessitar de atualização específicas e constantes no algoritmo.

Destaca-se também a possibilidade do desenvolvimento de uma ontologia quanto sua descrição formal em relação a termos linguísticos.

É possível também realizar a verificação de atualizações para compor o modelo de referência, considerando estudos posteriores à 2019. Deste modo, percebe-se que a seção 2.10.3 contribui com estes aspectos. Neste sentido, observa-se que incluir aspectos da Lei Geral de Proteção de Dados (LGPD) pode ampliar e agregar maior relevância ao modelo de referência.

Também é possível ampliar o escopo do modelo de referência adicionando o indicador preservação a fim de verificar se há exclusão dos conjuntos de dados e recursos. No entanto, este indicador está condicionado a ser utilizado após duas ou mais coletas do portal de dados. Isso ocorre, pois geralmente esses problemas relacionados a preservação são perceptíveis após a comparação entre coletas. Desta mesma forma, sugere-se como continuidade a implementação de um módulo de monitoramento. Assim, como é possível expandir o modelo ou propor um novo modelo para avaliar o conteúdo dos dados, a fim de verificar se estão de acordo com os padrões recomendados.

Utilizar do banco de dados construído à partir das coletas realizadas e aplicar técnicas de *bussines intelligence* ou *data analytics* com intuito de evidenciar outras variáveis ou temas de pesquisas, por exemplo.

Aprimorar a etapa de coleta e armazenamento dos dados de modo a proporcionar melhor desempenho, um exemplo, seria a utilização do *Hierarchical Data Format*. Outro ponto que pode ser explorado é o a verificação de segurança nos portais, além de verificar o emprego de *blockchain* neste contexto.

Utilizar de LOD para a exploração dos dados coletados a fim de identificar padrões ou mesmo inconsistências a nível de dados.

Realizar pesquisas que possibilite compor uma metodologia para propor padrões de avaliação de modo a possibilitar a avaliação conforme o perfil do *stakeholder*, considerando os contextos de cada portal de dados abertos.

Por fim, tais proposições têm como finalidade a expansão dos conhecimentos sobre o tema e o modelo de referência aqui exposto, sendo assim as oportunidades de possíveis pesquisas têm como intuito fornecer sugestões para o avanço do conhecimento e o aprimoramento dos portais de dados abertos.

## REFERÊNCIAS

- ABELLA, Alberto; ORTIZ-DE-URBINA-CRIADO, Marta; DE-PABLOS-HEREDERO, Carmen. Meloda 5: A metric to assess open data reusability. **El profesional de la información (EPI)**, v. 28, n. 6, 2019.
- AGHAEI, Sareh; NEMATBAKHS, Mohammad Ali; FARSANI, Hadi Khosravi. Evolution of the world wide web: From WEB 1.0 TO WEB 4.0. **International Journal of Web & Semantic Technology**, Academy & Industry Research Collaboration Center (AIRCC), v. 3, n. 1, p. 1, 2012.
- ALBINO, João Pedro. An approach to value creation in open data for small and medium-sized enterprises using “R” ecosystem. *In*: IEEE. **COMPUTING CONFERENCE, 2017**. [S.l.: s.n.], 2017. P. 346–351.
- ALMEIDA, Thábata Clezar de. **O Compartilhamento do Conhecimento em Portais de E-Participação Legislativa**. Jun. 2017. Diss. (Mestrado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2017/06/Thabata-Clezar-de-Almeida.pdf>.
- ALMEIDA NASCIMENTO, Heluiza Ormelez de. **Acessibilidade e Usabilidade em Plataformas MOOC: Indicadores de Boas Práticas**. Mai. 2019. Diss. (Mestrado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2019/05/Heluiza-Ormelez-de-Almeida-Nascimento.pdf>.
- ALSHEHHI, Maryam; LEIDA, Marcello; HIRSCH, Benjamin; YOO, Paul D; TAHA, Kamal. Visual analytics in the web of data. *In*: IEEE. **ELECTRONICS, CIRCUITS, AND SYSTEMS (ICECS), 2013 IEEE 20TH INTERNATIONAL CONFERENCE ON**. [S.l.: s.n.], 2013. P. 102–103.
- ANTONIOU, G.; HARMELEN F. van. **A Semantic Web Primer**. Cambridge: MIT Press, 2004. 238p.
- ARMAN, Arry Akhmad; RAMADHAN, Gilang; FAJRIN, Muhammad. Design of Data Management Guideline for Open Data Implementation: Case Study in Indonesia. *In*: **PROCEEDINGS OF THE 2015 2ND INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNANCE AND OPEN SOCIETY: CHALLENGES IN EURASIA**. St. Petersburg, Russian Federation: ACM, 2015. (EGOSE '15), p. 17–23. DOI:

10.1145/2846012.2846024. Disponível em:  
<http://doi.acm.org/10.1145/2846012.2846024>.

ARQUERO AVILES, Rosario; MARCO CUENCA, Gonzalo. European Union Open Data Portal: analysis and evaluation. **REVISTA GENERAL DE INFORMACION Y DOCUMENTACION**, UNIV COMPLUTENSE MADRID, SERVICIO PUBLICACIONES CIUDAD UNIV, OBISPO TREJO 3 ..., v. 24, n. 1, p. 99–118, 2014.

ASAT, Asrul Nizam; MAHAT, Amirul Faizal; HASSAN, Rosilah; AHMED, Amjed Sid. Development of dengue detection and prevention system (Deng-E) based upon open data in Malaysia. *In*: IEEE. **PROCEEDINGS OF THE 8TH INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF ELECTRONIC GOVERNANCE**. [S.l.: s.n.], 2017. P. 1–6.

ATTARD, Judie; ORLANDI, Fabrizio; SCERRI, Simon; AUER, Sören. A systematic review of open government data initiatives. **Government Information Quarterly**, Elsevier, v. 32, n. 4, p. 399–418, 2015.

AUER, Sören *et al.* Managing the life-cycle of linked data with the LOD2 stack. *In*: **THE SEMANTIC WEB–ISWC 2012**. [S.l.]: Springer, 2012. P. 1–16.

ÁVILA, Thiago. **O que faremos com os 40 trilhões de gigabytes de dados disponíveis em 2020?** [S.l.]: Open Knowledge Brasil, set. 2017. Disponível em:  
<https://www.ok.org.br/noticia/o-que-faremos-com-os-40-trilhoes-de-gigabytes-de-dados-disponiveis-em-2020/>.

BARRIO, Manuel GÉRTRUDIX; GARCÍA, Sergio ÁLVAREZ; DOCTOR, Visitante; FERNÁNDEZ, Mario RAJAS; DOCTOR, Contratado. Open Data in Mobile Applications, New Models for Service Information. **Fonseca: Journal of Communication**, Directory of Open Access Journals, v. 12, n. 12, p. 117–131, 2016.

BATES, Jo. The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. **Government Information Quarterly**, Elsevier, v. 31, n. 3, p. 388–395, 2014.

BAUER, Florian; KALTENBÖCK, Martin. Linked open data: The essentials. **Edition mono/monochrom, Vienna**, v. 710, 2011.

BELLO, Olayiwola; AKINWANDE, Victor; JOLAYEMI, Oluwatoyosi; IBRAHIM, Ahmed. Open data portals in africa: An analysis of open government data initiatives. **African Journal of Library, Archives & Information Science**, African Journal of Library, Archival & Information Science, v. 26, n. 2, p. 97, 2016.

BENITEZ-PAEZ, Fernando; DEGBELO, Auriol; TRILLES, Sergio; HUERTA, Joaquin. Roadblocks hindering the reuse of open geodata in Colombia and Spain: A data user's perspective. **ISPRS International Journal of Geo-Information**, Multidisciplinary Digital Publishing Institute, v. 7, n. 1, p. 6, 2018.

BERNASCONI, Anna; CERI, Stefano; CAMPI, Alessandro; MASSEROLI, Marco. Conceptual Modeling for Genomics: Building an Integrated Repository of Open Data. *In*: SPRINGER. **INTERNATIONAL CONFERENCE ON CONCEPTUAL MODELING**. [S.l.: s.n.], 2017. P. 325–339.

BERNERS-LEE, Tim. Linked Data Design Issues (July 2006). **URL= <http://www.w3.org/DesignIssues/LinkedData.html>**, última visita julho de 2018, 2010.

BERNERS-LEE, Tim. Linked Data Design Issues (July 2006). **URL= <http://www.w3.org/DesignIssues/LinkedData.html>**, última visita julho de 2018, 2011.

BERNERS-LEE, Tim; FIELDING, Roy; MASINTER, Larry. **Uniform resource identifiers (URI): generic syntax**. [S.l.]: RFC 2396, August, 1998. P. 4–8.

BIZ, Alexandre Augusto. **Avaliação dos Portais Turísticos Governamentais Quanto ao Suporte a Gestão do Conhecimento**. Abr. 2009. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2009/04/Alexandre-Augusto-Biz.pdf>.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked data: The story so far. *In*: **SEMANTIC SERVICES, INTEROPERABILITY AND WEB APPLICATIONS: EMERGING CONCEPTS**. [S.l.]: IGI Global, 2011. P. 205–227.

BOGDANOVIĆ-DINIĆ, Sanja; VELJKOVIĆ, Nataša; STOIMENOV, Leonid. How open are public government data? An assessment of seven open data portals. *In*: **MEASURING E-GOVERNMENT EFFICIENCY**. [S.l.]: Springer, 2014. P. 25–44.

BREMER, Carlos Frederico; LENZA, Rogério de Paula. Um modelo de referência para gestão da produção em sistemas de produção assembly to order–ATO e suas múltiplas aplicações. **Gestão & Produção**, v. 7, n. 3, p. 269–282, 2000.

CAI, Li; ZHU, Yangyong. The challenges of data quality and data quality assessment in the big data era. **Data science journal**, Ubiquity Press, v. 14, 2015.

CAMPOS, Paula Assumpção. **Dados Abertos Governamentais: Desafios na Publicação**. 2018. Diss. (Mestrado) – Universidade Federal de Santa Catarina.

CAO, Tuan-Dung; NGUYEN, Quang-Minh; NGUYEN, Anh-Duc; LE, Tan-Hung. Integrating open data and generating travel itinerary in semantic-aware tourist information system. *In*: ACM. **PROCEEDINGS OF THE 13TH INTERNATIONAL CONFERENCE ON INFORMATION INTEGRATION AND WEB-BASED APPLICATIONS AND SERVICES**. [S.l.: s.n.], 2011. P. 214–221.

CASAES, Júlio César Costa. **Governança de Dados Abertos Governamentais: Framework Conceitual para as Universidades Federais, Baseado em uma Visão Sistêmica**. Mai. 2019. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2019/05/J%C3%BAlio-C%C3%A9sar-Costa-Casaes-Tese.docx.pdf>.

CAVANILLAS, José Mariéa; CURRY, Edward; WAHLSTER, Wolfgang. The big data value opportunity. *In*: **NEW HORIZONS FOR A DATA-DRIVEN ECONOMY**. [S.l.]: Springer, Cham, 2016. P. 3–11.

CC, Creative Commons. **Frequently Asked Questions**. Abr. 2021. Disponível em: <https://creativecommons.org/faq/#is-creative-commons-against-copyright>.

CC, Creative Commons. **Licenças**. Fev. 2019. Disponível em: <https://br.creativecommons.org/licencas/>.

CEWEB. **Incentivo ao reuso de dados abertos**. *In*: [S.l.]: São Paulo, 2016. v. 1 6. Gestão de dados abertos governamentais. Acessado em: 30/05/2019. Disponível em: <https://ceweb.br/guias/incentivo-ao-reuso-de-dados-abertos/capitulo-6/>.

CHAN, Michael FS; CHUNG, Walter WC. A framework to develop an enterprise information portal for contract manufacturing. **International Journal of Production Economics**, Elsevier, v. 75, n. 1-2, p. 113–126, 2002.

CHARVAT, Karel; ESBRI PALOMARES, Miguel; MAYER, Walter; CHARVAT JR, Karel; CAMPOS, Antonio; PALMA, Raul; KRIVANEK, Zbynek. **FOODIE - Open data for agriculture**. In. DOI: 10.1109/ISTAFRICA.2014.6880647.

CHEN, Longbiao; ZHANG, Daqing; PAN, Gang; WANG, Leye; MA, Xiaojuan; CHEN, Chao; LI, Shijian. Container throughput estimation leveraging ship GPS traces and open data. In: ACM. **PROCEEDINGS OF THE 2014 ACM INTERNATIONAL JOINT CONFERENCE ON PERVASIVE AND UBIQUITOUS COMPUTING**. [S.l.: s.n.], 2014. P. 847–851.

CKAN, Comprehensive Knowledge Archive Network. **About Comprehensive Knowledge Archive Network**. Fev. 2019. Disponível em: <http://ckan.org>.

COELHO, Claudio. Compliance na Administração Pública. **Revista de Direito da Faculdade Guanambi**, v. 3, n. 01, p. 75–95, ago. 2017. DOI: 10.29293/rdfg.v3i01.103. Disponível em: <http://177.38.182.246/revistas/index.php/Revistadedireito/article/view/103>.

COLPAERT, Pieter; JOYE, Sarah; MECHANT, Peter; MANNENS, Erik; VAN DE WALLE, Rik. The 5 stars of open data portals. In: **PROCEEDINGS OF THE 7TH INTERNATIONAL CONFERENCE ON METHODOLOGIES, TECHNOLOGIES AND TOOLS ENABLING E-GOVERNMENT (METTEG13), UNIVERSITY OF VIGO, SPAIN**. [S.l.: s.n.], 2013. P. 61–67.

CORBETTA, Piergiorgio. **Metodología y técnicas de investigación social**. [S.l.]: McGraw-Hill, 2003.

CORREA, Andreiuid Sheffer; SOUZA, Raul Mendes de; SILVA, Flavio Soares Correa da. Towards an automated method to assess data portals in the deep web. **Government Information Quarterly**, Elsevier, v. 36, n. 3, p. 412–426, 2019.

CORRÊA, Andreiuid Sh; PAULA, Evandro Couto de; CORREA, Pedro Luiz Pizzigatti; SILVA, Flávio Soares Corrêa da. Transparency and open government data: a wide national assessment of data openness in Brazilian local governments. **Transforming Government: People, Process and Policy**, Emerald Publishing Limited, v. 11, n. 1, p. 58–78, 2017.

CORRÊA, Andreiuid Sheffer; CORRÊA, Pedro Luiz Pizzigatti; SILVA, Flávio Soares Corrêa da. Transparency portals versus open government data: an assessment of openness in Brazilian municipalities. *In*: ACM. **PROCEEDINGS OF THE 15TH ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH**. [S.l.: s.n.], 2014. P. 178–185.

CUCA, Branka. Geospatial future is open: lessons learnt from applications based on open data. *In*: SPRINGER. **INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ITS APPLICATIONS**. [S.l.: s.n.], 2016. P. 491–502.

CUNHA, Rodrigo Rafael. **Rankings e Indicadores para Smart Cities: uma proposta de cidades inteligentes autopoieticas**. Mai. 2019. Diss. (Mestrado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2019/05/Rodrigo-Rafael-Cunha-disserta%C3%A7%C3%A3o.pdf>.

DADOS ABERTOS, Portal Brasileiro de. **O que são dados abertos**. [S.l.: s.n.], 2018. <http://dados.gov.br/dados-abertos/>. 25/07/2018.

DE HAES, Steven; VAN GREMBERGEN, Wim; DEBRECENY, Roger S. COBIT 5 and enterprise governance of information technology: Building blocks and research opportunities. **Journal of Information Systems**, American Accounting Association, v. 27, n. 1, p. 307–324, 2013.

DE VOCHT, Laurens; VERBORGH, Ruben; MANNENS, Erik; VAN DE WALLE, Rik; VAN DEN BOSCH, Wouter; BUYLE, Raf; KONINCKX, Bruno. Providing interchangeable open data to accelerate development of sustainable regional mobile tourist guides. *In*: ACM. **PROCEEDINGS OF THE 9TH INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF ELECTRONIC GOVERNANCE**. [S.l.: s.n.], 2016. P. 195–198.

DEGBELO, Auriol; TEKA, Brhane Bahrishum. Spatial search strategies for open government data: A systematic comparison. *In*: **PROCEEDINGS OF THE 13TH WORKSHOP ON GEOGRAPHIC INFORMATION RETRIEVAL**. [S.l.: s.n.], 2019. P. 1–10.

DEIRDRE LEE, Derilinx; LÓSCIO, Bernadette Farias; ARCHER, Phil. **Data on the Web Best Practices Use Cases e Requirements**. Fev. 2019. Disponível em: <https://www.w3.org/TR/dwbp-ucr/>.

DEMSKI, Hans; GARDE, Sebastian; HILDEBRAND, Claudia. Open data models for smart health interconnected applications: the example of openEHR. **BMC medical informatics and decision making**, BioMed Central, v. 16, n. 1, p. 137, 2016.

DERGUECH, Wassim; BRUKE, Eanna; CURRY, Edward. **An autonomic approach to real-time predictive analytics using open data and internet of things**. *In*: IEEE. 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence & Computing and 2014 IEEE 11th Intl Conf on Autonomic & Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom). [S.l.: s.n.], 2014. P. 204–211.

DIN, Siti Haslinda Mohd; MANSOR, Nur Aziha; ARUS, Faiza Rusrianti Tajul. OPEN DATA AND CHALLENGES FACED NY NATIONAL STATISTICS OFFICES. *In*: **MALAYSIA STATISTICS CONFERENCE (MYSTATS 2015)**. [S.l.: s.n.], nov. 2015. P. 77.

DOMINGUE, John; FENSEL, Dieter; HENDLER, James A. **Handbook of semantic web technologies**. [S.l.]: Springer Science & Business Media, 2011.

DRESCH, Aline; LACERDA, Daniel Pacheco; JÚNIOR, José Antonio Valle Antunes. **Design science research: método de pesquisa para avanço da ciência e tecnologia**. [S.l.]: Bookman Editora, 2015.

EGC. **EGC - Linhas de Pesquisa**. [S.l.: s.n.], 2021. Disponível em: <https://ppgegc.paginas.ufsc.br/linhas-de-pesquisa/>. Acesso em: 22 mai. 2021.

ENRIÉQUEZ-REYES, Robert; CADENA-VELA, Susana; FUSTER-GUILLÓ, Andrés; MAZÓN, Jose-Norberto; IBÁÑEZ, Luis Daniel; SIMPERL, Elena. Systematic Mapping of Open Data Studies: Classification and Trends From a Technological Perspective. **IEEE Access**, IEEE, v. 9, p. 12968–12988, 2021.

EUROPEAN DATA PORTAL. **Benefits of Open Data**. [S.l.: s.n.], nov. 2018. <https://www.europeandataportal.eu/en/using-data/benefits-of-open-data>.

EUROPEAN DP, European Data Portal. **Vantagens da utilização dos dados**. Fev. 2019. Disponível em: <https://www.europeandataportal.eu/pt/using-data/benefits-of-open-data>.

FIEP, SESI-PR, Serviço Social da Indústria. Departamento Regional do Estado do Paraná. Observatório Regional Base de Indicadores de Sustentabilidade.

### **CONSTRUÇÃO E ANÁLISE DE INDICADORES.**

<http://www.portalodm.com.br/dnfile/epusInpigu4wbwym5cev/pdf/publicacoes/1/construcao-e-analise-de-indicadores.pdf>: Serviço Social da Indústria. Departamento Regional do Estado do Paraná., 2010. v. 1.

FITRIANI, Widia Resti; HIDAYANTO, Achmad Nizar; PURWANDARI, Betty; NAZIEF, Bobby A. A.; HARDIAN, Bob. **Evaluating the implementation of open data principles on government websites in Indonesia.** *In*: 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS). [S.l.]: IEEE, out. 2017. DOI: 10.1109/icacsis.2017.8355032. Disponível em: <https://doi.org/10.1109/icacsis.2017.8355032>.

GILL, Mark; CORBETT, Jon. Downscaling: Understanding the influence of open data initiatives in smaller and mid-sized cities in British Columbia, Canada. **The Canadian Geographer/Le Géographe canadien**, Wiley Online Library, v. 61, n. 3, p. 346–359, 2017.

GLOBAL PARTNERSHIP FOR SUSTAINABLE DEVELOPMENT DATA. **SDG Data Revolution Roadmaps Toolbox: Open Data Module.** [S.l.: s.n.], jul. 2016. Disponível em: <http://www.data4sdgs.org/sites/default/files/2017-09/0pen%20Data%20for%20Sustainable%20Development%20-%20Data4SDGs%20Toolbox%20Module.pdf>.

GÓIS, Veruska Sayonara de. A Lei de Compliance e sua Configuração Enquanto Política Pública Regulatória para o Setor Privado Brasileiro. **Revista Controle - Doutrina e Artigos**, v. 12, n. 2, p. 98–117, nov. 2016. DOI: 10.32586/rcda.v12i2.52. Disponível em: <https://revistacontrole.tce.ce.gov.br/index.php/RCDA/article/view/52>.

GOMES, Francisco Carlos da Rocha. Arquitetura de repositório semântico de dados para organização de pesquisa agropecuária. **Universidade Federal do Rio de Janeiro/Instituto Brasileiro de Informação em Ciência e Tecnologia**, 2013.

GOMES, Murilo Silveira. **Proposta de Arquitetura para Ecossistema de Inovação em Dados Abertos.** 2017. Diss. (Mestrado) – Universidade Federal de Santa Catarina.

GOÑI, Jorge Luis; FERNANDES, Maria Cristina Pfeiffer; LUCENA, Carlos José Pereira de. **E-learning e a Web Semântica**. [S.l.]: PUC, 2002. P. 4, 5.

HAO, B; SUN, W; YU, Y; XIE, G. Developing Healthcare Data Analytics APPs with Open Data Science Tools. **Studies in health technology and informatics**, v. 235, p. 176–180, 2017.

HARRIS, Ray; BAUMANN, Ingo. Open data policies and satellite Earth observation. **Space Policy**, Elsevier, v. 32, p. 44–53, 2015.

HARTH, Andreas; JANIK, Maciej; STAAB, Steffen. Semantic Web architecture. *In*: **HANDBOOK OF SEMANTIC WEB TECHNOLOGIES**. [S.l.]: Springer, 2011. P. 43–75.

HOUNSELL, Nick; SHRESTHA, Birendra; MCDONALD, Mike; WONG, Alan. Open data and the needs of older people for public transport information. **Transportation Research Procedia**, 2016.

HUSEIN, Inne Gartina; SUNINDYO, Wikan Danar; BAHAWARES, Rizal; NAINGGOLAN, Yulius; AKBAR, Saiful. Open data strategy for enhancing the productivity and competitiveness of fishery SMEs in Indonesia. *In*: IEEE. **ELECTRICAL ENGINEERING AND INFORMATICS (ICEEI), 2015 INTERNATIONAL CONFERENCE ON**. [S.l.: s.n.], 2015. P. 490–495.

IBÁÑEZ, Luis-Daniel; MILLARD, Ian; GLASER, Hugh; SIMPERL, Elena. An assessment of adoption and quality of linked data in European open government data. *In*: SPRINGER. **INTERNATIONAL SEMANTIC WEB CONFERENCE**. [S.l.: s.n.], 2019. P. 436–453.

INFORMATION SOCIETY, European Commission. Directorate General for the *et al*. **Creating value through open data: study on the impact of re use of public data resources**. [S.l.]: Publications Office, 2015. DOI: 10.2759/328101. Disponível em: <https://data.europa.eu/doi/10.2759/328101>.

JAPPUR, Rafael Feyh. **Modelo conceitual para criação, aplicação e avaliação de jogos educativos digitais**. Mai. 2014. Tese – •. Acessado em: 06/06/2019. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/129458/328363.pdf?sequence=1&isAllowed=y>.

JAROLIÉMEK, J; MARTINEC, R. Analysis of Open Data Availability in Czech Republic Agrarian Sector. **Agris On-Line Papers in Economics & Informatics**, v. 8, n. 3, 2016.

JUNIOR, Vanderlei Freitas. **Ontologia para Representação Semântica de Indicadores de Desempenho Considerando Aspectos de Vaguidade, Temporalidade e Relacionamento entre Indicadores**. Fev. 2016. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2016/03/Vanderlei-Freitas-Junior.pdf>.

JUNIOR, Vanderlei FREITAS; FLAVIO, CECI; WOSZEZENKI, Cristiane Raquel; LEOPOLDO, Alexandre. Design Science Research Methodology Enquanto Estratégia Metodológica para a Pesquisa Tecnológica. **Revistas Espacios 38 (6)**, p. 25, 2017.

KAPOOR, Kawaljeet; WEERAKKODY, Vishanth; SIVARAJAH, Uthayasankar. Open data platforms and their usability: Proposing a framework for evaluating citizen intentions. *In*: SPRINGER. **CONFERENCE ON E-BUSINESS, E-SERVICES AND E-SOCIETY**. [S.l.: s.n.], 2015. P. 261–271.

KAZEMI, Samira; ABGHARI, Shahrooz; LAVESSON, Niklas; JOHNSON, Henric; RYMAN, Peter. Open data for anomaly detection in maritime surveillance. **Expert Systems with Applications**, Elsevier, v. 40, n. 14, p. 5719–5729, 2013.

KEGELMAN, John C; HARBOTT, Lene K; GERDES, J Christian. Insights into vehicle trajectories at the handling limits: analysing open data from race car drivers. **Vehicle system dynamics**, Taylor & Francis, v. 55, n. 2, p. 191–207, 2017.

KIRSTEIN, Fabian; DITTWALD, Benjamin; DUTKOWSKI, Simon; GLIKMAN, Yury; SCHIMMLER, Sonja; HAUSWIRTH, Manfred. Linked data in the European data portal: A comprehensive platform for applying dcat-ap. *In*: SPRINGER. **INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT**. [S.l.: s.n.], 2019. P. 192–204.

KITSIOS, Fotis; KAMARIOTOU, Maria. Open data hackathons: an innovative strategy to enhance entrepreneurial intention. **International Journal of Innovation Science**, Emerald Publishing Limited, 2018.

KOIVUNEN, Marja-Riitta; MILLER, Eric. W3c semantic web activity. **Semantic Web Kick-Off in Finland**, p. 27–44, 2001.

KORFIATI, Athina; GKONOS, Charalampos; VERONESI, Fabio; GAKI, Ariadni; GRASSI, Stefano; SCHENKEL, Roland; VOLKWEIN, Stephan; RAUBAL, Martin; HURNI, Lorenz. Estimation of the Global Solar Energy Potential and Photovoltaic Cost with the use of Open Data. **International Journal of Sustainable Energy Planning and Management**, v. 9, p. 17–30, 2016.

KRASIKOV, Pavel; OBRECHT, Timo; LEGNER, Christine; EURICH, Markus. **Is Open Data Ready for Use by Enterprises? Learnings from Corporate Registers.**, p. 109–120, 2020.

KUBLER, Sylvain; ROBERT, Jérémy; LE TRAON, Yves; UMBRICH, Jürgen; NEUMAIER, Sebastian. Open data portal quality comparison using AHP. *In: ACM. PROCEEDINGS OF THE 17TH INTERNATIONAL DIGITAL GOVERNMENT RESEARCH CONFERENCE ON DIGITAL GOVERNMENT RESEARCH.* [S.l.: s.n.], 2016. P. 397–407.

KUBLER, Sylvain; ROBERT, Jérémy; NEUMAIER, Sebastian; UMBRICH, Jürgen; LE TRAON, Yves. Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. **Government Information Quarterly**, Elsevier, v. 35, n. 1, p. 13–29, 2018.

KUCERA, Jan; CHLAPEK, Dusan; KLIÉMEK, Jakub; NECASKÝ, Martin. Methodologies and Best Practices for Open Data Publication. *In: DATESO.* [S.l.: s.n.], 2015. P. 52–64.

LAUFER, Carlos. **Guia de web semântica.** 2015. Disponível em: [https://nic.br/media/docs/publicacoes/13/Guia\\_Web\\_Semantica.pdf](https://nic.br/media/docs/publicacoes/13/Guia_Web_Semantica.pdf).

LE MOIGNE, Jean-Louis. Les épistémologies constructivistes. **Presses universitaires de France Paris**, 1995.

LIMA, Júnio César de; CARVALHO, Cedric L de. **Resource description framework (rdf).** [S.l.], 2005.

LNENICKA, Martin. An in-depth analysis of open data portals as an emerging public e-service. **International Journal of Social, Education, Economics and Management Engineering**, v. 9, n. 2, p. 589–599, 2015.

LÓSCIO, Bernadette; BURLE, Caroline; CALEGARI, Newton. Boas práticas para dados na WEB: desafios e benefícios. **Revista Principia**, Divulgação Científica e Tecnológica do IFPB, n. 32, p. 9–18, 2016.

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton. **Data on the Web Best Practices**. [S.l.: s.n.], 2017. <https://www.w3.org/TR/dwbp/>.

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton. Data on the web best practices: Challenges and benefits. *In*: **OPEN DATA RESERACH SYMPOSIUM (ODRS 2016)**. [S.l.: s.n.], 2016.

LÓSCIO, Bernadette Farias; OLIVEIRA, Marcelo lury S.; BITTENCOURT, Ig lbert. Publicação e Consumo de Dados na Web: Conceitos e Desafios. *In*: COMPUTAÇÃO, Sociedade Brasileira de (Ed.). **Simpósio Brasileiro de Banco de Dados**. <<http://dexl.Incc.br/sbbd2015/anais/ShortCourses.pdf>>: [s.n.], fev. 2015. (1), p. 39–68. Acessado em: 08/02/2019.

LOURENÇO, Rui Pedro. An analysis of open government portals: A perspective of transparency for accountability. **Government Information Quarterly**, Elsevier, v. 32, n. 3, p. 323–332, 2015.

LOURENÇO, Rui Pedro. Open government portals assessment: a transparency for accountability perspective. *In*: SPRINGER. **INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT**. [S.l.: s.n.], 2013. P. 62–74.

MACEDO, Dirceu; LEMOS, Daniela. Dados abertos governamentais: iniciativas e desafios na abertura de dados no Brasil e outras esferas internacionais. **AtoZ: novas práticas em informação e conhecimento**, v. 10, n. 2, p. 14–26, 2021. ISSN 2237-826X. DOI: 10.5380/atoz.v10i2.77737. Disponível em: <https://revistas.ufpr.br/atoz/article/view/77737>.

MACHADO, Luís Miguel Oliveira; SOUZA, Renato Rocha; GRAÇA SIMÕES, Maria da. Semantic web or web of data? a diachronic study (1999 to 2017) of the publications of tim berners-lee and the world wide web consortium. **Journal of the Association for Information Science and Technology**, Wiley Online Library, v. 70, n. 7, p. 701–714, 2019.

MACHADO, Viviann; MANTINI, Gabriel; VITERBO, José; BERNARDINI, Flavia; BARCELLOS, Raissa. An instrument for evaluating open data portals: a case study in

Brazilian cities. *In*: ACM. **PROCEEDINGS OF THE 19TH ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH: GOVERNANCE IN THE DATA AGE**. [S.l.: s.n.], 2018. P. 19.

MÁCHOVÁ, Renata; LNÉNICKA, Martin. Evaluating the quality of open data portals on the national level. **Journal of theoretical and applied electronic commerce research**, Universidad de Talca, v. 12, n. 1, p. 21–41, 2017.

MÁCHOVÁ, Renáta; HUB, Miloslav; LNENICKA, Martin. Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective. **Aslib Journal of Information Management**, Emerald Publishing Limited, v. 70, n. 3, p. 252–268, 2018.

MACK, Robert; RAVIN, Yael; BYRD, Roy J. Knowledge portals and the emerging digital knowledge workplace. **IBM systems journal**, IBM, v. 40, n. 4, p. 925–955, 2001.

MANYIKA, James; CHUI, Michael; FARRELL, Diana; KUIKEN, Steve Van; GROVES, Peter; DOSHI, Elizabeth Almasi. **Open data: Unlocking innovation and performance with liquid information**. [S.l.: s.n.], out. 2013.  
<https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>.

MARKOVIĆ, Marko; GOSTOJIĆ, Stevan. Open judicial data: a comparative analysis. **Social Science Computer Review**, SAGE Publications Sage CA: Los Angeles, CA, v. 38, n. 3, p. 295–314, 2020.

MARMIER, Auriane; METTLER, Tobias. Developing an index for measuring OGD publisher compliance to good practice standards: Insights from opendata. swiss. **Information Polity**, IOS Press, v. 25, n. 1, p. 91–110, 2020.

MASUZZO, Paola; MARTENS, Lennart *et al.* An open data ecosystem for cell migration research. **Trends in cell biology**, Elsevier, v. 25, n. 2, p. 55–58, 2015.

MCCRAE, John P.; ABELE, Andrejs. **The Linked Open Data Cloud**. [S.l.: s.n.], 2021.  
<https://lod-cloud.net/>. Acessado em: 03/06/2021.

MCNAUGHTON, Maurice; MCLEOD, Michelle T; BOXILL, Ian. An Actor Network Perspective of Tourism Open Data. *In*: **TOURISM AND HOSPITALITY MANAGEMENT**. [S.l.]: Emerald Group Publishing Limited, 2016. P. 47–60.

MEDJROUBI, Wided; MÜLLER, Ulf Philipp; SCHARF, Malte; MATKE, Carsten; KLEINHANS, David. Open data in power grid modelling: new approaches towards transparent grid models. **Energy Reports**, Elsevier, v. 3, p. 14–21, 2017.

MEZZAROBA, Mariana Pessini. **Framework para Avaliação de Portais do Poder Judiciário Brasileiro a Partir de Mecanismos de Gestão do Conhecimento**. Mai. 2019. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2019/05/Mariana-Pessini-Mezzaroba-Tese.pdf>.

MEZZAROBA, Mariana Pessini. **Requisitos para a Avaliação de Portais de Governo Eletrônico do Poder Judiciário a partir das Resoluções de Metas do CNJ**. Mai. 2015. Diss. (Mestrado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2015/05/Mariana-Pessini-Mezzaroba.pdf>.

MOURA SPERONI, Rafael de. **Modelo de Referência Para Indicadores de Inovação Regional Suportado por Dados Ligados**. 2016. Tese (Doutorado) – Universidade Federal de Santa Catarina.

NEUMAIER, Sebastian. **Open Data Quality Assessment and Evolution of (Meta-)Data Quality in the Open Data Landscape**. 2015. Tese (Doutorado) – Technische Universität Wien.

NEUMAIER, Sebastian; UMBRICH, Jürgen; POLLERES, Axel. Automated quality assessment of metadata across open data portals. **Journal of Data and Information Quality (JDIQ)**, ACM, v. 8, n. 1, p. 2, 2016.

NOVA URIARTE, Flavia Maia da. **PORTAL CORPORATIVO COMO CANAL PARA GESTÃO DO CONHECIMENTO**. Fev. 2006. Tese Doutorado – Universidade Federal de Santa Catarina, <http://btd.egc.ufsc.br/wp-content/uploads/2010/06/Flavia-Maia-da-Nova-Uriarte.pdf>.

ODD, Open Data Day. **Open Data Day**. Fev. 2019. Disponível em: [http://opendataday.org/pt\\_br/#resources](http://opendataday.org/pt_br/#resources).

ODM, Open Data Monitor. **Methodology of OpenDataMonitor**. Mar. 2019. Disponível em: <https://opendatamonitor.eu/frontend/web/index.php?r=site%2Fmethodology>.

OKF, Open Knowledge Foundation. **Formatos de Arquivo**. Abr. 2020. Disponível em: [http://opendatahandbook.org/guide/pt\\_BR/appendices/file-formats/](http://opendatahandbook.org/guide/pt_BR/appendices/file-formats/).

OKF, Open Knowledge Foundation. **Network**. Abr. 2021. Disponível em: <https://okfn.org/network/>.

OPC, Open Data Commons. **Licenças**. Fev. 2019. Disponível em: <https://opendatacommons.org/licenses/>.

OPEN DEFINITION. **Open Definition**. [S.l.: s.n.], 2021. <http://opendefinition.org/>. 04/06/2021.

OPEN DEFINITION. **Open Definition 2.1**. [S.l.: s.n.], 2018. 10/11/2018. Disponível em: <http://opendefinition.org/od/2.1/en/#14-open-format>.

OPENDATA, Open Data Hand Book. **Machine readable**. Fev. 2019. Disponível em: <http://opendatahandbook.org/glossary/en/terms/machine-readable/>.

OPENDATASOFT. **About us**. Fev. 2019a. Disponível em: <https://www.opendatasoft.com/about-us/>.

OPENDATASOFT. **Stop observing data! Turn them into revenues with OpenDataSoft**. Fev. 2019b. Disponível em: <https://www.eurocloud.fr/doc/opendatasoft.pdf>.

OPENDATASOFT. **User manual of the OpenDataSoft platform**. Fev. 2019c. Disponível em: <https://help.opendatasoft.com/platform/en/>.

OPENDATASOFT, OpenDataSoft: Data Network. **Data**. Fev. 2019. Disponível em: <https://data.opendatasoft.com/pages/home/>.

OSAGIE, Edobor; WAQAR, Mohammad; ADEBAYO, Samuel; STASIEWICZ, Arkadiusz; PORWOL, Lukasz; OJO, Adegboyega. Usability evaluation of an open data platform. *In*: ACM. **PROCEEDINGS OF THE 18TH ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH**. [S.l.: s.n.], 2017. P. 495–504.

OWOLABI, K; SONOIKI, D; SALET, F; GIGNAC, N. Application of FOSS4G and Open Data to Support Polio Eradication, Vaccine Delivery and Ebola Emergency Response in West Africa. **International Journal of Geoinformatics**, v. 11, n. 3, 2015.

PACHECO, R. C. S. **Em Interdisciplinaridade: Universidade e Inovação Social e Tecnológica**. In: [S.l.]: Organizado por Joana Maria Pedro e Patrícia de Sá Freire, Curitiba, CRV, 2016. Coprodução em Ciência, Tecnologia e Inovação: Fundamentos e Visões, p. 21–62.

PARYCEK, Peter; HOCHTL, Johann; GINNER, Michael. Open government data implementation evaluation. **Journal of theoretical and applied electronic commerce research**, Universidad de Talca, v. 9, n. 2, p. 80–99, 2014.

PEDRINACI, Carlos; DOMINGUE, John *et al.* Toward the next wave of services: Linked services for the web of data. **J. ucs**, v. 16, n. 13, p. 1694–1719, 2010.

PEFFERS, Ken; TUUNANEN, Tuure; ROTHENBERGER, Marcus A; CHATTERJEE, Samir. A design science research methodology for information systems research. **Journal of management information systems**, Taylor & Francis, v. 24, n. 3, p. 45–77, 2007.

PEREIRA, Larissa Mariany Freiberger. **OGDPub: Uma ontologia para Publicação de Dados Abertos Governamentais**. 2017. Diss. (Mestrado) – Universidade Federal de Santa Catarina.

PETYCHAKIS, Michael; VASILEIOU, Olga; GEORGIS, Charilaos; MOUZAKITIS, Spiros; PSARRAS, John. A state-of-the-art analysis of the current public data landscape from a functional, semantic and technical perspective. **Journal of theoretical and applied electronic commerce research**, Universidad de Talca, v. 9, n. 2, p. 34–47, 2014.

PIDD, Michael. Just modeling through: A rough guide to modeling. **Interfaces, INFORMS**, v. 29, n. 2, p. 118–132, 1999.

PINHEIRO, Emerson B; COUTINHO, Emanuel F; MOREIRA, Leonardo O; BEZERRA, Carla IM; PAILLARD, Gabriel AL. cARFleet: A mobile application for vehicle fleet visualization based on augmented reality and open data. In: IEEE. **TELEMATICS AND INFORMATION SYSTEMS (EATIS), 2016 8TH EURO AMERICAN CONFERENCE ON**. [S.l.: s.n.], 2016. P. 1–5.

PIRES, Marco Túlio. **Melhoria do ambiente de negócios por meio da transparência no Estado de São Paulo**. Fev. 2019. Disponível em: [https://nic.br/media/docs/publicacoes/13/Guia\\_Dados\\_Abertos.pdf](https://nic.br/media/docs/publicacoes/13/Guia_Dados_Abertos.pdf).

PORTAL DE DADOS GOV, Governo Federal. **Portal de Dados Abertos do Governo Federal**. Fev. 2019. Disponível em: <http://dados.gov.br/>.

RAMALHO, Rogério Aparecido Sá; OUCHI, Marcos Teruo. TECNOLOGIAS SEMÂNTICAS: novas perspectivas para a representação de recursos informacionais. **Informação & Informação**, v. 16, n. 3, p. 60–75, 2011a.

RAMALHO, Rogério Aparecido Sá; OUCHI, Marcos Teruo. TECNOLOGIAS SEMÂNTICAS: novas perspectivas para a representação de recursos informacionais. **Informação & Informação**, v. 16, n. 3, p. 60–75, 2011b.

ROCCA, Gusseppe Bravo; CASTILLO-CARA, Manuel; LEVANO, Raul Arias; HERRERA, Javier Villegas; OROZCO-BARBOSA, Luis. Citizen security using machine learning algorithms through open data. *In*: IEEE. **COMMUNICATIONS (LATINCOM), 2016 8TH IEEE LATIN-AMERICAN CONFERENCE ON**. [S.l.: s.n.], 2016. P. 1–6.

ROSSI, Bruno; RUSSO, Barbara; SUCCI, Giancarlo. Open source software and open data standards as a form of technology adoption: a case study. *In*: SPRINGER. **IFIP INTERNATIONAL CONFERENCE ON OPEN SOURCE SYSTEMS**. [S.l.: s.n.], 2007. P. 325–330.

ROTTA, Maurício José Ribeiro. **As Plataformas de Governo Eletrônico e seu Potencial para a Promoção dos Princípios dos Commons: O Caso dos Municípios Brasileiros**. Dez. 2018. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2018/12/Maur%C3%ADcio-Rotta.pdf>.

ROYO-MONTANES, Sonia; BENITEZ-GOMEZ, Alberto. **Open data portals. Methodology of analysis and application to spanish municipalities [Portales de datos abiertos. Metodología de análisis y aplicación a municipios españoles]**. [S.l.], 2019.

SALES DOS SANTOS, Jose Carlos; GONCALVES DA SILVA, Rubens Ribeiro. ELECTRONIC GOVERNMENT AND POLITICAL PARTICIPATION IN THE MEMBERS' WEBSITES OF THE LEGISLATIVE ASSEMBLY OF BAHIA. **Informacao &**

**Sociedade-Estudos**, UNIV FED PARAIBA CCSA CIDADE UNIV, PROG POS-GRAD CIENC INFORM, JOAO PESSOA . . . , v. 21, n. 1, p. 215–224, 2011.

SALM JÚNIOR, José Francisco. **Padrão de projeto de ontologias para inclusão de referências do novo serviço público em plataformas de governo aberto**. Mai. 2012. Tese – Universidade Federal de Santa Catarina. Acessado em: 06/06/2019. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2012/08/JOSEFRANCISCOTESEPDF.pdf>.

SANTOS, Emanuella; NICOLAU, Marcos. Web do Futuro: a cibercultura e os caminhos trilhados rumo a uma web semântica ou web 3.0. *In*: **XXXV CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO**. [S.l.: s.n.], 2012.

SANTOS, Renato Almeida dos; HOYOS GUEVARA, Arnaldo José de; AMORIM, Maria Cristina Sanches; FERRAZ-NETO, Ben-Hur. Compliance e liderança: a suscetibilidade dos líderes ao risco de corrupção nas organizações. **Einstein (16794508)**, v. 10, n. 1, 2012.

SANTOS ALBUQUERQUE, Almir dos. **Modelo de Apoio a Gestão da Qualidade da Informação em Portais do Poder Judiciário Utilizando Fuzzy Quality Function Deployment**. Jan. 2020. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2019/05/J%C3%BAlio-C%C3%A9sar-Costa-Casaes-Tese.docx.pdf>.

SAXENA, Stuti. Open government data (OGD) in six Middle East countries: an evaluation of the national open data portals. **Digital Policy, Regulation and Governance**, Emerald Publishing Limited, v. 20, n. 4, p. 310–322, 2018.

SAYOGO, Djoko Sigit; PARDO, Theresa A; COOK, Meghan. A framework for benchmarking open government data efforts. *In*: IEEE. **2014 47TH HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES**. [S.l.: s.n.], 2014. P. 1896–1905.

SCHLAGWEIN, Daniel; CONBOY, Kieran; FELLER, Joseph; LEIMEISTER, Jan Marco; MORGAN, Lorraine. “Openness” with and without Information Technology: A Framework and a Brief History. **Journal of Information Technology**, SAGE Publications, v. 32, n. 4, p. 297–305, dez. 2017. DOI: 10.1057/s41265-017-0049-3. Disponível em: <https://doi.org/10.1057/s41265-017-0049-3>.

SHADBOLT, Nigel; BERNERS-LEE, Tim; HALL, Wendy. The semantic web revisited. **IEEE intelligent systems**, v. 21, n. 3, p. 96–101, 2006.

SILVA, Antonio Waldimir Leopoldino. **Governança de Sistemas de Indicadores de Sustentabilidade em Processos de Avaliação Ambiental Estratégica sob Mediação da Gestão do Conhecimento**. Jun. 2014. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2014/09/Antonio-Waldimir-L-da-Silva.pdf>.

SILVA, Deborah Bernett Leal da. **Indicadores para Avaliação da Influência dos Ambientes de Empreendedorismo Inovador na Geração de Capital Social**. Fev. 2016. Tese (Doutorado) – Universidade Federal de Santa Catarina. Disponível em: <http://btd.egc.ufsc.br/wp-content/uploads/2015/04/Deborah-Bernett.pdf>.

SILVA, Marco Antônio da. **Prova de conceito (PoC) em projetos**. [S.l.: s.n.], jun. 2014. Disponível em: <https://pmkb.com.br/artigos/prova-de-conceito-poc-em-projetos/>.

SILVA, Sivaldo Pereira da; SOARES, Ana Thereza Nogueira; CESAR, Daniel Jorge Teixeira; RABELO, Leon Eugênio Monteiro. Indicators for Open Data qualitative assessment: intelligibility, operability and interactivity of Federal Government datasets on the Brazilian Open Data Portal. **Informação & Sociedade: Estudos**; v. 30 n. 3 (2020); 1-19, v. 30, n. 3, p. 1–19, 2020.

SISTO, Raffaele; GARCÍA LÓPEZ, Javier; PAÉZ, José Manuel; MÚGICA, Elena Maté. Open Data Assessment in Italian and Spanish Cities. *In*: SPRINGER. **SMART AND SUSTAINABLE PLANNING FOR CITIES AND REGIONS: RESULTS OF SSPCR 2017 2**. [S.l.: s.n.], 2018. P. 121–131.

SOCRATA, Socrata Foundation. **Open Data & Citizen Engagement**. Fev. 2019a. Disponível em: <https://socrata.com/solutions/open-data-citizen-engagement/>.

SOCRATA, Socrata Foundation. **Socrata Data-as-a-Service Platform**. Acessado em: 11/02/2019. Fev. 2019b. Disponível em: [https://empower.tylertech.com/rs/015-NUU-525/images/Socrata\\_Data\\_as\\_a\\_Service\\_Platform\\_White\\_Paper.pdf](https://empower.tylertech.com/rs/015-NUU-525/images/Socrata_Data_as_a_Service_Platform_White_Paper.pdf).

SRIMUANG, Chatipot; COOHAROJANANONE, Nagul; TANLAMAI, Uthai; CHANDRACHAI, Achara. Open government data assessment model: An indicator development in Thailand. *In*: IEEE. **2017 19TH INTERNATIONAL CONFERENCE ON**

**ADVANCED COMMUNICATION TECHNOLOGY (ICACT)**. [*S.l.: s.n.*], 2017.  
P. 341–347.

STUERMER, Matthias; DAPP, Marcus M. Measuring the promise of open data: Development of the Impact Monitoring Framework. *In: IEEE. 2016 CONFERENCE FOR E-DEMOCRACY AND OPEN GOVERNMENT (CEDEM)*. [*S.l.: s.n.*], 2016.  
P. 197–203.

SUSHA, Iryna; GRÖNLUND, Åke; JANSSEN, Marijn. Driving factors of service innovation using open government data: An exploratory study of entrepreneurs in two countries. **Information Polity**, IOS Press, v. 20, n. 1, p. 19–34, 2015a.

SUSHA, Iryna; ZUIDERWIJK, Anneke; JANSSEN, Marijn; GRÖNLUND, Åke. Benchmarks for evaluating the progress of open data adoption: usage, limitations, and lessons learned. **Social Science Computer Review**, SAGE Publications Sage CA: Los Angeles, CA, v. 33, n. 5, p. 613–630, 2015b.

TAMIMI, Hatem; HOSHANG, Salam Amir; AL BLOOSHI, Essa Jasem. Analysis of UAE open government data usability within mobile application development. *In: IEEE. 2017 IEEE 2ND INTERNATIONAL CONFERENCE ON BIG DATA ANALYSIS (ICBDA)*. [*S.l.: s.n.*], 2017. P. 437–441.

THOMA, Matthias; KAKANTOUSIS, Theofilos; BRAUN, Torsten. Rest-based sensor networks with OData. *In: IEEE. WIRELESS ON-DEMAND NETWORK SYSTEMS AND SERVICES (WONS), 2014 11TH ANNUAL CONFERENCE ON*. [*S.l.: s.n.*], 2014.  
P. 33–40.

TORINO, Emanuelle; TREVISAN, Gustavo Lunardelli; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Dados abertos capes: A look at the challenges of web data publishing. **Ciencia da Informacao**, p. 38–46, 2019.

UMBRICH, Jürgen; NEUMAIER, Sebastian; POLLERES, Axel. Quality assessment and evolution of open data portals. *In: IEEE. 2015 3RD INTERNATIONAL CONFERENCE ON FUTURE INTERNET OF THINGS AND CLOUD*. [*S.l.: s.n.*], 2015.  
P. 404–411.

UTAMACHANT, Piriya; ANUTARIYA, Chutiporn. An analysis of high-value datasets: a case study of Thailand's open government data. *In: IEEE. 2018 15TH*

**INTERNATIONAL JOINT CONFERENCE ON COMPUTER SCIENCE AND SOFTWARE ENGINEERING (JCSSE).** [S.l.: s.n.], 2018. P. 1–6.

VAN BELLEN, Hans Michael. **Indicadores de sustentabilidade: uma análise comparativa.** [S.l.]: FGV Editora, 2005.

VICENTE-PANOS, Adrian; JORDAN-ALFONSO, Aurea. Access to and reuse of public information in Spanish regions: evaluation of open data reuse. **PROFESIONAL DE LA INFORMACION**, EPI APARTADO 32 280, BARCELONA, 08080, SPAIN, v. 26, n. 3, p. 381–391, 2017.

VIRGINIA BEACH, City of Virginia Beach. **Catalog.** Fev. 2019. Disponível em: <https://data.vbgov.com/browse>.

VON ALAN, R Hevner; MARCH, Salvatore T; PARK, Jinsoo; RAM, Sudha. Design science in information systems research. **MIS quarterly**, Springer, v. 28, n. 1, p. 75–105, 2004.

VOSTROVSKY, V; TYRYCHTR, J. Consistency of Open Data as Prerequisite for Usability in Agriculture. **Scientia agriculturae bohemica**, Sciendo, v. 49, n. 4, p. 333–339, 2018.

W3C. **About W3C.** Abr. 2021. Disponível em: <https://www.w3.org/Consortium/>.

W3C. **Standards.** Acessado em: 07/02/2019. Fev. 2019a. Disponível em: <https://www.w3.org/standards/>.

W3C. **W3C Arquitetura Web Semântica.** [S.l.: s.n.], 2019b. <https://www.w3.org/2007/03/layerCake.svg>. Acessado em: 28/01/2019.

W3C. **W3C Internationalized Resource Identifiers (IRIs).** [S.l.: s.n.], 2016. <https://www.w3.org/International/0-URL-and-ident.html>. Acessado em: 01/07/2016.

W3C, Consortium. **XML Technology.** [S.l.: s.n.], 2016. <http://www.w3.org/standards/xml/>. Acessado em: 29/06/2016.

WAGNER FILHO, Fernando; LÓSCIO, Bernadette Farias. **Web Semântica: Conceitos e Tecnologias**, 2016. Acessado em: 04/07/2016.

WEBER, Thomas; MITÖHNER, Johann; NEUMAIER, Sebastian; POLLERES, Axel. ODArchive—Creating an Archive for Structured Data from Open Data Portals. *In*: SPRINGER. **INTERNATIONAL SEMANTIC WEB CONFERENCE**. [S.l.: s.n.], 2020. P. 311–327.

WIENER, Martin; SOMMER, Friedrich T; IVES, Zachary G; POLDRACK, Russell A; LITT, Brian. Enabling an open data ecosystem for the neurosciences. **Neuron**, Elsevier, v. 92, n. 3, p. 617–621, 2016.

WIERINGA, Roel J. **Design science methodology for information systems and software engineering**. [S.l.]: Springer, 2014.

ZHANG, Hui; XIAO, Jianying. Quality assessment framework for open government data. **The Electronic Library**, Emerald Publishing Limited, 2020.

ZHU, Xiaohua; FREEMAN, Mark Antony. An evaluation of US municipal open data portals: A user interaction framework. **Journal of the Association for Information Science and Technology**, Wiley Online Library, v. 70, n. 1, p. 27–37, 2019.

## APÊNDICE A – MODELOS DE AVALIAÇÃO DE PORTAIS DE DADOS ABERTOS

### A.0.1 Proposta de Andreiuid Sh Corrêa *et al.* (2017)

Este trabalho propõe avaliar portais de dados abertos de 561 municípios brasileiros escolhidos aleatoriamente. Para a escolha dos municípios utilizou-se cálculo de amostragem, com base nos municípios registrados no Instituto Brasileiro de Geografia e Estatística (IBGE). Para a construção do instrumento de avaliação, os autores fizeram um relacionamento entre a Lei de Acesso à Informação (LAI), juntamente com os oito princípios de dados abertos. Deste modo, propuseram uma descrição curta dos critérios de avaliação. O Quadro 22 apresenta esse relacionamento entre LAI e os oito princípios, bem como a descrição curta denominada pelos autores.

Quadro 22 – Requisitos técnicos da lei brasileira e o mapeamento dos princípios da OGD

<b>Requisito técnico em linguagem jurídica (art. 8º § 3º da lei)</b>	<b>Princípios de OGD</b>	<b>Descrição Curta</b>
I - Contendo pesquisa de conteúdo que permite o acesso a informações de forma objetiva, transparente, clara e de fácil compreensão	4. Acessível 6. Não discriminatório	I - Pesquisa e acesso
II - Capacidade de salvar relatórios em múltiplos formatos, incluindo formatos abertos e não proprietários, como planilhas e texto para facilitar a análise das informações.	5. Processável por máquina 7. Não proprietário	II - Multi-formato
III - Permitir o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina.	5. Processável por máquina	III - legível por máquina
IV - Divulgar detalhes sobre os formatos utilizados para estruturação de informações.	1. Completo 2. Primária	IV - estrutura de dados
V - Garantir a autenticidade e integridade das informações disponíveis para acesso.	Nenhum mapeamento direto	-
VI - Manter as informações mais recentes disponíveis para acesso	3. Atuais	VI - dados atualizados

VII - Indicar localização e instruções para permitir que as pessoas se comuniquem com o proprietário da fonte de dados pela internet ou telefone	1. Completo 4. Acessível	VII - Informação de contato
VIII - Adotar os recursos necessários para garantir a acessibilidade de conteúdos para pessoas com deficiência	6. Não-Discriminatório	VIII- Acessibilidade
Não coberto pela lei	8. Os dados devem estar isentos de licença	-

Fonte: (CORRÊA, Andreiwid Sh *et al.*, 2017), tradução do autor

Como pode ser observado, a maioria dos princípios de OGD está relacionado com a LAI, contudo, o oitavo princípio (Livre de licença) não tem referência com a lei e segundo os autores este princípio confronta a lei de direitos autorais. Em relação à LAI o item (V - Garantir a autenticidade e a integridade das informações disponíveis para acesso) não está descrito na OGD e ficou fora da proposta de avaliação.

Para efetuar esta avaliação, os autores adotaram os critérios de avaliação como **Sim** para os dados divulgados de maneira correta, **Não** para os dados divulgados de maneira errada ou **Parcial** para os dados divulgados de maneira que não cumpre inteiramente a lei e/ou forneça informações incompletas. Por sua vez, o critério que avalia a acessibilidade, utiliza o *AccessMonitor*<sup>1</sup>, onde esta ferramenta efetua uma avaliação de forma automática de portais utilizando de um índice simples que varia de 1 à 10 com base em critérios de acessibilidade recomendados *WCAG*, sendo esta uma iniciativa da W3C .

É importante salientar que foi encontrado outro trabalho deste mesmo autor (CORRÊA, Andreiwid Sheffer *et al.*, 2014), todavia, não foi apresentado nesta subseção, pois o mesmo apresenta os mesmos critérios de avaliação.

#### A.0.2 Proposta de Máchová e Lnénicka (2017)

Este trabalho apresenta uma proposta de um *Framework de Benchmarking* para avaliar a qualidade dos portais de dados abertos. Esta proposta está sendo apresentada, pois a estrutura de avaliação é consistente e segue um rigor científico. No entanto, os aspectos que envolvem *Benchmark* não serão abordados por não serem o foco desta tese. O objetivo dos autores é avaliar diversos portais de dados

<sup>1</sup> <<http://www.acessibilidade.gov.pt/accessmonitor/>>

abertos em escala global utilizando uma série de critérios retirados da literatura. A seguir o Quadro 23 apresenta oriundo de uma revisão da literatura realizada pelos autores.

Quadro 23 – O framework de benchmarking para avaliação de qualidade de portais de dados abertos

<b>I.Características gerais de portal de dados abertos</b>	
Lista de métricas	Descrição dos requisitos para a avaliação da qualidade
<b>1. Dimensão técnica</b>	
1.1 Autoridade e responsabilidade	Portais devem fornecer informações sobre a autoridade, que hospeda o portal e do modelo de governo ou quadro institucional de apoio ao modelo de fornecimento de dados.
1.2 Dados de sistema de gestão	Portais devem fornecer informações sobre o sistema de gestão de dados, o que é utilizado para alimentar o portal de dados abertos.
1.3 Idioma	Portais devem oferecer mais versões linguísticas para ganhar mais usuários (atenção) e melhorar a qualidade geral deste portal.
1.4 Grátis	Portais devem prever que todos os conjuntos de dados e serviços estão disponíveis de forma gratuita e sem quaisquer restrições sob licenças abertas.
<b>2. Disponibilidade e dimensão de acesso</b>	
2.1 Número de conjuntos de dados	Portais devem fornecer o número de conjuntos de dados que incluem.
2.2 Número de pedidos (re-usos)	Portais devem fornecer o número de aplicações desenvolvidos com base nos dados abertos re-utilizado.
2.3 Pesquisa do motor (filtro)	Portais devem aprovar e tornar visível uma estrutura global de organização e fornecer recursos de pesquisa aos conjuntos de dados e ferramentas de seleção usando critérios diferentes para navegar através de categorias e navegar através de filtros.
2.4 API	Portais devem proporcionar API para partes interessadas de desenvolver aplicações utilizando dados abertos.
2.5 Conta de Usuário	Portais devem apoiar a criação de conta de usuário, a fim de personalizar pontos de vista e informações mostradas.
2.6 Categorias temáticas	Portais devem fornecer categorias temáticas dos conjuntos de dados fornecidos. O portal deve distinguir claramente as categorias (temas) de tags (palavras-chave)
2.7 Etiquetas (palavras-chave)	
<b>3. Comunicação e dimensão participação</b>	

3.1 Fórum (feedback)	Portais devem proporcionar uma oportunidade para enviar comentários sobre os dados dos usuários para provedores e fórum para discutir e trocar ideias entre os usuários.
3.2 Forma de Pedido	Portais devem proporcionar uma forma de solicitar ou sugerirem novos tipos ou formatos de dados abertos.
3.3 Ajuda (usabilidade)	Portais devem incluir alta qualidade da funcionalidade de documentação e ajuda para aprender a usar o portal e melhorar a usabilidade.
3.4 Perguntas Frequentes (FAQ)	Portais devem fornecer uma seção FAQ para ajudar a resolver possíveis problemas.
3.5 As mídias sociais	Portais devem ser conectados a uma plataforma de mídia social para criar um canal social de distribuição de dados abertos. Assim, os usuários de OGD podem se informar sobre o que fizeram e o que aprenderam a partir de um conjunto de dados.
<b>II.Características gerais do conjunto de dados</b>	
1. Título e descrição	Conjuntos de dados devem ser fornecidos juntamente com sua descrição e título.
2. Publicador	Conjuntos de dados devem ser fornecidos juntamente com a sua editora para verificar a autenticidade da sua origem.
3. Data de lançamento	Conjuntos de dados devem ser explicitamente associada à um momento específico ou um período. Todas as informações do conjunto de dados deve ser atualizado.
4. Licença	Conjuntos de dados devem fornecer informações sobre a licença relacionada com a utilização dos conjuntos de dados. Os conjuntos de dados que não têm explicitamente uma licença aberta não são dados aberto.
5. Cobertura geográfica	Conjuntos de dados devem ter explicito se os dados são a nível nacional, regional ou local.
6. URL	As URLs dos conjuntos de dados devem estar disponíveis.
7. Conjunto de dados (arquivo) tamanho	O tamanho dos conjuntos de dados (arquivo) devem estar disponível.
8. Número de visitas	Número total de visualizações on-line deve estar disponível para um conjunto de dados.
9. Número de Download	O número total de downloads devem estar disponíveis para um conjunto de dados.
10. Formatos de leitura	Conjuntos de dados devem ser fornecidos em formatos de fácil análise e em formatos modificáveis.

11. Visualização	Dispor da visualização dos conjuntos de dados ( <i>on-line</i> ), como visualizações em gráficos ou visualizações em mapas.
12. Comentários	Conjuntos de dados devem fornecer recursos que permitam coletar avaliações do usuário e comentários sobre um conjunto de dados ou para discutir conclusões com base no uso de dados.

Fonte: (MÁCHOVÁ; LNÉNICKA, 2017), tradução do autor

Para efetuar a avaliação dos portais e aplicar os critério apresentados no Quadro 23 os autores utilizaram um questionário que foi distribuído aos usuários. Os 28 critérios elencados foram utilizados em formas de perguntas onde foram avaliadas em uma escala *Likert* de cinco pontos para medir a concordância ou discordância, sendo esta, **1) Discordo Completamente, 2) Discordo, 3) Neutro, 4) Concordo e 5) Concordo completamente**. Desse modo, cada questão fornece uma pontuação de uma a cinco, sendo um resultado total variável de 28 a 140 pontos. Com esta pontuação, os autores realizam um comparativo e uma posterior análise.

### A.0.3 Proposta de Sisto *et al.* (2018)

Nesta proposta os autores avaliam portais de dados de algumas cidades espanholas e italianas. Para isso, apresentam uma estrutura com quatro dimensões, um score total do portal, quantidade de conjunto de dados, estrutura e um total de cada dimensão.

Para a estruturação do score sobre a quantidade de conjuntos de dados, os autores consideraram um range da quantidade de conjuntos de dados, sendo, 0-49 = 0 pontos 50-99 = 1 ponto, 100-199 = 2 pontos, 200-299 = 3 pontos e mais de 300 = 5. No quesito de estrutura, os autores levam em consideração quatro categorias, sendo estas:

- **Apresentação:** Acesso simplificado e funcional aos conjuntos de dados e descrição da missão e visão. Para representar esta categoria os autores utilizaram o critério "Apresentação";
- **Conjuntos de dados:** Possibilitar a pesquisa por palavras-chave, filtros, tópicos, tags, formatos e frequência de atualizações para garantir acessibilidade e navegabilidade, além de uma breve descrição dos formatos usados. Para representar esta categoria os autores utilizaram os critérios "Buscar, Filtros e Descrição de formatos";

- **Boas Práticas:** Participação de terceiros devem ser promovidas e medidas através de aplicativos oriundos de publicações, melhores práticas e exemplos de desenvolvimento baseados em aplicações de conjuntos de dados. Para representar esta categoria os autores utilizaram o critério "Boas Práticas";
- **Participação do cidadão:** Inclusão de seções como, perguntas frequentes, contato, dicas e fornecer uma área específica para publicação de aplicativos desenvolvidos por e/ou para os cidadãos. Para representar esta categoria os autores utilizaram os critérios "Solicitações de Dados, Ideia e resultados, Termos de uso e contato";
- **Informações gerais:** Refere-se aos termos, licenças e condições de uso dos conjuntos de dados que devem ser listados. Para representar esta categoria os autores utilizaram o critério "Avisos Legais";

Para que possa ser atribuído um score sobre a estrutura, os autores somam 1 ponto a cada critério presente no portal avaliado, podendo variar de 1-10 pontos.

Por sua vez, cada dimensão possui um conjunto de critério, sendo estes apresentados a seguir.

- **Dimensão Qualidade de Conteúdo**

- **Código Aberto:** os conjuntos de dados devem estar de acordo com o conceito de dados abertos (reutilização, redistribuição e uso livre permitido);
- **Livre:** inexistência de taxa para acessar os conjuntos de dados;
- **Oficial:** os conjuntos de dados devem ser publicados a partir de uma fonte do governo ou de uma fonte oficialmente reconhecida;
- **Cinco estrelas de dados abertos:** baseia-se no formato de dados do Tim Berners-Lee, apresentado na Figura 6 anteriormente;
- **Atualização:** os conjuntos de dados devem se manter atualizados levando em conta os metadados, descrição do nome do arquivo, as tags e a data de publicação;

- **Dimensão Acessibilidade**

- **Download em massa:** os conjuntos de dados podem ser baixados de uma só vez e em um único arquivo;
- **Idiomas:** Disponibilidade de vários idiomas, sem levar em consideração o idioma nativo;
- **Site do município:** um link direto para o portal do conselho da cidade;

- **Dados on-line e para download:** os dados não podem ser enviados por um canal específico, exemplo, envio por e-mail;

- **Dimensão Natureza Transversal**

- **Economia:** comércio, finanças, emprego, impostos, indústria, turismo e habitação.
- **Mobilidade:** transporte e infraestrutura urbana.
- **Meio ambiente:** energia e meio ambiente.
- **Cidadãos:** demografia e educação.
- **Qualidade de vida:** ciências, tecnologia, cultura, lazer, segurança, saúde, esporte, sociedade e bem-estar.
- **Governança:** lei, justiça, setor público.

- **Dimensão Visualização**

- **Painel on-line:** verifica a existência de um painel online para vários conjuntos de dados;
- **Pré-visualização dos dados:** verifica a possibilidade dos conjuntos de dados que podem ser visualizados antes do download;
- **Formato legível:** conjuntos de dados legíveis para usuários não técnicos, apenas formato de dados abertos considerados;
- **Download médio:** verifica a existência do download médio dos conjuntos de dados baixados;

As dimensões apresentam um total parcial, com base em um valor máximo, médio e mínimo dada por uma escala entre o total parcial e a quantidade de critérios sendo estas arredondadas.

#### **A.0.4 Proposta de Xiaohua Zhu e Freeman (2019)**

Neste trabalho os autores tem como objetivo avaliar os portais de OGD de 34 agências governamentais municipais dos EUA. Para realizar esta avaliação os autores propuseram um *Framework* de interação de usuário, deste modo, é possível avaliar além de algumas funcionalidades do portal também alguns aspectos de usabilidade. Para a construção do instrumento de avaliação proposto, os autores realizaram uma pesquisa na literatura e após a identificação dos critérios, realizaram uma análise considerando o contexto de OGD. Deste modo, possibilitou a identificação das dimensões com base em estudos encontrados na literatura e na observação dos autores. Com

isso, os autores chegaram em uma estrutura com 5 dimensões 30 critérios e 42 perguntas que auxiliaram na avaliação dos portais. A seguir o Quadro 24 apresenta a estrutura proposta pelos autores.

Quadro 24 – Quadro de avaliação de envolvimento do usuário

Dimensão	Critérios	Operacionalização (itens de medição)
Acesso	Organização de dados	A1a. Os usuários podem procurar conjuntos de dados por categorias?
		A1b. Os usuários podem ver uma lista de todos os conjuntos de dados disponíveis?
		A1c. Os usuários podem filtrar e / ou classificar esses conjuntos de dados?
	Pesquisabilidade	A2a. Os usuários podem pesquisar conjuntos de dados?
		A2b. O portal oferece opções avançadas de pesquisa (por exemplo, booleano, pesquisa de campo, filtragem de resultados e / ou classificação de resultados)?
	Livre de restrições	A3 Os usuários NÃO precisam se inscrever para usar os dados?
	Licença	A4 Existe uma licença explícita de cc-by ou equivalente?
	Múltiplos idiomas	A5 O portal é apresentado em outra língua que não o inglês?
	Processabilidade da máquina	A6 Todos os conjuntos de dados estão disponíveis em formato(s) processável(s) por máquina?
	Formatos abertos	A7a Todos os conjuntos de dados estão disponíveis em ao menos um formato aberto?
		A7b. Todos os conjuntos de dados estão disponíveis em vários formatos abertos?
		A7c Algum dos conjuntos de dados inclui algum arquivo .exe?
	URI permanente	A8 Cada conjunto de dados tem um URI permanente?

Confiabilidade	Completude	B1a O número total de conjuntos de dados está disponível?
		B1b. O portal inclui as categorias mais comuns - saúde, segurança pública e transporte?
		B1c. O portal diferencia entre vários tipos de dados (como conjuntos de dados, documentos, formulários, calendários e assim por diante)?
	Atualidade	B2 Alguma coisa foi atualizada ou adicionada nos últimos 3 meses?
	Disponibilidade de política de dados	B3 A política de dados está disponível no portal?
	Fontes	B4 Todos os conjuntos de dados podem ser rastreados até sua fonte primária (como agências governamentais específicas que produziram os dados)?
	Granularidade dos dados	B5 Todos os conjuntos de dados incluem dados mais granulares que os dados no nível da cidade?
Relevância	B6a. Existe alguma indicação de quantas vezes cada conjunto de dados foi baixado?	
	B6b. Existe alguma indicação de quantas vezes cada conjunto de dados foi acessado?	
Entendimento	Suporte ao usuário	C1a. As informações de ajuda / FAQ estão disponíveis no portal?
		C1b. Há informações de contato para os usuários entrarem em contato com os gerenciadores de portais e / ou gerenciadores de conjuntos de dados?
	Vitrine de aplicativos	C2. Existem alguns exemplos de aplicativos disponíveis?
Documentação	C3a. Algum dos conjuntos de dados é acompanhado por documentação de suporte especificando os métodos de coleta de dados?	

		C3b. Algum dos conjuntos de dados são acompanhados por um dicionário de dados detalhado com definições de termos / variáveis / campos?
	Metadados	C4. Os metadados estão disponíveis para todos os conjuntos de dados?
Engajar-integrar	Disponibilidade de análises	D1. O portal fornece análises sobre como os dados ou o portal são usados / acessados?
	Disponibilidade de api	D2. A API está presente no portal?
	Disponibilidade de formato de citação	D3. O portal sugere um formato de citação?
	Personalização	D4. Os usuários podem personalizar o portal para permitir recursos como salvar, histórico de pesquisa e assim por diante?
	Download	D5a. Todos os conjuntos de dados estão disponíveis para download?
		D5b. Uma opção de impressão é fornecida para cada conjunto de dados?
		D5c. O portal fornece uma opção de download em lote?
	Manipulação online	D6. O portal fornece ferramentas para manipulação on-line, como filtragem e classificação em todos os conjuntos de dados?
Visualização online	D7. O portal fornece mapa ou outras ferramentas de visualização para qualquer um dos conjuntos de dados?	
Conjuntos de dados comparativos	D8. Para qualquer conjunto de dados, o portal fornece conjuntos de dados comparativos ou recomenda conjuntos de dados relacionados?	
Participar	Engajamento proativo	E1. Existe envolvimento proativo com a comunidade (boletins informativos / e-mails / podcasts / mídias sociais)?

	Compartilhamento	E2. Existem ferramentas disponíveis para a comunidade compartilhar conjuntos de dados?
	Participação	E3. A comunidade pode comentar / discutir / avaliar conjuntos de dados?
	Feedback do usuário	E4. O portal solicita que os usuários sugiram alterações nos conjuntos de dados existentes e / ou sugiram novos conjuntos de dados?

Fonte: (ZHU, X.; FREEMAN, 2019), tradução do autor

Para realizar as avaliações os autores consideraram como resposta dos critérios **Sim (1)** e **Não (0)** e para cada dimensão foi ponderada como 20% do índice de 100 pontos. Além da análise descritiva, a análise estatística básica foi realizada com base no score total de cada dimensão.

#### A.0.5 Proposta de ODM (2019)

O *Open Data Monitor (ODM)* é um portal de dados criado pela União Europeia de demais parceiros, a fim de apresentar informações sobre os portais de dados e que possibilite a análise por parte da população e a realização de *benchmark*. Os aspectos de *benchmark* não serão considerados.

Para atender este objetivo o ODM propôs uma metodologia na qual utiliza para apresentar essas informações aos usuários, possibilitando a comparação de portais de dados abertos. É importante salientar que este estudo foi selecionado porquê possui critérios que auxiliam em aspectos quantitativos dos portais de dados. Esta metodologia é dividida em três partes, sendo estas:

- **Colheita:** É o processo automatizado de coleta de metadados dos portais de dados abertos europeus.
- **Harmonização:** Esta etapa realiza a transformação dos metadados para que possa ser posteriormente analisados.
- **APIs:** A ODM conta com um serviço de API, disponibilizando os resultados a terceiros que desejam usufruir da plataforma.

A análise dos metadados através da plataforma ODM utiliza dois agrupamentos, sendo um de quantidade e outro de qualidade. A métrica de qualidade, remete-se à tentativa de quantificar quão bons e úteis são os metadados coletados para os

usuários finais. A seguir serão abordadas as métricas de qualidade e como é realizada a avaliação de cada uma, sendo estas:

- **Licença Aberta:** Representa a contagem total de licenças existentes em relação ao total de metadados com uso de licença.
- **Legível por máquina:** Representa a contagem da legibilidade de um conjunto de dados ao calcular a legibilidade por máquina de seus formatos.
- **Completeness dos metadados:** É a média de metadados ausentes em um conjunto definido de campos: licença, autor, organização e a existência de uma das datas: criada ou atualizada.
- **Disponibilidade:** É o número de conjuntos de dados qualificados como disponíveis publicamente sobre o número total de conjuntos de dados em um catálogo.
- **Descoberta:** É uma estimativa da importância de um catálogo na Web com base em dois sistemas de classificação de tráfego: Google e Alexa.
- **Formatos abertos:** É a contagem total da diversidade dos recursos dos conjuntos de dados com um formato não proprietário sobre a contagem total de recursos com um formato;
- **Pontuação geral de qualidade:** É definida como a média das métricas mais relevantes, como licenças abertas, legibilidade por máquina, disponibilidade e integridade de metadados.

A métrica de quantidade tem por objetivo apresentar um relatório de estatística contando ou somando os atributos selecionados dos metadados coletados. A seguir serão apresentadas as métricas referentes ao agrupamento de quantidade.

- **Tamanho total da distribuição:** Tamanho total de todos os recursos independente do seu formato;
- **Número de distribuição:** Quantidade total de todos os conjuntos de dados, referente à diversidade de seus recursos, presente em um catálogo específico;
- **Número de conjunto de dados:** Quantidade total de conjuntos de dados disponíveis em um catálogo;
- **Número de editores exclusivos:** Quantidade total de organizações que publicam em um catálogo próprio;
- **Catálogos colhidos:** Quantidade de catálogos por país coletado e harmonizado;

### A.0.6 Proposta de Abella *et al.* (2019)

A proposta das Métricas para Liberar Dados Abertos (MELODA) tem como objetivo auxiliar os editores de dados a liberarem dados reutilizáveis para que os dados possam ser úteis para a criação de novos produtos e serviços. Com isso, o MELODA possibilita avaliar os conjuntos de dados independente da origem de publicação (governo, empresas privadas, dados científicos, entre outros). Para que seja realizada a avaliação, o MELODA conta com uma estrutura de seis dimensões e alguns critérios e estes por sua vez com pesos específicos. A seguir no Quadro 25 serão apresentadas as dimensões bem como seus critérios e pesos.

Quadro 25 – Dimensões e critérios MELODA

<b>Dimensão Legal</b>	
Descrição	Peso %
Nível 1. Direitos autorais. Fontes de dados neste nível: Restringir a reutilização sem acordo por escrito. Ele restringe o uso não autorizado Exemplos: 1) Copyright	0
Nível 2. Uso Privado: As fontes de dados nesse nível permitirão o uso dos dados sem processos de aprovação, mas apenas para usos privados. Não uso público disponível.	10
Nível 3. Reutilização não comercial: As fontes de dados neste nível permitirão a reutilização de dados, mas não permitirão usos comerciais da reutilização de informações. Exemplos: 1) (CC BY-NC-ND 4.0) - Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 2) (CC BY-NC 4.0) Atribuição-Uso Não-Comercial 4.0 Internacional 3) (CC BY-NC-SA 4.0) Atribuição-Uso Não-Comercial-Compartilhamento pela mesma licença 4.0 Internacional	25
Nível 4. As fontes de dados neste nível permitirão a reutilização de dados, incluindo a reutilização comercial. (ou seja, para qualquer empresa, a fim de criar nova visualização de dados) Exemplos: 1) (CC BY-ND 4.0) Atribuição-SemDerivativos 4.0 Internacional 2) Atribuição-Compartilhamento pela mesma licença 4.0 Internacional (CC BY-SA 4.0)	90
Nível 5. Sem restrições ou somente atribuição: As fontes de dados nesse nível só solicitarão aos reutilizadores a atribuição dos dados. Exemplos: 1) Atribuição 4.0 Internacional (CC BY 4.0)	100
<b>Dimensão Padrões Técnicos</b>	
Descrição	Peso %
Nível 1. Padrão fechado não reutilizável: As fontes de dados neste nível são lançadas sobre padrões proprietários não focados na reutilização de dados (ou seja, formato de imagem para dados ou padrões de documentos). Exemplos: 1) imagem pdf 2) doc	10

Nível 2. Padrão fechado reutilizável e aberto não reutilizável: Outros formatos fechados de alguma forma adequado para reutilização (formatos proprietários de dados 3) shp 4) xls y xlsx com macro / fórmula	35
Nível 3. Padrão reutilizável aberto: As fontes de dados nesse nível são publicadas em padrões abertos, mas como arquivos individuais. Exemplos: 1) csv sem mais informações 2) txt 3) odb 4) odt 5) ods 6) WMS 4) xls y xlsx sem macro / fórmula	60
Nível 4. Metadados individuais padrão abertos: As fontes de dados nesse nível incluem aquelas que fornecem metadados anexados a quaisquer dados. Dados semânticos estão neste nível Exemplos: 1) rdf, 2) rss, 3) json 4) xml (com metatags)	100
<b>Dimensão Acesso a Informação</b>	
Descrição	Peso %
Nível 1. Sem acesso à Web ou solicitação manual: O acesso à informação requer um processo de aprovação não automático para acesso a informações de dados ou para registrar dados em um formulário não digital. Exemplos 1) É necessário um email não automático ou uma interação pessoal para obter acesso	0
Nível 2. URL de acesso à Web com registro ou com interação na Web: Acesso a informações pela Web, mas requer interação do usuário para selecionar a fonte de dados. Exemplos 1) É necessário interagir manualmente para obter acesso 2) Existe uma limitação no número de dados a serem acessados em um único disparo	10
Nível 3. Acesso à Web ou parâmetros de URL exclusivos ao conjunto de dados: Acesso a informações pela Web, mas permite que cada um dos conjuntos de dados seja acessado individualmente ou por meio de um URL exclusivo ou por parâmetros específicos na chamada de consulta Exemplos 1) um URL único e constante 2) Um URL com parâmetros (mas sem variação nos parâmetros) para o mesmo conjunto de dados	50
Nível 4. Acesso Web exclusivo com parâmetros para dados únicos: Acesso a informações pela Web, mas permite que cada dado dentro dos conjuntos de dados seja acessado individualmente ou por meio de um URL abreviado ou por parâmetros específicos na consulta e inclua a data, a versão ou a última atualização. Dados vinculados atendem a esse nível e o próximo Exemplos 1) url com parâmetros pode endereçar dados individuais sem fazer o download de todo o conjunto de dados	90
Nível 5. API ou linguagem de consulta: O acesso a informações fornece acesso a dados específicos do conjunto de dados, seja chamando uma API documentada ou por meio de uma linguagem de consulta de fontes de dados Exemplos 1) Um ponto SPARQL a ser consultado ou uma API documentada (CKAN, socrata, ODS API)	100

<b>Dimensão Compartilhamento de dados</b>	
Descrição	Peso %
Nível 1. Modelo de dados desconhecido: O modelo de dados não é liberado nem explicado.	15
Nível 2. Modelo próprio de dados ad hoc: Existem campos criados pelo editor. Mas os campos são identificados apenas, não explicando os intervalos, tipo e resto de características, etc. Ele é usado apenas pela entidade editora.	35
Nível 3. Modelo próprio de dados ad hoc publicado: Embora o modelo seja projetado pelo editor, a especificação dos campos da informação está disponível como uma informação separada, e é livremente permitida a ser usada por outros. Ontologia caseira pertence a este nível. Exemplos 1) Existe uma ontologia ou descrição de publicação disponível gratuitamente	50
Nível 4. Modelo de dados abertos locais: Existe um modelo padronizado por um organismo de normalização / entidade local (ou seja, nacional), mas é mal adotado (ou seja, novos padrões, padrões não amplamente adotados). O uso de uma ontologia pública nacional pertence a este nível.	90
Nível 5. Modelo global de dados abertos: Existe um modelo de dados padronizado divulgado por uma entidade global (ou seja, supranacional, ISO) e / ou há uma ampla adoção (ou seja, quando é obrigatório por lei, etc.)	100
<b>Dimensão Informação geolocalizada</b>	
Descrição	Peso %
Nível 1. Nenhuma informação geográfica: A informação divulgada não possui nenhum campo referente ao local	15
Nível 2. Campo de texto simples: As informações geográficas são apenas um campo de texto ou um ID próprio que dificulta a conexão dessas informações com outros bancos de dados. Exemplos 1) Nome do campo: Cidade, valor: Astorga	30
Nível 3. Campo de texto complexo: As informações geográficas são vários campos com descrição de texto (o número de uma rua seria considerado como um texto). E os campos são hierárquicos Exemplos 1) Campo1: País, valor: Espanha, Campo2: Cidade, valor: Astorga	50
Nível 4. Coordenadas: A informação geográfica inclui dois campos com as coordenadas (em algum sistema padronizado de coordenadas) Exemplos 1) Campo1, Latitude, valor: 42.605556, longitude de campo2, valor: -5.5	90
Nível 5. Informação geográfica completa: Inclui os níveis 3 e 4 Exemplos 1) Campo1, Latitude, valor: 42.605556, longitude 2, valor: -5.5, campo 3: País, valor: Espanha, campo 4: cidade, valor: Leão	100
<b>Dimensão Informação em tempo real</b>	
Descrição	Peso %

Nível 1. Mais do que semana: O período de atualização é maior que uma semana. 1) A diferença entre uma versão e a próxima é maior que 7 dias 2) ou seja, mensal, anual	15
Nível 2. Dias : O período de atualização varia de 1 dia a 7 dias. Exemplos 1) A diferença entre uma versão e a próxima é maior que 1 dia mas menor que 7 dias 2) Lançamento semanal	40
Nível 3. Horas: O período de atualização varia de 1 hora a 24 horas. Exemplos 1) A diferença entre uma versão e a próxima é maior que 1 dia mas menor que 7 dias 2) Lançamento diário	70
Nível 4. Minutos: O período de atualização varia de 1 hora a 1 minuto. Exemplos 1) A diferença entre uma versão e a próxima é maior que 1 minuto mas menor que 1 hora 2) liberação por hora	90
Nível 5. Segundos: O período de atualização é menor que 1 minuto. É um dataJet. Exemplos 1) A diferença entre uma versão e a próxima é menor que 1 minuto 2) Período de atualização medido em segundos	100

Fonte: (ABELLA *et al.*, 2019), tradução do autor

Para efetuar a avaliação é necessário escolher um nível de cada dimensão e após aplicar o seguinte calculo:

$$100.\sqrt[6]{D1.D2.D3.D4.D5.D6}$$

Onde, D1 a D6 correspondem ao nível selecionado para cada uma das dimensão com seus respectivos pesos. Após aplicar o cálculo e obter-se os resultados, utiliza-se do Quadro 26 a fim de verificar em qual escala esta os conjuntos de dados avaliados.

Quadro 26 – Qualificação global baseada em resultados de avaliação

Escala	Categorias	Termo simples
0 a 25	Inadequado para reutilização	Reutilização limitada
25 a 50	Reutilização básica possível	Reutilização básica
50 a 75	Reutilização possível, mas com alguma característica de melhoria	Boa reutilização
75 a 100	O melhor para reutilizar	Reutilização ideal

Fonte: (ABELLA *et al.*, 2019), tradução do autor

Durante o processo de pesquisa, foram encontrados trabalhos que abordam o *benchmark* como meio de avaliar portais de dados abertos, como é o caso da *Open Data Barometer*<sup>2</sup> e o *Global Open Data Index*<sup>3</sup>. Estes trabalhos tem como objetivo

<sup>2</sup> <<https://opendatabarometer.org/>>

<sup>3</sup> <<https://index.okfn.org/>>

comparar e avaliar o desempenho entre países a fim de identificar qual se sobressai em relação aos outros, baseado em alguns critérios. Deste modo o *benchmark* pode ser definido como "o uso de métricas de desempenho para avaliar e comparar diferentes organizações ou países" (SUSHA *et al.*, 2015b). Portanto, trabalhos que utilizam deste contexto não serão utilizado nesta pesquisa.

## APÊNDICE B – DEFINIÇÕES EM DETALHES DO MODELO DE REFERÊNCIA PROPOSTO

### B.0.1 Dimensão Acesso

Esta dimensão tem como objetivo verificar a capacidade dos *stakeholders* acessarem os dados publicados no portal. Garantir o acesso aos dados publicados a seres humanos e máquinas é de extrema importância, pois desta forma, os dados abertos podem ser explorados e descobertos.

#### B.0.1.1 Indicador Atualização e suas variáveis

Este indicador tem por objetivo verificar a atuação das organizações em disponibilizar novos conjuntos de dados e seus recursos, bem como a sua atualização. Isso ajuda, por exemplo, a compreender se as organizações possuem um equilíbrio entre criar e atualizar os conjuntos de dados ou são tendenciosas a somente criar os conjuntos de dados. Outro ponto a ser destacado é ao avaliar um portal de dados somente considerando a atualização existe a possibilidade de ocasionar uma avaliação imprecisa, visto que o portal pode adicionar mais conjuntos de dados novos do que atualizar. Isso geralmente ocorre em portais que possuem diretrizes que a cada nova atualização cria-se um novo conjunto de dados.

A literatura aponta que as atualizações dos dados sejam sempre que possíveis disponibilizadas e que se utilize versionamento para não prejudicar os usuários que utilizam dos conjuntos de dados (LÓSCIO, B. F. *et al.*, 2017; CORRÊA, Andreiuid Sh *et al.*, 2017; SISTO *et al.*, 2018). No entanto, Xiaohua Zhu e Freeman (2019) consideram dados atualizados apenas os últimos três meses. Contudo, no estudo realizado por Global Partnership for Sustainable Development Data (2016), é mencionado que as organizações precisam manter a quantidade de dados disponibilizados. Desta forma, para este estudo será considerado o período de doze meses para verificação da atualização.

Dados as diferentes recomendações da literatura, entende-se que as colocações de Global Partnership for Sustainable Development Data (2016), Bernadette Farias Lóscio *et al.* (2017), Andreiuid Sh Corrêa *et al.* (2017) e Sisto *et al.* (2018) são satisfatórias, pois, possibilitam vislumbrar a atividade do portal em relação a criação e atualização dos conjuntos de dados, visto que ambas as colocações se complementam. O Quadro 27 descreve as variáveis que irão compor este indicador.

Quadro 27 – Descrição das variáveis, indicador atualização

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Conjuntos de Dados - Atualização nos últimos 12 meses	Dados com uma atualização regular e distribuída mostram o comprometimento do órgão em relação aos seus dados e transparece confiança aos <i>stakeholders</i> . Esta variável verifica a quantidade de conjuntos de dados e recursos atualizados em doze meses.	Dados atualizados	Fornecer dados Atualizados (BP 21)
Recursos - Adicionados nos últimos 12 meses			
Conjuntos de Dados - Adicionados nos últimos 12 meses	Dados adicionados demonstram a atividade do portal e o comprometimento em disponibilizar novos dados, mesmo sendo dados que complementam (atualizam) conjuntos de dados existentes. Esta variável verifica a quantidade de conjuntos de dados e recursos adicionados em doze meses.	—	Fornecer dados Atualizados (BP 21)
Recursos - Adicionados nos últimos 12 meses			

Fonte: O Autor

Na concepção do modelo de referência percebeu-se que alguns portais adicionam novos conjuntos de dados com objetivo de complementar outros já existentes ou com pequenas alterações, originando assim novas versões. Sendo assim, entende-se que um novo conjunto de dados, sendo este uma variação de outro conjunto de dados só que de diferente versão, também pode ser considerado uma atualização.

### B.0.1.2 Indicador Download em Massa e suas variáveis

Este indicador tem por objetivo verificar a existência de recursos disponibilizados em formatos que possibilitem o *download* de grandes volumes de dados em uma única requisição. Desta forma, traz maior confiança aos dados, dificultando uma manipulação inconsistente ao criar conjuntos de dados oriundos de uma coleta individual, por exemplo. Isso proporciona melhor organização dos dados, aumentando a sua consistência e confiança perante os *stakeholders*.

A literatura aponta algumas abordagens para o emprego de *download* em massa, sendo estas, a possibilidade de dados compactados (MÁCHOVÁ; LNÉNICKA, 2017; ZHU, X.; FREEMAN, 2019; ABELLA *et al.*, 2019), disponibilização dos dados através de meio diferentes ao HTTP, como exemplo do *File Transfer Protocols* (FTP) e acesso via API (LÓSCIO, B. F. *et al.*, 2017). O Quadro 28 irá descrever as variáveis que irão compor este indicador.

Quadro 28 – Descrição das variáveis, indicador download em massa

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Recursos disponíveis - API ou FTP	Recursos disponibilizados em API possibilitam a captura instantânea de dados completos dinâmicos ou mesmo o acesso aos dados quase que instantaneamente. A disponibilização em FTP permite o acesso a grandes volumes de dados em uma única solicitação. Esta variável verifica a quantidade de recursos disponíveis em API ou FTP.	Disponibilização de API	Disponibilizar dados por meio de uma API – (BP 23); Fornecer dados em tempo real (BP 20);

Recursos compactados disponibilizados no portal	Recursos disponibilizados com formato de arquivo compactado (ZIP, RAR, TAR, GZIP, outros) possibilitam à <i>stakeholders</i> coletar grandes volumes de dados de forma simples e mantêm a organização dos dados no portal. Esta variável verifica a quantidade de recursos compactados.	Download	Fornecer download em massa (BP 17)
---	---	----------	------------------------------------

Fonte: O Autor

É importante salientar que a disponibilização de recursos em API ou FTP demandam uma infraestrutura e conhecimentos técnicos. Esta maneira é mais adequada visto que possui toda uma arquitetura para a transmissão de grandes volumes de dados. Contudo, a disponibilização de recursos em formato compactado pode ser criada de forma simples e sem conhecimento técnico aprofundado e consegue manter uma organização de vários dados em um único arquivo compactado, facilitando também a transmissão de grandes volumes de dados.

Entende-se que ambas as variáveis aqui apresentadas não são excludentes e como a literatura não explicita uma quantidade, compreende-se que o portal de dados quando possível deve disponibilizar meios que possibilitem o download em massa.

### B.0.1.3 Indicador Metadados e suas variáveis

Este indicador verifica a existência de metadados em formato *Hypertext Markup Language* (HTML) e em formatos serializados (legível por máquina). Os metadados são um meio necessário e fundamental para a descoberta dos dados na internet; por este motivo, é importante que os conjuntos de dados e os recursos tenham a maior quantidade de metadados possíveis (LÓSCIO, B. F. *et al.*, 2017).

Todas as informações atreladas aos metadados contribuem com a compreensão do seu significado, conteúdo e de sua estrutura isso possibilita aos *stakeholders* localizarem e utilizarem conjuntos de dados, além de esclarecer questões sobre: licença, frequência de atualização, a quem pertence os dados, dentre outros (LAUFER, 2015). O Quadro 29 descreve as variáveis que irão compor este indicador.

Quadro 29 – Descrição das variáveis, indicador Metadados

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Conjuntos de Dados - Metadados (Formato HTML)	Estes metadados estão visíveis em páginas web, de modo, a proporcionar a visualização dos metadados pelos <i>stakeholders</i> logo ao acessarem o conjunto de dados ou os recursos no portal. Esta variável verifica a quantidade de metadados em HTML disponíveis em conjuntos de dados e recursos.	Metadados	Fornecer metadados – (BP 01)
Recursos - Metadados (Formato HTML)			
Conjuntos de Dados - Metadados (Formato serializado)	Estes metadados estão acessíveis normalmente em plataformas que dão suporte aos portais de dados e permitem o acesso legível por máquinas, possibilitando o seu fácil processamento. Esta variável verifica a quantidade de metadados serializados disponíveis em conjuntos de dados e recursos.		
Recursos - Metadados (Formato serializado)			

Fonte: O Autor

A literatura destaca os metadados como um requisito fundamental para a publicação de conjuntos de dados na Web, sendo imprescindível sua presença. Desta forma, o emprego de metadados em todos os conjuntos de dados e recursos disponibilizados no portal se faz necessário (LÓSCIO, B. F. *et al.*, 2017; ZHU, X.; FREEMAN, 2019; ABELLA *et al.*, 2019; ODM, 2019).

#### B.0.1.4 Indicador Metadados Descritivos e suas variáveis

Este indicador verifica os metadados descritivos, referentes aos conjuntos de dados e seus recursos. Com isso, é possível identificar a sua natureza, possibilitando a compreensão, bem como, revelar informações sobre o proprietário dos dados, o autor, título e descrição, dentre outros (CORRÊA, Andreiuid Sh *et al.*, 2017; MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018; ZHU, X.; FREEMAN, 2019; ODM, 2019).

Este indicador é um complemento do indicador anterior – seção B.0.1.3 desta forma é possível descrever os conjuntos de dados e recursos utilizando geralmente estrutura chave-valor, possibilitando um melhor refinamento para a descoberta e exploração dos dados. O Quadro 30 apresenta os metadados descritivos citados pela Bernadette Farias Lóscio *et al.* (2017) e a descrição das variáveis que irão compor o indicador.

Quadro 30 – Descrição das variáveis, indicador metadados descritivos

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Conjuntos de Dados - Metadados (Nome)	A atribuição de um título (nome) ou uma descrição contendo informações complementares sobre o conjunto de dados e recursos, é extremamente importante para a sua localização e acesso. Esta variável verifica a quantidade de metadados que possuem valores relacionados ao nome.	Título e Descrição; Granularidade dos dados	Fornecer Metadados descritivos (BP 02) – Título e a Descrição
Metadados descrição (Conjunto de Dados e Recursos)			
Metadados data criação (Conjunto de Dados e Recursos)	A atribuição da data que foi criado, permite identificar temporalmente os conjuntos de dados e seus recursos. Esta variável verifica a quantidade de metadados que possuem valores relacionados a data criação.	Granularidade dos dados	Fornecer Metadados descritivos (BP 02) – Data de publicação de dados

Metadados Palavras-chave	Atribuir palavras-chave que auxiliem na localização do conjunto de dados. Esta variável verifica a quantidade de metadados que possuem valores relacionados a palavra-chave.	Palavra-chave; Granularidade dos dados	Fornecer Metadados descritivos (BP 02) – Palavras-chave
Metadados mantenedor	A atribuição do nome da entidade ou pessoa responsável sobre o conjunto de dados publicados, possibilita identificar quem responderá oficialmente pelos dados disponibilizados. Esta variável verifica a quantidade de metadados que possuem valores relacionados ao mantenedor.	Informações de publicadores; Granularidade dos dados	Fornecer Metadados descritivos (BP 02) → Entidade responsável (publicadora de dados)
Metadados Autor	A atribuição do nome do autor que criou o conjunto de dados, permite identificar o responsável pela criação dos conjuntos de dados e recursos. Esta variável verifica a quantidade de metadados que possuem valores relacionados ao autor.		Fornecer Metadados descritivos (BP 02) – Ponto de contato
Metadados mantenedor e-mail	A atribuição de dados que identifiquem a entidade responsável pelo conjunto de dados, como exemplo, e-mail, telefone, etc, facilita o contato entre publicadores e <i>stakeholders</i> . Esta variável verifica a quantidade de metadados que possuem valores relacionados ao e-mail do mantenedor.		Fornecer Metadados descritivos (BP 02) → Entidade responsável (publicadora de dados)

Metadados autor e-mail	A atribuição de dados que identifiquem o autor responsável pelo conjunto de dados, como exemplo, e-mail, telefone, etc, permite uma comunicação direta com o autor possibilitando a solução de dúvidas dos <i>stakeholders</i> , por exemplo. Esta variável verifica a quantidade de metadados que possuem valores relacionados ao e-mail do autor.		Fornecer Metadados descritivos (BP 02) – Ponto de contato
Metadados formato de arquivo	Atribuir o formato de arquivo (recurso) que está sendo disponibilizado. Esta variável verifica a quantidade de metadados que possuem valores relacionados ao formato de arquivo.	Formato de dados; Granularidade dos dados	Fornecer Metadados descritivos (BP 02) – Formato de mídia
Metadados - Tamanho de arquivo	Atribuir o tamanho total em KB, MB ou GB dos recursos disponibilizados. Esta variável verifica a quantidade de metadados que possuem valores relacionados ao tamanho de arquivo.	Tamanho do conjunto de dados; Granularidade dos dados	—

Fonte: O Autor

Segundo Bernadette Farias Lóscio *et al.* (2017) recomenda-se o uso de alguns atributos para explicitar as informações sobre os conjuntos de dados, mas também é possível adicionar outros para complementar e melhor descreve-los. Devido a importância deste indicador, faz-se necessário a aplicação deste em todos os conjuntos de dados e recursos.

#### B.0.1.5 Indicador Metadados Estrutural e suas variáveis

Este indicador verifica se existem metadados estruturais, que apresentam informações sobre a estrutura dos recursos. Os metadados estruturais podem ser visualizados de três maneiras, sendo: **i)** através de uma página em HTML, que normalmente

está localizado próximo aos metadados descritivos; **ii)** através de arquivos, sendo estes, planilhas, documentos de texto, arquivo de texto e em alguns casos links HTML ou **iii)** através de formato serializado, possibilitando a legibilidade por máquinas.

A literatura coloca que ao fornecer os metadados estruturais facilita-se a compreensão da estrutura interna de um recurso, possibilitando a exploração dos dados e o entendimento sobre o significado dos dados (LÓSCIO, B. F. *et al.*, 2017). O Quadro 31 descreve as variáveis que irão compor este indicador.

Quadro 31 – Descrição das variáveis, indicador metadados estruturais

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Metadados estrutural (Formato HTML)	Este tipo de disponibilidade de metadados estrutural está visível na página web. De modo, a proporcionar aos <i>stakeholders</i> a compreensão da estrutura dos recursos, possibilitando uma interpretação adequada dos dados disponibilizados. Esta variável verifica a quantidade de metadados estruturais em HTML que estão disponíveis.	—	Fornecer Metadados Estruturais (BP 3)
Metadados estrutural (Formato serializado)	Este tipo de disponibilidade de metadados serializado está normalmente disponível em plataformas que dão suporte aos portais de dados. Este meio permite o acesso legível por máquinas, facilitando o processamento. Esta variável verifica a quantidade de metadados estruturais serializados que estão disponíveis.	—	

Metadados estrutural (Documentos separados)	Este tipo de disponibilidade de metadados estrutural é geralmente disponibilizado como um recurso (arquivo de dados). Desta forma, caso o portal não permita disponibilizar estes metadados de forma integrada com a plataforma, ainda sim, será possível disponibilizar os metadados estruturais. Esta variável verifica a quantidade de metadados estruturais disponibilizados em arquivos(recursos).	Documentação de apoio aos dados	
--	---	---------------------------------	--

Fonte: O Autor

As variáveis apresentadas no Quadro 31 possibilitam descrever a estrutura dos recursos, como por exemplo, o nome campo, o título, a descrição e o tipo de dado. Este indicador também é um complemento do indicador metadados – seção B.0.1.3.

Percebe-se a importância deste indicador quando utilizado por exemplo, para a identificação da atualização entre versões de conjuntos de dados, uma vez que, os *stakeholders* possam comparar as estruturas e perceber quais foram as mudanças realizadas, dentre outros benefícios.

#### B.0.1.6 Indicador Versionamento e suas variáveis

Este indicador tem como objetivo verificar se os conjuntos de dados apresentam algum controlador de versão ou mesmo o uso de datas em seu URI para diferenciá-los, bem como verificar a disponibilidade do histórico de versões. Desta forma, permite-se aos *stakeholders* contemplar as diferenças entre os conjuntos de dados e não prejudica o acesso e o uso aos conjuntos de dados previamente criados.

A literatura não apresenta um consenso sobre qual momento deve ser efetuada a atualização de versão, seja através da alteração de um conjunto de dados ou a criação de um novo. O que é recomendado pela literatura, é a enumeração dos conjuntos de dados ou mesmo o uso de datas na URI, caso não tenha como atribuir uma versão. O Quadro 32 descreve as variáveis que irão compor este indicador.

Quadro 32 – Descrição das variáveis, indicador versionamento

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Conjuntos de dados com controle de versão ou com data no URI	Atribuição com controle de versão normalmente é uma solução fornecida pelas plataformas. Por sua vez, o uso de datas nos URIs é uma atribuição menos eficaz que o controle de versão, contudo, caso não seja possível utilizar controle de versão, o emprego de datas no URI é válido para identificar as diferenças entre os conjuntos de dados. Esta variável verifica a quantidade de conjuntos de dados que possuem controle de versão ou data atrelada a URI.	—	Fornecer o indicador de versão (BP 7).
Histórico de versão disponível	A disponibilidade de histórico de versão auxilia a compreensão e diferenciação entre as versões dos conjuntos de dados ao longo do tempo. Esta variável verifica se está disponível o histórico de versão para os conjuntos de dados, para isso utiliza-se conceito de variável binária.	—	Fornecer o histórico de versão (BP 8).

Fonte: O Autor

Percebe-se que a atribuição de versões aos conjuntos de dados é importante, sendo assim, é extremamente recomendada a atribuição do versionamento a todos os conjuntos de dados.

### B.0.1.7 Indicador URI e suas variáveis

Este indicador verifica o acesso e a disponibilidade de URI pertencente aos conjuntos de dados e recursos. Um dos meios de se garantir o acesso é através da URI, dado que é a maneira na qual é garantido o acesso e a descoberta.

O uso de URI deve ser exclusivo para cada conjunto de dados e recursos, e que estejam sempre acessíveis. Não existem problemas em utilizar recursos hospedados em outros servidores, contudo, é de extrema importância que as URIs estejam sempre acessíveis (LÓSCIO, B. F. *et al.*, 2017; MÁCHOVÁ; LNÉNICKA, 2017; ZHU, X.; FREEMAN, 2019; ABELLA *et al.*, 2019) e que seja utilizado *HTTP Status Code* recomendados pela W3C conforme cada situação específica <sup>1</sup>. O Quadro 33 descreve as variáveis que irão compor este indicador.

Quadro 33 – Descrição das variáveis, indicador URI

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
URIs encontradas ou redirecionadas (Conjuntos de Dados)	Esta variável é utilizada para verificar se a URI foi encontrada com sucesso ou se houve algum redirecionamento. Isso garante que as requisições à URI estão funcionais, mesmo quando houver algum redirecionamento. Alguns redirecionamentos são temporários, sendo normalizados posteriormente, contudo, todo redirecionamento deve ser "Encontrado". Esta variável verifica a quantidade de URI acessíveis ou redirecionadas com resposta "Encontrada" em relação aos conjunto de dados e recursos.	URI permanente	Usar URI persistentes como identificadores de conjuntos de dados (BP 9); Usar URI persistentes como identificadores dentro de conjuntos de dados (BP 10); Atribuir URI para as versões dos conjuntos de dados e séries (BP 11).
URI encontradas ou redirecionadas (Recursos)			

<sup>1</sup> <https://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html>

Fonte: O Autor

Percebe-se que a importância da URI e que sem o seu funcionamento adequado impossibilita o acesso e a exploração dos dados. Portanto, entende-se que todos os conjuntos de dados e recursos devem estar acessíveis, contudo, compreende-se que ao longo do tempo, podem ocorrer problemas com o acesso, sendo importante utilizar do *HTTP Status Code* para sinalizar o que ocorre com as URIs durante cada solicitação.

## **B.0.2 Dimensão Formatos**

Esta dimensão tem como objetivo verificar os formatos de arquivos que são disponibilizados no portal de dados, sendo estes compreensíveis por humanos e máquinas. Os formatos disponibilizados são importantes para garantir que serão utilizados pelos *stakeholders*, porque não adianta garantir o acesso aos dados se estes não possibilitarem o seu uso e reuso

### **B.0.2.1 Indicador Formatos de dados e suas variáveis**

Este indicador possibilita verificar os diferentes recursos (arquivos de dados) disponibilizados com os conjuntos de dados e preferencialmente formatos abertos: PDF, TXT, ODT, ODS, etc e também dos formatos abertos legíveis por máquinas: CSV, XML, HDF5, JSON, RDF/XML, JSON-Linked Data, dentre outros. A disponibilização de diferentes formatos de dados contribui com a redução de custos (tempo e processamento) e reduz erros que podem ocorrer em um processo de transformação (LÓSCIO, B. F. *et al.*, 2017). Além disso, possibilita diferentes visualizações ou mesmo a uma melhor compreensão dos dados publicados.

A literatura apresenta alguns formatos de arquivos, sendo que não existe um consenso sobre a quantidade de formatos que um conjunto de dados pode possuir. Recomenda-se utilizar a maior variedade possível e que seja preferencialmente disponibilizado em formatos abertos e preferivelmente formato aberto legíveis por máquinas, de modo a alcançar o maior número de *stakeholders* (LÓSCIO, B. F. *et al.*, 2017; CORRÊA, Andreiuid Sh *et al.*, 2017; MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018; ZHU, X.; FREEMAN, 2019). O Quadro 34 descreve as variáveis que irão compor este indicador.

---

Quadro 34 – Descrição das variáveis, indicador formato de dados

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Cobertura de formatos de dados	Esta variável permite identificar a variedade de formatos não repetidos por conjuntos de dados. Para isso, será realizada a soma total de formatos não repetidos por conjuntos de dados e este valor divide-se pelo total de conjuntos de dados. Para a normalização da variável o valor obtido previamente será dividido pelo maior valor de formatos não repetidos identificados nos conjuntos de dados.	—	Usar formatos de dados padronizados legíveis por máquinas (BP 12); Fornecer dados em formatos múltiplos (BP 14)
Arquivos (recursos) - formato aberto	Esta variável verifica a quantidade de recursos em formato aberto (PDF, ODS, ODT e TXT). A disponibilização de formatos abertos oferece facilidade aos <i>stakeholders</i> em utilizarem os dados disponibilizados.	Formato de dados	—
Arquivos (recursos) - formato aberto legível por máquina	Esta variável verifica a quantidade de recursos em formatos legíveis por máquinas, sendo estes: CSV, XML, HDF5, JSON, RDF/XML, JSON-LD e Turtle.	Usar formatos de dados padronizados legíveis por máquinas (BP 12);	

<p>Formatos com URI condizente ao formato cadastrado</p>	<p>Esta variável verifica a quantidade de recursos cadastrados que são condizentes com as URI disponibilizadas. Isso traz maior credibilidade ao portal e a veracidade sobre os dados que estão sendo publicados perante aos <i>stakeholders</i>.</p>	<p>—</p>	<p>—</p>
--	---	----------	----------

Fonte: O Autor

No que se refere a variável **Formatos com URI condizente ao formato cadastrado**, pode-se observar que não há conceitos ou boas práticas que fazem referência. Isso ocorre, pois esta variável foi identificada quando o modelo de referência estava em construção. Onde, percebeu-se que algumas URIs de formatos de dados eram diferentes dos formatos informados pelo portal, ou seja, o portal informava que era um arquivo CSV, mas quando acessado era um arquivo PDF, por exemplo.

### B.0.3 Dimensão Direitos

Esta dimensão tem como objetivo abordar questões relacionadas às licenças que remetem ao direito sobre os dados. Questões legais ou aspectos da LGPD não são abordados nesta dimensão, apenas questões relacionadas a licença de dados.

#### B.0.3.1 Indicador Licença e suas variáveis

Este indicador verifica se há licença empregada nos conjunto de dados. No âmbito de dados abertos a licença é de extrema importância, pois é desta maneira que determina-se o nível de restrições, o compartilhamento e reutilização dos dados disponibilizados. Outro ponto importante sobre o uso de licenças é que os dados publicados sem licença aberta não podem ser considerados dados abertos (MÁCHOVÁ; LNÉNICKA, 2017).

A literatura recomenda preferencialmente o uso da licença aberta sem restrições (CC0, PDDL) ou o emprego da licença aberta por atribuição (CC-BY, CC-BY-SA, ODC-BY ) que possibilita o livre uso, mas com algumas restrições, como por exemplo, informar quem é o real proprietário dos dados (LÓSCIO, B. F. *et al.*, 2017). Contudo, a escolha de qual licença utilizar é determinada conforme a estratégia e do grau de importância definida pela organização. Deste modo, quem for fazer uso dos conjuntos

de dados com licença aberta terá total liberdade para usar e até mesmo comercializar os produtos ou serviços oriundos destes dados.

As licenças são divididas em dois grupos, as licenças de conteúdo *CC* e as licenças para *Open Database License (ODbL)*. Ambos os grupos possuem diferentes derivações, como apresentado na seção 2.5. O Quadro 35 descreve as variáveis que irão compor este indicador.

Quadro 35 – Descrição das variáveis, indicador licenças

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Licença Aberta	Esta variável verifica a quantidade de licenças CC-ZERO, ODC-PDDL ou “outra licença aberta”. Estas licenças se remetem ao uso irrestrito dos dados.	Licenças	Fornecer informações sobre a licença de dados (BP 4)
Licença Aberta com Restrição	Esta variável verifica a quantidade de licenças CC-BY, CC-BY-SA, CC-BY-ND, ODC-ODBL, ODC-BY, outra-AT, UK-ODL e GFDL. Estas licenças se enquadram no livre uso desde que seja informado o proprietário dos dados, por exemplo.		
Licença Fechada	Esta variável verifica a quantidade de licenças CC-NC, CC-BY-NC, CC-BY-NC-SA, CC-BY-NC-ND, outra-fechada e outra-NC. Estas licenças se enquadram os dados que não podem ser utilizados, principalmente com o foco comercial.		

Licença Específica	Esta variável verifica a quantidade de licenças que foram criadas por governos a fim de atender requisitos específicos de acesso e uso dos dados.		
--------------------	---	--	--

Fonte: O Autor

#### **B.0.4 Dimensão Semântica**

Esta dimensão tem como objetivo evidenciar o uso e os tipos de vocabulários vinculados aos conjuntos de dados, conforme as recomendações da literatura. Outra questão que esta dimensão aborda é a proveniência dos dados, de modo a evidenciar as fontes oficiais a quais os conjuntos de dados pertencem, por exemplo. A semântica contribui para que máquinas possam compreender o significado dos dados e ao mesmo tempo apoiar a Web de dados.

##### **B.0.4.1 Indicador Proveniência e suas variáveis**

Este indicador verifica a existência de informações que identifiquem a autenticidade das publicações dos conjuntos de dados por um órgão ou entidade oficial. Esta informação é dada normalmente através de um endereço Web oficial principalmente em caso de publicações de terceiros (SISTO *et al.*, 2018).

Esta abordagem é comum em casos de organizações terceiras não possuírem a infraestrutura para efetuar as publicações, por exemplo, universidades que publicam seus dados diretamente no portal do governo federal. Outro aspecto é o vínculo dos conjuntos de dados a uma organização isso demonstra a seriedade e confiabilidade dos dados publicados. Esta prática é recomendada pela literatura, caso as entidades parceiras não possuam a estrutura necessária para a publicação dos dados de forma autônoma (SISTO *et al.*, 2018).

No que lhe concerne, os portais de dados que possuem conjuntos de dados de órgãos/entidades parceiras devem explicitar a quem pertence, bem como o site destes órgãos ou entidades. O quadro 36 apresenta as variáveis que irão compor o indicador.

Quadro 36 – Descrição das variáveis, indicador proveniência

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Conjuntos de dados que estão sem vínculo a alguma organização	Conjuntos de dados sem vínculo à uma organização não são confiáveis e podem descreditar o portal de dados. Portanto, esta variável verifica a quantidade de conjuntos de dados que foram publicados com vínculo à uma organização.	Site parceiros	Fornecer informações de procedência dos dados (BP5)
Conjuntos de dados que possuem informações da fonte oficial dos dados	Conjuntos de dados que possuem vínculo com uma organização e transparece um endereço Web da organização ou mesmo da origem dos dados, proporcionam confiança e segurança. Esta variável verifica a quantidade de conjuntos de dados que possuem um link para a fonte oficial que originou a postagem.	Fonte Oficial	

Fonte: O Autor

#### B.0.4.2 Indicador Vocabulários e suas variáveis

Este indicador verifica a existência de vocabulários que dão suporte a compreensão do significado dos dados e a flexibilização no uso dos conjuntos de dados por agentes computacionais.

A literatura recomenda fortemente o uso dos vocabulários padrões para que além de melhorar a compreensão dos conjuntos de dados, contribua-se com comunicação entre pessoas de diferentes domínios, a formalização na representação do conhecimento eliminando as contradições e inconsistências, a representação de conhecimento e reutilização.

Os vocabulários padrões que podem ser utilizados: *Dublin Core (DC)*, *Simple*

*Knowledge Organization System (SKOS)*, *Vcard*, *FOAF*, dentre outros<sup>2</sup>, sendo importante salientar que cada vocabulário possui um contexto de aplicação e em determinados casos é possível a criação de um vocabulário próprio ou mesmo o aprimoramento de um existente para atender o contexto específico (LÓSCIO, B. F. *et al.*, 2017). O Quadro 37 descreve as variáveis que irão compor este indicador.

Quadro 37 – Descrição das variáveis, indicador vocabulário

Variáveis	Definição	Conceitos da literatura	Boas Práticas W3C
Disponibiliza informações sobre vocabulários próprios	Esta variável verifica se algum conjunto de dados faz uso de vocabulário que foi criado ou aperfeiçoado pela própria organização. Para isso, utiliza-se conceito de variável binária.	Vocabulário	Reutilizar vocabulários, dando preferência aos padronizados (BP 15)
Vocabulários nos conjuntos de dados	Esta variável verifica quantos conjuntos de dados utilizam de vocabulários.		
Vocabulários Licenças	Esta variável verifica quantos conjuntos de dados fazem uso de vocabulários que se remetem às licenças. São verificadas a existência das seguintes classes: <code>dct:license</code> , <code>cc:license</code> , <code>schema:license</code> e <code>xhtml:license</code> .	—	Fornecer informações sobre a licença de dados (BP 4)
Vocabulários Proveniência	Esta variável verifica quantos conjuntos de dados fazem uso de vocabulários que se remetem a proveniência. É verificada a existência do seguinte vocabulário PROV <sup>3</sup> .	—	Fornecer informações de procedência dos dados (BP5)

<sup>2</sup> Outros vocabulários <<https://lov.linkeddata.es/dataset/lov/>>

<sup>3</sup> <https://www.w3.org/TR/prov-dm/>

Vocabulários Qualidade	Esta variável verifica quantos conjuntos de dados fazem uso de vocabulários que se remetem a qualidade. É verificado a existência do seguinte vocabulário DQV <sup>4</sup> .	—	Fornecer informações de qualidade de dados (BP 6)
Vocabulários Versionamento	Esta variável verifica quantos conjuntos de dados fazem uso de vocabulários que se remetem aos versionamento. São verificadas a existência das seguintes classes (owl:versionInfo, pav:version).	—	Fornecer indicador de versão (BP 7)
Vocabulários Histórico de Versão	Esta variável verifica quantos conjuntos de dados fazem uso de vocabulários que se remetem ao histórico de versão. São verificadas a existência das seguintes classes: (dct:isVersionOf, pav:previousVersion).	—	Fornecer o histórico de versão (BP 8)
Vocabulários Atualização	Esta variável verifica quantos conjuntos de dados fazem uso de vocabulários que se remetem as atualizações. É verificado a existência da classe: (dct:accrualPeriodicity).	—	Fornecer dados atualizados (BP 21)
Vocabulários Feedback Usuário	Esta variável verifica quantos conjuntos de dados fazem o uso de vocabulários que se remetem ao <i>feedback</i> de usuário. É verificado o seguinte vocabulário DUV <sup>5</sup> .	—	Compartilhar o <i>feedback</i> disponível (BP 30)

Fonte: O Autor

<sup>4</sup> <https://www.w3.org/TR/vocab-dqv/>

<sup>5</sup> <https://www.w3.org/TR/vocab-duv/>

### B.0.5 Dimensão Estrutura

Esta dimensão tem por objetivo verificar os aspectos que se remetem à estrutura do portal de dados, sendo este, o meio para a divulgação, catalogação, hospedagem, localização e interação com os dados.

#### B.0.5.1 Indicador Pesquisa e suas variáveis

Este indicador tem como objetivo verificar a existência de um campo de busca, de modo a realizar pesquisas no portal de dados. Outro aspecto que o indicador pesquisa verifica é o refinamento, onde permite aos usuários filtrarem os resultados, chegando mais próximo do desejado ou mesmo encontrando com maior precisão os conjuntos de dados, ou recursos. Em caso de não existir a possibilidade de pesquisa, torna-se difícil o acesso e a localização dos conjuntos de dados.

A literatura recomenda a existência de barra de busca e a opção de refinamento preferencialmente em um local evidente, de modo a facilitar a procura pelos conjuntos de dados ou recursos (CORRÊA, Andreiuid Sh *et al.*, 2017; MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018; ZHU, X.; FREEMAN, 2019). Portanto, quanto maior a variedade de filtros para o refinamento, melhor, como exemplo: filtro por data de publicação, por formato de arquivos, por categorias de licença, por entidades, por grupos, dentre outros. O Quadro 38 apresenta as variáveis que irão compor o indicador.

Quadro 38 – Descrição das variáveis, indicador pesquisa

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Barra de pesquisa disponível	Esta variável verifica se está disponível a barra de pesquisa, para isso utiliza-se conceito de variável binária.	Pesquisa	—
Refinamento disponível	Esta variável verifica se está disponível a opção de refinamento, para isso utiliza-se conceito de variável binária.	Filtros	—

Fonte: O Autor

### B.0.5.2 Indicador Conta Usuário e suas variáveis

Este indicador verifica a existência da opção de conta de usuário. Normalmente esta conta disponibilizada é considerada uma conta de visitante, possibilitando acompanhar as atividades que são realizadas no portal ou mesmo atividades específicas (CKAN, 2019). Esta opção de conta de usuário aumenta a interação e o interesse dos *stakeholders* sobre o portal de dados, por possibilitar notificações das atividades por e-mail, a possibilidade de acompanhar publicações ou organizações de interesse.

A literatura recomenda que o portal disponibilize o registro e o acesso de usuários, de modo a proporcionar uma melhor experiência ao utilizar o portal de dados (MÁ-CHOVÁ; LNÉNICKA, 2017). Contudo, o perfil de conta disponibilizada é de critério da organização. O Quadro 39 apresenta as variáveis que irão compor o indicador.

Quadro 39 – Descrição das variáveis, indicador conta usuário

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Disponibilidade de cadastro de usuário	Esta variável verifica se está disponível o cadastro de usuário, para isso utiliza-se conceito de variável binária.	Conta de usuários	—
Disponibilidade de acesso de usuário	Esta variável verifica se está disponível acesso de usuário ( <i>login</i> ), para isso utiliza-se conceito de variável binária.		

Fonte: O Autor

### B.0.5.3 Indicador Recomendação e suas variáveis

Este indicador verifica a existência de um sistema de recomendação de conjuntos de dados. Deste modo, o usuário ao pesquisar um conjunto de dados pode receber as recomendações sobre outros conjuntos de dados ou mesmo relacionadas com a palavra-chave de busca.

A literatura aponta que ao recomendar conjuntos de dados, melhora a experiência do cidadão, proporcionando um melhor aproveitamento dos conjuntos de dados, aumentando a possibilidade de gerar novos negócios, dentre outros benefícios (ZHU, X.; FREEMAN, 2019). O Quadro 40 apresenta as variáveis que irão compor este indicador.

Quadro 40 – Descrição das variáveis, indicador recomendação

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Sugestões na barra de pesquisa	Identifica se ao buscar por uma palavra-chave o portal lista outras semelhantes ao contexto de busca. Esta variável verifica se está disponível a sugestão de busca, para isso utiliza-se conceito de variável binária.	Recomendação de conjunto de dados	—
Conjunto de dados semelhantes	Identifica se ao navegar entre os conjuntos de dados o portal mostra outros resultados semelhantes ao que esta sendo acessado. Esta variável verifica se está disponível a recomendação de conjuntos de dados semelhantes, para isso utiliza-se conceito de variável binária.		

Fonte: O Autor

#### B.0.5.4 Indicador Acessibilidade e suas variáveis

Este indicador verifica se existem recursos de acessibilidade do portal baseado no WCAG<sup>6</sup>. Este indicador utiliza com base a nota gerada pelo software *Access Monitor*<sup>7</sup>, que verifica Web sites com base na WCAG 2.1. Desta forma é possível obter uma nota e verificar o quão adequado o portal de dados está em relação a aspectos de acessibilidade.

O quesito, acessibilidade é importante, pois possibilita a interação de pessoas com alguma deficiência, de modo que permita exercer o seu papel de cidadão. Para isso, a literatura recomenda utilizar das diretrizes de acessibilidade para o conteúdo da Web disponibilizada pela W3C no portal de dados abertos (CORRÊA, Andreiuid Sh *et al.*, 2017). O Quadro 41 apresenta as variáveis que irão compor este indicador.

<sup>6</sup> Documentação da WCAG <<https://www.w3.org/TR/WCAG21/>>

<sup>7</sup> Software Access Monitor <<http://accessmonitor.acessibilidade.gov.pt/>>

Quadro 41 – Descrição das variáveis, indicador acessibilidade

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Nota WCAG 2.1	Utiliza da nota gerada pelo sistema <i>access monitor</i> , baseado nas recomendações e métricas da WCAG 2.1.	Acessibilidade	—

Fonte: O Autor

Vale destacar que a nota obtida pelo WCAG pode variar de um a dez pontos, sendo quanto maior a pontuação melhor serão os aspectos de acessibilidade apresentados pelo portal.

#### B.0.5.5 Indicador Multi-Idioma e suas variáveis

Este indicador verifica a existência dos diferentes idiomas além do idioma nativo do portal. Isso possibilita uma maior difusão dos dados e do portal por vários *stakeholders* de diferentes países. Neste indicador está sendo verificada a disponibilidade de diferentes idiomas em relação ao portal, aspecto relacionado aos dados ou mesmo ao conteúdo dos recursos disponibilizados não são considerados.

A literatura aponta que ao disponibilizar o portal em mais de um idioma, possibilita a oportunidade de uso dos conjuntos de dados, o conhecimento da organização/entidade por qualquer cidadão em escala global (MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018; ZHU, X.; FREEMAN, 2019). O Quadro 42 apresenta as variáveis que irão compor este indicador.

Quadro 42 – Descrição das variáveis, indicador multi-idioma

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Disponibilidade em outros idiomas além do nativo	Identifica se o portal dispõe da opção de múltiplos idiomas.	Multi-Idioma	—

Fonte: O Autor

### B.0.5.6 Indicador Visualização dos dados e suas variáveis

Este indicador verifica a existência de mecanismos que possibilitem a visualização dos recursos pertencentes aos conjuntos de dados sem a necessidade de baixar o arquivo ou utilizar algum software específico, podendo visualizar os dados pelo próprio portal (*on-line*).

A literatura aponta que a possibilidade de visualizar os recursos, melhora a experiência do cidadão, proporciona um maior envolvimento com as atividades políticas e sociais, melhora o aspecto de transparência e responsabilização (MÁCHOVÁ; LNÉ-NICKA, 2017; SISTO *et al.*, 2018; ZHU, X.; FREEMAN, 2019). A Quadro 43 apresenta as variáveis que irão compor indicador.

Quadro 43 – Descrição das variáveis, indicador visualização de dados

<b>Variável</b>	<b>Descrição</b>	<b>Conceitos da literatura</b>	<b>Boas Práticas W3C</b>
Visualizador de conteúdo HTML	Esta variável verifica se está disponível visualização de conteúdo HTML, para isso utiliza-se conceito de variável binária.	Visualização dos conjuntos de dados e/ou recursos interativos	Fornecer visualizações complementares (BP 32)
Visualizador de imagens	Esta variável verifica se está disponível visualização de imagens, para isso utiliza-se conceito de variável binária.		
Visualizador de texto	Esta variável verifica se está disponível visualização de texto, para isso utiliza-se conceito de variável binária.		
Visualizador com uso de mapas	Esta variável verifica se está disponível visualização de conteúdo georreferenciados, para isso utiliza-se conceito de variável binária.		
Visualizador de PDF	Esta variável verifica se está disponível visualização de documentos em PDF, para isso utiliza-se conceito de variável binária.		

Visualizador de gráficos	Esta variável verifica se está disponível visualização de gráficos, para isso utiliza-se conceito de variável binária.		
Visualizador de arquivos JSON	Esta variável verifica se está disponível visualização de arquivo JSON, para isso utiliza-se conceito de variável binária.		

Fonte: O Autor

#### B.0.5.7 Indicador Temática e suas variáveis

Este indicador verifica a existência de categorias temáticas para melhora organizar os conjuntos de dados. O uso de temas ou grupos traz maior organização dos conjuntos de dados, facilita a localização e permite um panorama de áreas que o portal abrange.

Deve-se utilizar de temas nos conjuntos de dados. Sendo que os conjuntos de dados devem estar vinculado a um ou mais grupos. Salienta-se a necessidade de fazer uso de temas de natureza transversal (comércio, saúde, meio-ambiente, educação, etc...) e deve-se cuidar para não confundir palavras-chaves com temas, pois possuem diferentes objetivos (MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018; ZHU, X.; FREEMAN, 2019). O Quadro 44 apresenta as variáveis que irão compor indicador.

Quadro 44 – Descrição das variáveis, indicador temática

Variável	Descrição	Conceitos da literatura	Boas Práticas W3C
Conjunto de dados com pelo menos um grupo vinculado	Esta variável verifica a quantidade de conjuntos de dados que possuem ao menos um grupo vinculado.	Categoria de conjunto de dados	Fornecer metadados descritivos (BP 2)

Fonte: O Autor

#### B.0.5.8 Indicador API|SPARQL e suas variáveis

Este indicador verifica se o portal disponibiliza o acesso por via API ou SPARQL. A disponibilização de API é importante, pois simplifica o acesso aos dados e informa-

ções através de requisições em HTTP. O SPARQL também possui sua importância, por ser uma linguagem de consulta de grafos em RDF e por ser uma das tecnologias base para emprego da Web semântica.

A literatura recomenda a disponibilização do acesso via API, por oferecer maior flexibilidade e processabilidade, de modo que os *stakeholders* trabalhem diretamente com os dados em um alto nível de detalhamento (LÓSCIO, B. F. *et al.*, 2017; ZHU, X.; FREEMAN, 2019; MÁCHOVÁ; LNÉNICKA, 2017). O uso do SPARQL também é recomendado visto que desta forma possibilita a descoberta e a exploração de dados na Web semântica (LÓSCIO, B. F. *et al.*, 2017; ABELLA *et al.*, 2019), ambos não precisam ser excludentes. O Quadro 45 apresenta as variáveis que irão compor o indicador.

Quadro 45 – Descrição das variáveis, indicador API | SPARQL

Variável	Descrição	Conceitos da literatura	Boas Práticas W3C
Disponibilidade API	Esta variável verifica se está disponível consultas via API, para isso utiliza-se conceito de variável binária.	Disponibilização de API	Disponibilizar dados por meio de uma API (BP 23)
Disponibilidade SPARQL	Esta variável verifica a disponibilidade de consultas via SPARQL para isso utiliza-se conceito de variável binária.	Consultas SPARQL	—
Informações sobre API	Esta variável verifica se está disponível a documentação da API para isso utiliza-se conceito de variável binária.	—	Fornecer documentação completa para as APIs (BP 25)

Fonte: O Autor

#### B.0.5.9 Indicador Interatividade e suas variáveis

Este indicador verifica aspectos que possibilitam a interação entre *stakeholders*, o portal e seus responsáveis. Para isso, é verificada a existência de formulários para solicitar aos órgãos conjuntos de dados que não estejam publicados ou acessíveis, um espaço para fórum, comentários ou mesmo a possibilidade de efetuar avaliações sobre o portal, ou os conjuntos de dados e seus recursos.

Outro ponto verificado é o compartilhamento através das mídias sociais isto possibilita aos *stakeholders* compartilhar os dados publicados para além de seu ecos-

sistema, amplificando assim suas ideias e refletindo também em quesitos de responsabilização e transparência do órgão a qual o portal pertence.

É verificada a possibilidade de utilizar citações em relação aos conjuntos de dados isso possibilita a pesquisadores, jornalistas, dentre outros referenciar os dados utilizados e por consequência aumentando as chances de descoberta do portal e de seus dados.

Toda interação proporcionada pelo portal pode potencializar o engajamento da sociedade em relação aos órgãos e ao uso dos dados. É compreensível que as organizações não saibam o que os *stakeholders* necessitam em relação aos dados, mas é necessário que se mantenha um canal de comunicação e sempre que possível atender às solicitações e promover a interação entre os *stakeholders* e os responsáveis pelo portal.

Com relação à solicitação de dados, a literatura aponta que a existência de um espaço onde o cidadão possa entrar em contato via formulário com a organização solicitando dados, traz confiabilidade e proporciona o vislumbre para a sociedade referente à competência/responsabilidade. Portanto, um canal de comunicação, que possibilite o cidadão entrar em contato solicitando um conjunto de dados, é relevante. Outro ponto que a literatura salienta é a possibilidade de disponibilizar as solicitações e as respostas ao público desta forma além de mostrar o envolvimento e dedicação com a sociedade, proporciona ao supervisor do portal uma visão geral sobre as solicitações e atendimentos (MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018).

Práticas que possibilitam a discussão dos *stakeholders* ou mesmo uma breve opinião sobre os dados (comentários e fóruns) são recomendadas. Desta forma, melhoram a interatividade, instigam a troca de ideias e potencializam a geração de ideais, possíveis negócios ou mesmo o *network* entre os *stakeholders*. As avaliações dos conjuntos de dados proporcionam uma melhor percepção dos publicadores sobre o que está sendo disponibilizado e se estes estão de acordo com a expectativa dos *stakeholders*. Portanto, quanto mais recursos o portal possuir no âmbito de interação com o cidadão, mais notável será a interação da sociedade com o portal, sobre os fatos político-sociais, dentre outros (ZHU, X.; FREEMAN, 2019; MÁCHOVÁ; LNÉNICKA, 2017).

No que lhe concerne, as mídias sociais são recomendadas pela literatura por possibilitar o compartilhamento dos conjuntos de dados, sendo um meio de exercitar o trabalho de cidadania, cobrança sobre a responsabilidade e transparência, dentre outros (ZHU, X.; FREEMAN, 2019; MÁCHOVÁ; LNÉNICKA, 2017).

Sobre os formatos de citação, a literatura recomenda meios de facilitar a citação dos conjuntos de dados baseados nas normativas, como exemplo, *American Psychological Association (APA)*, Associação Brasileira de Normas Técnicas (ABNT), *Modern Language Association (MLA)*, *Institute of Electrical and Electronics Engineers (IEEE)*,

entre outros (ZHU, X.; FREEMAN, 2019). Desta forma, se proporciona a localização sobre os dados apresentados por pesquisas, jornais, de modo a evitar a distorção e promover a transparência do que esta sendo exposto. O Quadro 46 apresenta as variáveis que compõem este indicador.

Quadro 46 – Descrição das variáveis, indicador Interatividade

Variável	Descrição	Conceitos da literatura	Boas Práticas W3C
Formulários ou informações sobre a solicitação de conjuntos de dados	Esta variável verifica se está disponível opção para a solicitação de conjuntos de dados, para isso utiliza-se conceito de variável binária.	Fórum, comentários e avaliações.	—
Disponibilidade de fórum	Esta variável verifica se está disponível a opção de fórum, para isso utiliza-se conceito de variável binária.		Coletar <i>feedback</i> de consumidores de dados (BP 29); Compartilhar o <i>feedback</i> disponível (BP 30);
Disponibilidade de comentário	Esta variável verifica se está disponível a opção para comentários nos conjuntos de dados ou recursos, para isso utiliza-se conceito de variável binária.		—
Disponibilidade de avaliação	Esta variável verifica se está disponível a opção de avaliação dos conjuntos de dados ou recursos, para isso utiliza-se conceito de variável binária.		—
Compartilhamento no Facebook	Esta variável verifica se está disponível a opção de compartilhamento nas redes sociais, para isso utiliza-se conceito de variável binária.	Mídias sociais	—
Compartilhamento no Google			
Compartilhamento no Twitter			

Citações disponíveis	Esta variável verifica se está disponível a opção de criar citações sobre os conjuntos de dados, utilizando de normas, como exemplo, MLA, APA, ABNT, entre outras. Para isso utiliza-se conceito de variável binária.	Disponibilizar formato de citação	—
----------------------	---	-----------------------------------	---

Fonte: O Autor

### B.0.6 Dimensão Apresentação

Esta dimensão verifica aspectos que se remetem à divulgação de informações sobre o portal, considerando conceitos que se remetam a organização do portal, identificação dos responsáveis, resumos quantitativos sobre a rotina do portal, divulgação de projetos oriundos do uso dos dados, dentre outros.

#### B.0.6.1 Indicador Compreensão e suas variáveis

Este indicador verifica a existência de uma ou mais páginas Web que tragam informações sobre o portal, como exemplo: o seu objetivo, periodicidade de atualização, responsáveis, a disponibilização das perguntas frequentes ou mesmo a divulgação de projetos oriundos dos dados publicados.

Estas informações são importantes, pois facilitam a compreensão e o entendimento dos objetivos e a importância do portal para o seu ecossistema ou mesmo para a sociedade (ZHU, X.; FREEMAN, 2019; MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018). Desta forma, as informações disponibilizadas devem ser de fácil acesso, claras, objetivas e compreensíveis também é importante explicitar um canal de comunicação para qualquer eventual dúvida (SISTO *et al.*, 2018; ABELLA *et al.*, 2019).

O uso de perguntas frequentes é uma estratégia para auxiliar de modo rápido e eficaz o esclarecimento de dúvidas rotineiras que em muitas vezes são simples (ZHU, X.; FREEMAN, 2019; MÁCHOVÁ; LNÉNICKA, 2017). Deste modo, contribui-se com o autoatendimento e compreensão dos *stakeholders* sobre aspectos do portal e uso dos dados, refletindo também em um alívio nos canais de comunicação.

A literatura recomenda um espaço (página Web) para a divulgação dos eventos, projetos, negócios originados ou promovidos que fizeram uso dos dados abertos publicados pelo portal de dados. Isso demonstra que os dados publicados são normalmente adequados ou mesmo possuem algum valor para o uso (SISTO *et al.*, 2018). O Quadro 47 apresenta as variáveis que irão operacionalizar este indicador.

Quadro 47 – Descrição das variáveis, indicador compreensão

Variável	Descrição	Conceitos da literatura	Boas Práticas W3C
Introdução sobre o portal de dados	Esta variável verifica se está disponível um espaço que faça uma apresentação do portal, trazendo informações que auxiliem na compreensão dos <i>stakeholders</i> , para isso utiliza-se conceito de variável binária.	Apresentação do portal	—
Perguntas frequentes (FAQ)	Esta variável verifica se está disponível um espaço que apresente as dúvidas mais frequentes sobre o portal, para isso utiliza-se conceito de variável binária.	Perguntas Frequentes (FAQ)	—
Disponibilização de projetos	Esta variável verifica se está disponível um espaço que apresente os projetos (Aplicativos, <i>Hacktoons</i> e outros) que utilizaram dos dados ou mesmo resultaram em produtos, para isso utiliza-se conceito de variável binária.	Espaço para divulgação	—

Fonte: O Autor

#### B.0.6.2 Indicador Quantitativos e suas variáveis

Verifica se existem informações gerais e quantitativas sobre os conjuntos de dados e sobre aspectos do portal de dados abertos, como por exemplo, número de visitantes, número de downloads, entre outros. Este resumo dos dados quantitativo permitem aos *stakeholders* perceberem se há uma quantidade razoável de conjuntos de dados ou mesmo uma maneira simples de perceber o crescimento do portal.

A literatura recomenda a apresentação de números gerais sobre alguns pontos específicos, de modo que possibilite um resumo quantitativo sobre aspectos do portal, onde os indivíduos que fazem uso podem ter um panorama sobre o portal de

dados (ZHU, X.; FREEMAN, 2019; MÁCHOVÁ; LNÉNICKA, 2017; SISTO *et al.*, 2018; ODM, 2019). Além disso, não há uma definição sobre os mínimos exigidos para os quantitativos de cada variável assim optou-se pela verificação da existência ou não. Os quantitativos que se destacam nos trabalhos de Xiaohua Zhu e Freeman (2019), Máchová e Lnénicka (2017), Sisto *et al.* (2018) e ODM (2019) são:

- Número de downloads
- Total em MB de dados
- Total de conjunto de dados
- Total de editores
- Número de aplicativos criados
- Total de catálogos coletados
- Total de recursos
- Número de visitas

O Quadro 48 apresenta as variáveis que compõem este indicador.

Quadro 48 – Descrição das variáveis, indicador quantitativos

Variáveis	Descrição	Conceitos da literatura	Boas Práticas W3C
Quadro com o resumo das atividades do portal	Esta variável verifica se está disponível um quadro com informações gerais sobre as atividades do portal, para isso utiliza-se conceito de variável binária.	—	—
Disponibilizar Número de downloads	Esta variável verifica se está disponível a quantidade de downloads realizados, para isso utiliza-se conceito de variável binária.	Número de download	—
Disponibilizar Total de conjunto de dados	Esta variável verifica se está disponível a quantidade total de conjuntos de dados, para isso utiliza-se conceito de variável binária.	Disponibilidade dos conjuntos de dados	—
Disponibilizar Número de projetos desenvolvidos	Esta variável verifica se está disponível a quantidade de projetos que já foram criados com dados oriundos do portal, para isso utiliza-se conceito de variável binária.	Número de aplicativos criados	—

Disponibilizar Total de recursos	Esta variável verifica se está disponível a quantidade total de recursos cadastrados, para isso utiliza-se conceito de variável binária.	Total geral dos recursos	—
Disponibilizar Total em MB de dados	Identificar se o portal disponibiliza a quantidade total em Megabytes de dados disponibilizados no portal, para isso utiliza-se conceito de variável binária.	Número de distribuição	—
Disponibilizar Total de editores	Esta variável verifica se está disponível a quantidade total de editores, ou seja, quem publicam no portal (por exemplo, as organizações), para isso utiliza-se conceito de variável binária.	Número total de editores	—
Disponibilizar Total de catálogos coletados	Esta variável verifica se está disponível a quantidade de dados que são coletados de outros portais e divulgados, para isso utiliza-se conceito de variável binária.	Número total de catálogos coletados	—
Disponibilizar Número de visitantes	Esta variável verifica se está disponível a quantidade total de visitantes no portal, podendo ser quantificado as visitas por conjuntos de dados, para isso utiliza-se conceito de variável binária.	Número de visitas dos conjuntos de dados	—

Fonte: O Autor

No que se refere à variável **Quadro com o resumo das atividades do portal** pode-se observar que não há conceitos ou boas práticas que fazem referência. Isso ocorre, pois esta variável foi identificada quando o modelo de referência estava em construção observando alguns portais de dados que possuem um quadro informativo. Sendo assim, entendeu-se que ao verificar a existência de um quadro geral, este pode

conter outras informações além das outras variáveis que estão sendo verificadas e assim apresentado dados quantitativos que é o objetivo do indicador. Destaca-se ainda, que não há a necessidade da existência de todas as variáveis que compõe o indicador.

## APÊNDICE C – RESULTADO DETALHADO DOS PORTAIS ANALISADOS

Os resultados apresentados neste apêndice tem por objetivo complementar o entendimento da análise apresentada na seção 4.2. Todos os resultados exibidos sobre os portais analisados estão disponíveis na aplicação Web, vale ressaltar que a aplicação possui cerca de 25 portais coletados de diferentes níveis de governo e países.

### C.1 DIMENSÃO ACESSO

#### C.1.1 Indicador Atualização

Este indicador tem por objetivo verificar se os conjuntos de dados e os recursos estão atualizados, bem como, se houve a adição de novos. Desta forma, conjuntos de dados atualizados são mais atrativos e confiáveis, proporcionando o seu uso ou reuso. Além da atualização, a adição de novos conjuntos de dados/recursos mostram que o portal está ativo e publicando, além disso, novos conjuntos de dados podem ser considerados uma atualização. A Tabela 15 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Disponibilizar dados atualizados regularmente com pelo menos o mesmo nível de qualidade e quantidade.
- Manter os dados atualizados sempre que possível;

Tabela 15 – Variáveis Indicador Atualização

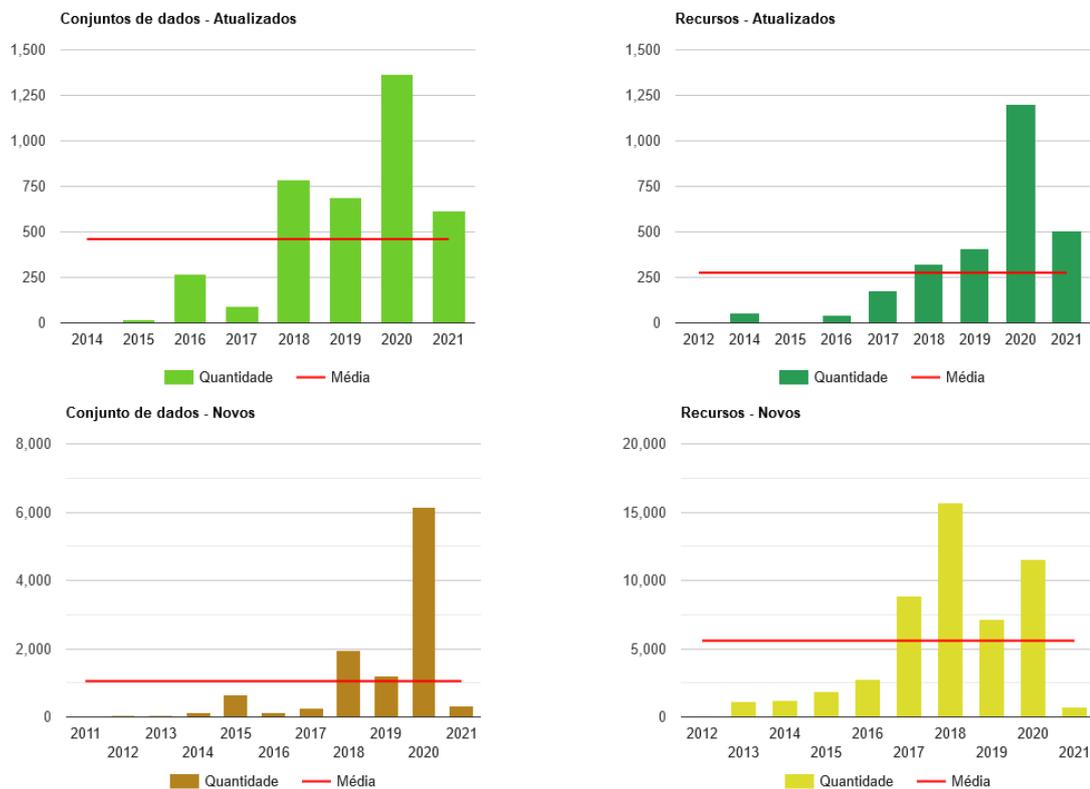
Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V01	Conjuntos de Dados - Atualização nos últimos 12 meses	1958	0.180	0.25	0.045
	V02	Recursos - Atualização nos últimos 12 meses	1673	0.033	0.25	0.008
	V03	Conjuntos de Dados - Adicionados nos últimos 12 meses	6421	0.592	0.25	0.148
	V04	Recursos - Adicionados nos últimos 12 meses	11686	0.229	0.25	0.057
	<b>Valor Indicador Atualização</b>					

Estado Rio Grande do Sul	V01	Conjuntos de Dados - Atualização nos últimos 12 meses	25	0.022	0.25	0.005
	V02	Recursos - Atualização nos últimos 12 meses	15	0.001	0.25	0.000
	V03	Conjuntos de Dados - Adicionados nos últimos 12 meses	2	0.002	0.25	0.000
	V04	Recursos - Adicionados nos últimos 12 meses	95	0.006	0.25	0.001
	<b>Valor Indicador Atualização</b>					
Município de São Paulo	V01	Conjuntos de Dados - Atualização nos últimos 12 meses	87	0.446	0.25	0.112
	V02	Recursos - Atualização nos últimos 12 meses	285	0.088	0.25	0.022
	V03	Conjuntos de Dados - Adicionados nos últimos 12 meses	7	0.036	0.25	0.009
	V04	Recursos - Adicionados nos últimos 12 meses	605	0.187	0.25	0.047
	<b>Valor Indicador Atualização</b>					

Fonte: do Autor

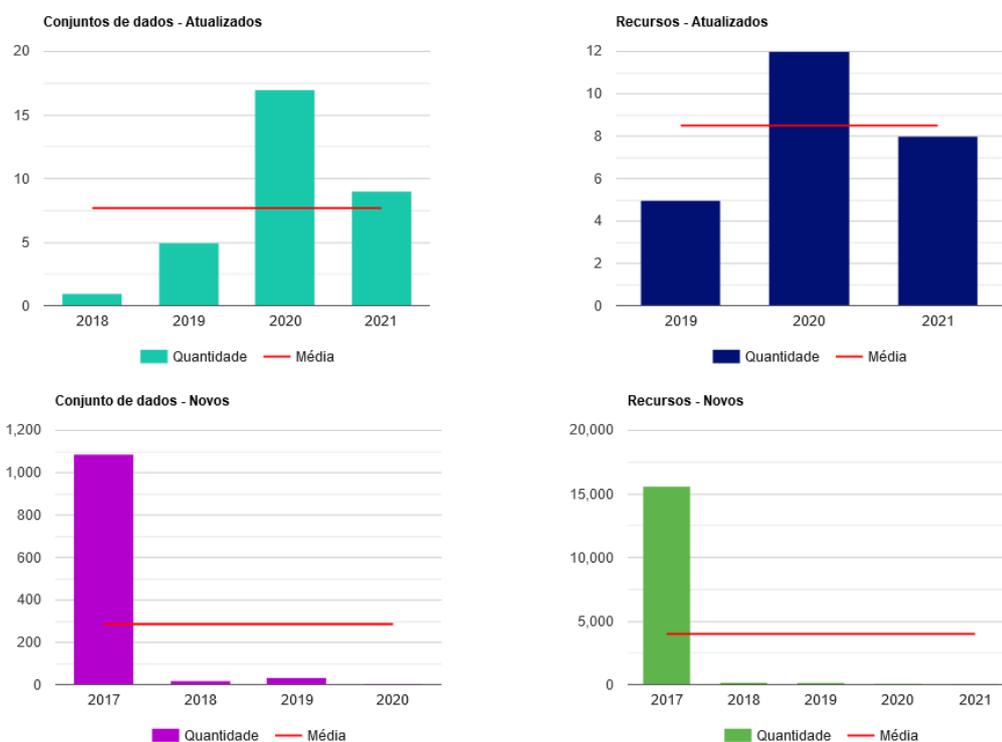
A seguir os Gráficos 26, 27 e 28 irão exibir a série histórica de atualizações e adições dos conjuntos de dados e recursos. Vale ressaltar que novos conjuntos de dados podem ser considerados atualizações, desde que atribuída a versão e destacado nos metadados e nas descrições.

Gráfico 26 – Série histórica dos conjuntos de dados e recursos - Portal Governo Federal



Fonte: O Autor

Gráfico 27 – Série histórica dos conjuntos de dados e recursos - Portal Estado do Rio Grande do Sul



Fonte: O Autor

Gráfico 28 – Série histórica dos conjuntos de dados e recursos - Portal Município de São Paulo



Fonte: O Autor

### C.1.2 Indicador Download em Massa

Este indicador tem por objetivo verificar a existência de recursos disponibilizados em formatos que possibilitem o download de grandes volumes de dados em uma única requisição ou mesmo em requisições segmentadas. Desta forma, traz maior confiança aos dados, dificultando uma manipulação inconsistente ao criar conjuntos de dados oriundos de uma coleta individual de dados, por exemplo. Proporcionando assim, uma melhor organização dos dados e aumentando a sua consistência. A Tabela 16 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Disponibilizar recursos em formato compactado, FTP ou API;
- Os formatos não são excludentes (compactado, FTP ou API);

Tabela 16 – Variáveis Indicador Download em Massa

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V05	Recursos disponíveis - API ou FTP	1081	0.021	0.5	0.011
	V06	Recursos compactados disponibilizados no portal	2867	0.056	0.5	0.028
	<b>Valor Indicador Download em Massa</b>					<b>0.039</b>
Estado Rio Grande do Sul	V05	Recursos disponíveis - API ou FTP	0	0	0.5	0.000
	V06	Recursos compactados disponibilizados no portal	160	0.01	0.5	0.005
	<b>Valor Indicador Download em Massa</b>					<b>0.005</b>
Município São Paulo	V05	Recursos disponíveis - API ou FTP	0	0	0.5	0.000
	V06	Recursos compactados disponibilizados no portal	689	0.213	0.5	0.107
	<b>Valor Indicador Download em Massa</b>					<b>0.107</b>

Fonte: do Autor

### C.1.3 Indicador Metadados

Este indicador verifica a existência de metadados em formato HTML e em formatos serializados. Os metadados são um meio necessário e fundamental para a

descoberta dos dados na Internet, por este motivo é tão importante que os conjuntos de dados e os recursos precisam possuir a maior quantidade de metadados possíveis. A Tabela 17 a seguir apresenta os dados coletados oriundos do portal selecionado.

### Recomendações

- Disponibilizar metadados para todos conjuntos de dados, bem como seus recursos;
- Disponibilizar metadados em diferentes formatos (não são excludentes);
- Disponibilizar metadados em formatos legíveis por máquinas;

Tabela 17 – Variáveis Indicador Metadados

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V07	Conjuntos de Dados - Metadados (Formato HTML)	10850	1	0.25	0.250
	V08	Recursos - Metadados (Formato HTML)	51100	1	0.25	0.250
	V09	Conjuntos de Dados - Metadados (Formato serializado)	10851	1	0.25	0.250
	V10	Recursos - Metadados (Formato serializado)	51100	1	0.25	0.250
	<b>Valor Indicador Metadados</b>					<b>1.000</b>
Estado Rio Grande do Sul	V07	Conjuntos de Dados - Metadados (Formato HTML)	1151	1	0.25	0.250
	V08	Recursos - Metadados (Formato HTML)	16039	1	0.25	0.250
	V09	Conjuntos de Dados - Metadados (Formato serializado)	1151	1	0.25	0.250
	V10	Recursos - Metadados (Formato serializado)	16039	1	0.25	0.250
	<b>Valor Indicador Metadados</b>					<b>1.000</b>
Município São Paulo	V07	Conjuntos de Dados - Metadados (Formato HTML)	193	0.99	0.25	0.247
	V08	Recursos - Metadados (Formato HTML)	3232	1	0.25	0.250
	V09	Conjuntos de Dados - Metadados (Formato serializado)	195	1	0.25	0.250

V10	Recursos - Metadados (Formato serializado)	3232	1	0.25	0.250
<b>Valor Indicador Metadados</b>					<b>0.997</b>

Fonte: do Autor

#### C.1.4 Indicador Metadados Descritivos

Este indicador verifica os metadados descritivos, referente aos conjuntos de dados e seus recursos. Com isso, é possível identificar a sua natureza, possibilitando a compreensão. Este indicador é um complemento do indicador anterior com isso é possível descrever os conjuntos de dados e recursos utilizando de chave-valor possibilitando um melhor refinamento para a descoberta dos dados. A literatura recomenda o uso de alguns atributos para explicitar as informações sobre os conjuntos de dados, mas também é possível adicionar outros para complementar e melhor descrever os conjuntos de dados ou recursos. A Tabela 18 a seguir apresenta os dados coletados oriundos do portal selecionado.

##### **Recomendações**

- \* Disponibilizar metadados descritivos:
  - O título e a descrição do conjunto de dados;
  - Palavras-chave que descrevem o conjunto de dados;
  - Data de publicação do conjunto de dados;
  - Entidade responsável (publicadora de dados) por disponibilizar o conjunto de dados;
  - Informações para contactar os responsáveis pelo conjunto de dados;
  - Temas/categorias cobertos por um conjunto de dados;
  - Formato de mídia da distribuição;
  - Outros que contribuem com a descrição do conjuntos de dados publicados;

Tabela 18 – Variáveis Indicador Metadados Descritivos

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V11	Conjuntos de Dados - Metadados (Nome)	10851	1	0.077	0.077
	V12	Recursos - Metadados (Nome)	50866	0.995	0.077	0.077
	V13	Conjuntos de Dados - Metadados (Descrição)	10850	1	0.077	0.077
	V14	Recursos - Metadados (Descrição)	50909	0.996	0.077	0.077
	V15	Conjuntos de Dados - Metadados (Data Criação)	10851	1	0.077	0.077
	V16	Recursos - Metadados (Data Criação)	50909	0.996	0.077	0.077
	V17	Metadados (Mantenedor)	9926	0.915	0.077	0.070
	V18	Metadados (Mantenedor e-mail)	6819	0.628	0.077	0.048
	V19	Metadados (Autor)	9975	0.919	0.077	0.071
	V20	Metadados (Autor e-mail)	6845	0.631	0.077	0.049
	V21	Metadados (Palavras-Chave)	10314	0.951	0.077	0.073
	V22	Metadados - Formato de arquivo	50909	0.996	0.077	0.077
	V23	Metadados - Tamanho de arquivo	4047	0.079	0.077	0.006
<b>Valor Indicador Metadados Descritivo</b>						<b>0.854</b>
	V11	Conjuntos de Dados - Metadados (Nome)	1151	1	0.077	0.077
	V12	Recursos - Metadados (Nome)	16022	0.999	0.077	0.077
	V13	Conjuntos de Dados - Metadados (Descrição)	1151	1	0.077	0.077
	V14	Recursos - Metadados (Descrição)	16022	0.999	0.077	0.077
	V15	Conjuntos de Dados - Metadados (Data Criação)	1143	0.993	0.077	0.076
	V16	Recursos - Metadados (Data Criação)	16022	0.999	0.077	0.077
	V17	Metadados (Mantenedor)	1143	0.993	0.077	0.076

Estado Rio Grande do Sul	V18	Metadados (Mantenedor e-mail)	1143	0.993	0.077	0.076
	V19	Metadados (Autor)	57	0.05	0.077	0.004
	V20	Metadados (Autor e-mail)	57	0.05	0.077	0.004
	V21	Metadados (Palavras-Chave)	1128	0.98	0.077	0.075
	V22	Metadados - Formato de arquivo	16022	0.999	0.077	0.077
	V23	Metadados - Tamanho de arquivo	0	0	0.077	0.000
<b>Valor Indicador Metadados Descritivo</b>						<b>0.773</b>
Município São Paulo	V11	Conjuntos de Dados - Metadados (Nome)	195	1	0.077	0.077
	V12	Recursos - Metadados (Nome)	2670	0.826	0.077	0.064
	V13	Conjuntos de Dados - Metadados (Descrição)	195	1	0.077	0.077
	V14	Recursos - Metadados (Descrição)	2670	0.826	0.077	0.064
	V15	Conjuntos de Dados - Metadados (Data Criação)	195	1	0.077	0.077
	V16	Recursos - Metadados (Data Criação)	2670	0.826	0.077	0.064
	V17	Metadados (Mantenedor)	195	1	0.077	0.077
	V18	Metadados (Mantenedor e-mail)	195	1	0.077	0.077
	V19	Metadados (Autor)	195	1	0.077	0.077
	V20	Metadados (Autor e-mail)	195	1	0.077	0.077
	V21	Metadados (Palavras-Chave)	187	0.959	0.077	0.074
	V22	Metadados - Formato de arquivo	2670	0.826	0.077	0.064
	V23	Metadados - Tamanho de arquivo	0	0	0.077	0.000
<b>Valor Indicador Metadados Descritivo</b>						<b>0.866</b>

Fonte: do Autor

A seguir são apresentadas as Figuras 26, 27 e 28 que correspondem as palavras-chaves mais usadas nos metadados nos portais analisados.





Governo Federal	V26	Metadados estrutural (Documentos separados)	877	0.017	0.333	0.006
	<b>Valor Indicador Metadados Estrutural</b>					<b>0.006</b>
Estado Rio Grande do Sul	V24	Metadados estrutural (Formato HTML)	0	0	0.333	0.000
	V25	Metadados estrutural (Formato serializado)	0	0	0.333	0.000
	V26	Metadados estrutural (Documentos separados)	0	0	0.333	0.000
	<b>Valor Indicador Metadados Estrutural</b>					<b>0.000</b>
Município São Paulo	V24	Metadados estrutural (Formato HTML)	0	0	0.333	0.000
	V25	Metadados estrutural (Formato serializado)	0	0	0.333	0.000
	V26	Metadados estrutural (Documentos separados)	87	0.027	0.333	0.009
	<b>Valor Indicador Metadados Estrutural</b>					<b>0.009</b>

Fonte: do Autor

### C.1.6 Indicador Versionamento

Este indicador tem como objetivo verificar se o conjunto de dados apresenta algum controlador de versão ou mesmo o uso de datas em seu URI para diferenciar os conjuntos de dados, bem como verificar a existência de histórico de versões. Desta forma, permite aos *stakeholders* identificarem as diferenças entre os conjuntos de dados, a repetibilidade na pesquisa e possibilita comparações entre versões. Naturalmente os conjuntos de dados podem mudar com o passar do tempo, sendo atualizações diárias ou quando a atualização seja possível. A Tabela 20 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Atribuir versão em todos os conjuntos de dados;
- Caso não tenha como atribuir versão aos conjuntos de dados, disponibilizar datas na URI;
- Disponibilizar um histórico de versões;

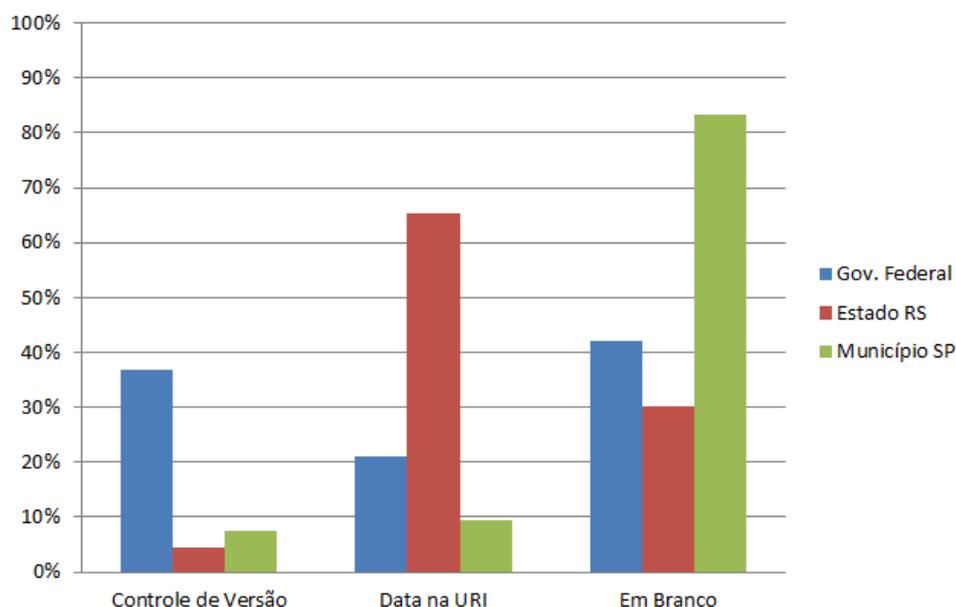
Tabela 20 – Variáveis Indicador Versionamento

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V27	Conjuntos de dados com controle de versão ou com data na URI	3892	0.359	0.5	0.179
	V28	Conjuntos de dados com histórico de versão	NÃO	0	0.5	0.000
	<b>Valor Indicador Versionamento</b>					<b>0.179</b>
Estado Rio Grande do Sul	V27	Conjuntos de dados com controle de versão ou com data na URI	46	0.04	0.5	0.020
	V28	Conjuntos de dados com histórico de versão	NÃO	0	0.5	0.000
	<b>Valor Indicador Versionamento</b>					<b>0.020</b>
Município São Paulo	V27	Conjuntos de dados com controle de versão ou com data na URI	34	0.174	0.5	0.087
	V28	Conjuntos de dados com histórico de versão	NÃO	0	0.5	0.000
	<b>Valor Indicador Versionamento</b>					<b>0.087</b>

Fonte: do Autor

O Gráfico 29 exibe algumas informações complementares, onde é possível comparar os tipos de versionamentos utilizados, bem como a quantidade de conjuntos de dados que não possuem nenhum tipo de versionamento atribuído.

Gráfico 29 – Informações complementares sobre o indicador versionamento



Fonte: O Autor

### C.1.7 Indicador URI

Este indicador verifica o acesso e a disponibilidade dos URIs pertencentes aos conjuntos de dados e recursos. Um dos meios de se garantir o acesso é através do URI, pois, é a maneira na qual é garantido o acesso e a descoberta dos conjuntos de dados e recursos. O uso de URI deve ser exclusivo de cada conjunto de dados e recursos, e que estejam sempre acessíveis. Não existem problemas em utilizar recursos hospedados em outros servidores, contudo, é de extrema importância que as URIs estejam sempre acessíveis. A Tabela 21 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Todas as URIs devem ser encontradas ou redirecionadas;
- Não pode ocorrer erros nas solicitações de URI;

Tabela 21 – Variáveis Indicador URI

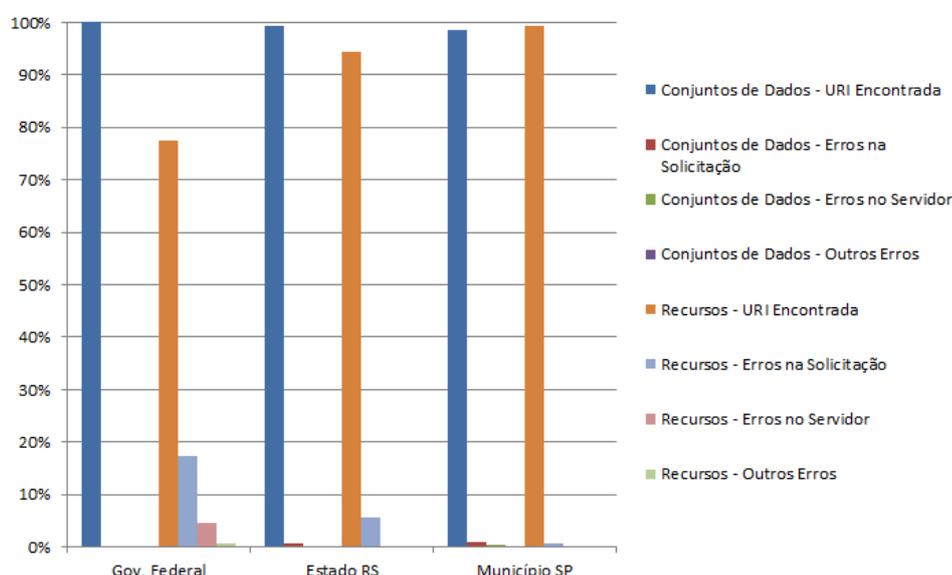
Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
	V29	Conjunto de Dados - URI encontrado ou redirecionado	10850	1	0.5	0.500

Governo Federal	V30	Recursos - URI encontrado ou redirecionado	39552	0.774	0.5	0.387
	<b>Valor Indicador URI</b>					<b>0.887</b>
Estado Rio Grande do Sul	V29	Conjunto de Dados - URI encontrado ou redirecionado	1143	0.993	0.5	0.497
	V30	Recursos - URI encontrado ou redirecionado	15124	0.943	0.5	0.471
	<b>Valor Indicador URI</b>					<b>0.968</b>
Município São Paulo	V29	Conjunto de Dados - URI encontrado ou redirecionado	192	0.985	0.5	0.492
	V30	Recursos - URI encontrado ou redirecionado	3211	0.994	0.5	0.497
	<b>Valor Indicador URI</b>					<b>0.989</b>

Fonte: do Autor

O Gráfico 30 exibe algumas informações complementares, onde é possível comparar os tipo de licenças utilizados nos portais.

Gráfico 30 – Informações complementares sobre o indicador URI - Gráficos HTTP Status Code



Fonte: O Autor

## C.2 DIMENSÃO FORMATOS

### C.2.1 Indicador Formatos de dados

Este indicador possibilita verificar os diferentes recursos (arquivos de dados) disponibilizados com os conjuntos de dados. As disponibilizações de diferentes formatos de dados contribuem com a redução de custo (tempo e processamento) e reduz os erros que podem ocorrer em um processo de transformação. Além disso, possibilita diferentes visualizações ou mesmo uma melhor compreensão dos dados publicados. Por sua vez, dados legíveis por máquinas são mais atrativos para os *stakeholders* técnicos por demandarem menos tempo de tratamento e processamento, bem como o acesso unitário aos dados e a facilidade de utilizar agentes computacionais. Dados legíveis por máquinas também são formatos de dados, sendo assim a literatura recomenda utilizar a maior variedade de dados legível por máquina e se possível em todos os conjuntos de dados. A Tabela 22 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Não pode conter formatos discrepantes com o URI;
- Deve possuir a maior variedade de formatos de dados (recursos) possível;
- Evitar a concentração de um único tipo de formato de dados;

- Disponibilizar a maior quantidade e variedade possível de arquivos de dados (recursos) em formato legível por máquinas;

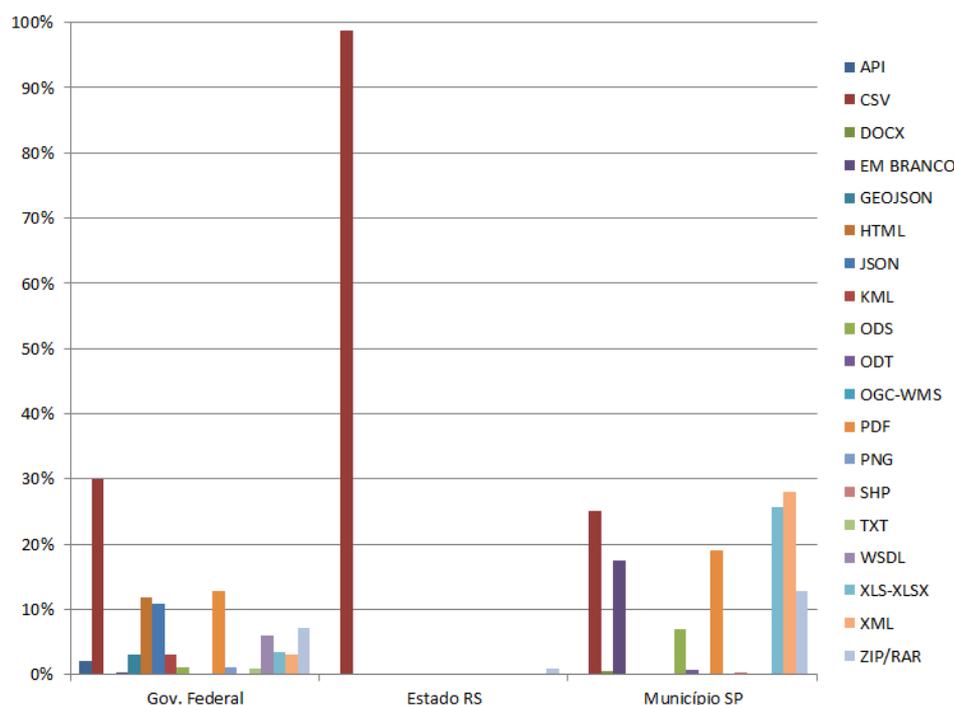
Tabela 22 – Variáveis Indicador Formatos de dados

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V31	Cobertura de formatos de dados	2.963	0.37	0.25	0.093
	V32	Arquivos (Recursos) - Formato aberto	7652	0.15	0.25	0.037
	V33	Arquivos (Recursos) - Formato aberto legível por máquina	22527	0.441	0.25	0.110
	V34	Formatos com URIs condizente ao recurso cadastrado	51100	1	0.25	0.250
	<b>Valor Indicador Formatos de Dados</b>					
Estado Rio Grande do Sul	V31	Cobertura de formatos de dados	1.014	0.338	0.25	0.084
	V32	Arquivos (Recursos) - Formato aberto	26	0.002	0.25	0.000
	V33	Arquivos (Recursos) - Formato aberto legível por máquina	15836	0.987	0.25	0.247
	V34	Formatos com URIs condizente ao recurso cadastrado	16039	1	0.25	0.250
	<b>Valor Indicador Formatos de Dados</b>					
Município São Paulo	V31	Cobertura de formatos de dados	2.441	0.271	0.25	0.068
	V32	Arquivos (Recursos) - Formato aberto	874	0.27	0.25	0.068
	V33	Arquivos (Recursos) - Formato aberto legível por máquina	838	0.259	0.25	0.065
	V34	Formatos com URIs condizente ao recurso cadastrado	3232	1	0.25	0.250
	<b>Valor Indicador Formatos de Dados</b>					

Fonte: do Autor

O Gráfico 31 exibe algumas informações complementares, onde é possível comparar os tipo de formatos utilizados entre os portais.

Gráfico 31 – Informações complementares sobre o indicador Formato de dados



Fonte: O Autor

## C.3 DIMENSÃO DIREITOS

### C.3.1 Indicador Licenças

Este indicador verifica a licença empregada ao conjunto de dados. No âmbito de dados abertos o quesito licença é de extrema importância, pois é desta maneira que determina o nível de restrições, o compartilhamento e reutilização dos dados disponibilizados. Outro ponto importante sobre o uso de licenças é que os dados publicados sem licença aberta não podem ser considerados dados abertos. A Tabela 23 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Utilizar licenças padrões e definidas por instituições de referência;
- Utilizar licenças abertas sem restrição;
- Em alguns casos pode utilizar licenças abertas com restrição;

- Se possível não utilizar licença do tipo fechada ou fora do padrão (Licenças específicas);
- Não deixar de atribuir licença nos conjuntos de dados;

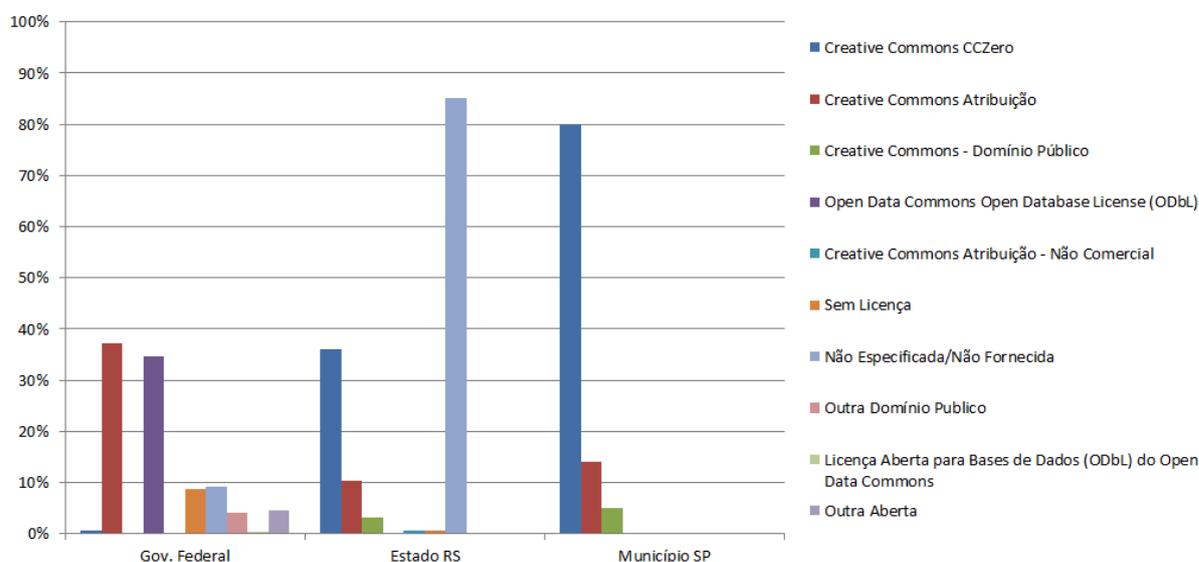
Tabela 23 – Variáveis Indicador Licenças

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V35	Licença Aberta	1085	0.1	0.25	0.025
	V36	Licença Aberta C/ Restrição	7862	0.725	0.25	0.181
	V37	Lic. Específica	0	0	0.25	0.000
	V38	Licença Fechada	10	0.001	0.25	0.000
	<b>Valor Indicador Licença</b>					
Estado Rio Grande do Sul	V35	Licença Aberta	36	0.031	0.25	0.008
	V36	Licença Aberta C/ Restrição	118	0.103	0.25	0.026
	V37	Lic. Específica	0	0	0.25	0.000
	V38	Licença Fechada	8	0.007	0.25	0.002
	<b>Valor Indicador Licença</b>					
Município São Paulo	V35	Licença Aberta	166	0.851	0.25	0.213
	V36	Licença Aberta C/ Restrição	29	0.149	0.25	0.037
	V37	Lic. Específica	0	0	0.25	0.000
	V38	Licença Fechada	0	0	0.25	0.000
	<b>Valor Indicador Licença</b>					

Fonte: do Autor

O Gráfico 32 exibe algumas informações complementares, onde é possível comparar os tipos de licenças utilizados.

Gráfico 32 – Informações complementares sobre o indicador Licença



Fonte: O Autor

## C.4 DIMENSÃO SEMÂNTICA

### C.4.1 Indicador Proveniência

Este indicador verifica a existência de informações que identifiquem a autenticidade da publicação do conjunto de dados por um órgão ou entidade oficial. Esta informação é dada normalmente através de um endereço Web oficial principalmente em caso de publicações de organizações de terceiros. Esta abordagem é comum em casos das organizações parceiras não possuírem a infraestrutura para efetuar as publicações, por exemplo, universidades que publicam seus dados diretamente no portal do governo federal. Outro aspecto é o vínculo dos conjuntos de dados a uma organização isso demonstra a seriedade e confiabilidade dos dados publicados. A Tabela 24 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Todo conjunto de dados deve possuir vínculo com uma organização;
- Disponibilizar endereço eletrônico (link) da fonte oficial que originou o conjunto de dados;

Tabela 24 – Variáveis Indicador Proveniência

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V39	Conjuntos de dados que possuem vínculo com uma organização	10718	0.988	0.5	0.494
	V40	Conjuntos de dados que possuem informações da fonte oficial dos dados	8361	0.771	0.5	0.385
	<b>Valor Indicador Proveniência</b>					<b>0.879</b>
Estado Rio Grande do Sul	V39	Conjuntos de dados que possuem vínculo com uma organização	1143	0.993	0.5	0.497
	V40	Conjuntos de dados que possuem informações da fonte oficial dos dados	35	0.03	0.5	0.015
	<b>Valor Indicador Proveniência</b>					<b>0.512</b>
Município São Paulo	V39	Conjuntos de dados que possuem vínculo com uma organização	195	1	0.5	0.500
	V40	Conjuntos de dados que possuem informações da fonte oficial dos dados	107	0.549	0.5	0.274
	<b>Valor Indicador Proveniência</b>					<b>0.774</b>

Fonte: do Autor

#### C.4.2 Indicador Vocabulários

Este indicador verifica a existência de vocabulários que dão suporte à compreensão do significado dos dados e a flexibilização no uso dos conjuntos de dados por agentes computacionais. A literatura recomenda fortemente o uso dos vocabulários padrões para que além de melhorar a compreensão dos conjuntos de dados, isso contribua com comunicação entre pessoas de um domínio, a formalização na representação do conhecimento eliminando as contradições e inconsistências, a representação de conhecimento e reutilização. A Tabela 25 a seguir apresenta os dados coletados oriundos do portal selecionado.

##### Recomendações

- Utilizar vocabulários padrões sempre que possível;

- Disponibilizar vocabulários a todos os conjuntos de dados;
- Em caso de utilizar vocabulário próprio, disponibilizar a documentação para compreensão;

Tabela 25 – Variáveis Indicador Vocabulários

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V41	Disponibiliza informações sobre vocabulários próprios	NÃO	0	0.111	0.000
	V42	Vocabulários nos conjuntos de dados	0	0	0.111	0.000
	V43	Vocabulário Licenças	0	0	0.111	0.000
	V44	Vocabulário Proveniência	0	0	0.111	0.000
	V45	Vocabulários Qualidade	0	0	0.111	0.000
	V46	Vocabulário Versionamento	0	0	0.111	0.000
	V47	Vocabulário Histórico de Versão	0	0	0.111	0.000
	V48	Vocabulário Atualização	0	0	0.111	0.000
	V49	Vocabulário Feedback Usuário	0	0	0.111	0.000
<b>Valor Indicador Vocabulários</b>						<b>0.000</b>
Estado Rio Grande do Sul	V41	Disponibiliza informações sobre vocabulários próprios	NÃO	0	0.111	0.000
	V42	Vocabulários nos conjuntos de dados	0	0	0.111	0.000
	V43	Vocabulário Licenças	0	0	0.111	0.000
	V44	Vocabulário Proveniência	0	0	0.111	0.000
	V45	Vocabulários Qualidade	0	0	0.111	0.000
	V46	Vocabulário Versionamento	0	0	0.111	0.000
	V47	Vocabulário Histórico de Versão	0	0	0.111	0.000
	V48	Vocabulário Atualização	0	0	0.111	0.000
	V49	Vocabulário Feedback Usuário	0	0	0.111	0.000
<b>Valor Indicador Vocabulários</b>						<b>0.000</b>
	V41	Disponibiliza informações sobre vocabulários próprios	NÃO	0	0.111	0.000
	V42	Vocabulários nos conjuntos de dados	0	0	0.111	0.000

Município São Paulo	V43	Vocabulário Licenças	0	0	0.111	0.000
	V44	Vocabulário Proveniência	0	0	0.111	0.000
	V45	Vocabulários Qualidade	0	0	0.111	0.000
	V46	Vocabulário Versionamento	0	0	0.111	0.000
	V47	Vocabulário Histórico de Versão	0	0	0.111	0.000
	V48	Vocabulário Atualização	0	0	0.111	0.000
	V49	Vocabulário Feedback Usuário	0	0	0.111	0.000
<b>Valor Indicador Vocabulários</b>						<b>0.000</b>

Fonte: do Autor

## C.5 DIMENSÃO ESTRUTURA

### C.5.1 Indicador Pesquisa

Este indicador tem como objetivo verificar a existência de um campo de busca, de modo a realizar pesquisas no portal de dados. Outro aspecto que o indicador pesquisa verifica é o refinamento, onde permite aos usuários filtrarem os resultados, chegando mais próximo do desejado ou mesmo encontrando com maior precisão os conjuntos de dados, ou recursos. Em caso de não existir a possibilidade de pesquisa, torna-se difícil o acesso e localização dos conjuntos de dados isso reflete em um desinteresse dos *stakeholders*. No que concerne, em relação à inexistência da possibilidade do refinamento, impacta na dificuldade de navegação pelos resultados e na descoberta de potenciais dados para uso. A Tabela 26 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Disponibilizar barra de busca;
- Disponibilizar opções de refinamento de pesquisa (filtros);

Tabela 26 – Variáveis Indicador Pesquisa

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V50	Barra de pesquisa disponível	SIM	1	0.5	0.500
	V51	Refinamento disponível	SIM	1	0.5	0.500
<b>Valor Indicador Pesquisa</b>						<b>1</b>

Estado Rio Grande do Sul	V50	Barra de pesquisa disponível	SIM	1	0.5	0.500
	V51	Refinamento disponível	SIM	1	0.5	0.500
	<b>Valor Indicador Pesquisa</b>					<b>1</b>
Município São Paulo	V50	Barra de pesquisa disponível	SIM	1	0.5	0.500
	V51	Refinamento disponível	SIM	1	0.5	0.500
	<b>Valor Indicador Pesquisa</b>					<b>1</b>

Fonte: do Autor

### C.5.2 Indicador Conta Usuário

Este indicador verifica a existência da opção de conta de usuário. Normalmente esta conta disponibilizada é considerada uma conta de visitante, onde possibilita aos *stakeholders* acompanhar as atividades que são realizadas no portal ou mesmo atividades específicas. Esta opção de conta de usuário aumenta a interação e o interesse dos *stakeholders* sobre o portal de dados por possibilitar notificações das atividades por e-mail, a possibilidade de acompanhar publicações ou organizações de interesse. A Tabela 27 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Disponibilizar opção de cadastro e acesso de usuário;

Tabela 27 – Variáveis Indicador Conta Usuário

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	52	Disponibilidade de cadastro de usuário	NÃO	0	0.5	0.000
	V53	Disponibilidade de acesso de usuário	NÃO	0	0.5	0.000
	<b>Valor Indicador Conta Usuário</b>					<b>0.000</b>
Estado Rio Grande do Sul	V52	Disponibilidade de cadastro de usuário	NÃO	0	0.5	0.000
	V53	Disponibilidade de acesso de usuário	SIM	1	0.5	0.500
	<b>Valor Indicador Conta Usuário</b>					<b>0.500</b>
Município São Paulo	V52	Disponibilidade de cadastro de usuário	NÃO	0	0.5	0.000
	V53	Disponibilidade de acesso de usuário	NÃO	0	0.5	0.000

	<b>Valor Indicador Conta Usuário</b>	<b>0.000</b>
--	--------------------------------------	--------------

Fonte: do Autor

### C.5.3 Indicador Recomendações

Este indicador verifica a existência de um sistema de recomendação de conjuntos de dados. Deste modo, o usuário ao pesquisar um conjunto de dados pode receber recomendações sobre outros conjuntos de dados que possam ser de seu interesse ou mesmo relacionadas com o seu tema de busca. Desta maneira possibilita a uma melhor navegação e exploração dos conjuntos de dados. A Tabela 28 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- Disponibilizar sugestão de pesquisa na barra de busca;
- Disponibilizar sugestão de conjuntos de dados do mesmo Tema, autor, por exemplo;

Tabela 28 – Variáveis Indicador Recomendação

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V54	Sugestões na barra de pesquisa	NÃO	0	0.5	0.000
	V55	Recomendação de conjunto de dados semelhantes	NÃO	0	0.5	0.000
	<b>Valor Indicador Recomendação</b>					<b>0.000</b>
Estado Rio Grande do Sul	V54	Sugestões na barra de pesquisa	NÃO	0	0.5	0.000
	V55	Recomendação de conjunto de dados semelhantes	NÃO	0	0.5	0.000
	<b>Valor Indicador Recomendação</b>					<b>0.000</b>
Município São Paulo	V54	Sugestões na barra de pesquisa	NÃO	0	0.5	0.000
	V55	Recomendação de conjunto de dados semelhantes	NÃO	0	0.5	0.000
	<b>Valor Indicador Recomendação</b>					<b>0.000</b>

Fonte: do Autor

### C.5.4 Indicador Acessibilidade

Este indicador verifica se existem recursos de acessibilidade do portal baseado no WCAG. Este indicador utiliza com base na nota gerada pelo software Access Monitor, que verifica web sites com base na WCAG 2.1. Desta forma é possível obter uma nota e verificar o quão adequado o portal de dados está em relação a aspectos de acessibilidade. A Tabela 29 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- A nota pode variar de um a dez pontos, sendo que quanto mais próximo de 10, melhor serão os aspectos de acessibilidade apresentados pelo portal;

Tabela 29 – Variáveis Indicador Acessibilidade

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V56	Nota WCAG 2.1	5.5	0.55	1	0.550
	<b>Valor indicador Acessibilidade</b>					<b>0.550</b>
Estado do Rio Grande do Sul	V56	Nota WCAG 2.1	3.6	0.36	1	0.360
	<b>Valor indicador Acessibilidade</b>					<b>0.360</b>
Município São Paulo	V56	Nota WCAG 2.1	4.4	0.44	1	0.440
	<b>Valor indicador Acessibilidade</b>					<b>0.440</b>

Fonte: do Autor

### C.5.5 Indicador Multi-Idioma

Este indicador verifica a existência de idiomas diferentes do idioma nativo do portal. Isso possibilita uma maior abrangência em relação à descoberta de dados por diferentes *stakeholders* de diferentes países. Neste indicador está sendo verificada a disponibilidade de diferentes idiomas em relação ao portal, aspecto relacionado aos dados ou mesmo ao conteúdo dos recursos disponibilizados não serão verificados. A Tabela 30 a seguir apresenta os dados coletados oriundos do portal selecionado.

### Recomendações

- O portal deve disponibilizar outros idiomas, além do idioma nativo.

Tabela 30 – Variáveis Indicador Multi-Idioma

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V57	Disponibilidade em outros idiomas além do nativo	NÃO	0	1	0
	<b>Valor indicador Multi-Idioma</b>					<b>0</b>
Estado do Rio Grande do Sul	V57	Disponibilidade em outros idiomas além do nativo	NÃO	0	1	0
	<b>Valor indicador Multi-Idioma</b>					<b>0</b>
Município São Paulo	V57	Disponibilidade em outros idiomas além do nativo	SIM	1	1	1
	<b>Valor indicador Multi-Idioma</b>					<b>1</b>

Fonte: do Autor

### C.5.6 Indicador Visualização dos dados

Este indicador verifica a existência de mecanismos que possibilitem a visualização dos recursos pertencentes aos conjuntos de dados, sem a necessidade de baixar o arquivo ou utilizar algum software específico, podendo visualizar os dados pelo próprio portal. Desta forma, viabiliza uma melhor interação dos *stakeholders* sobre a exploração dos dados. A Tabela 31 a seguir apresenta os dados coletados oriundos do portal selecionado.

### Recomendações

- Disponibilizar a maior variedade possível de visualizações dos formatos de arquivos on-line;

Tabela 31 – Variáveis Indicador Visualização dos Dados

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
	V58	Visualizador de conteúdo HTML	NÃO	0	0.143	0.000
	V59	Visualizador de imagens	NÃO	0	0.143	0.000

Governo Federal	V60	Visualizador de texto	NÃO	0	0.143	0.000
	V61	Visualizador com uso de mapas	NÃO	0	0.143	0.000
	V62	Visualizador de gráficos	NÃO	0	0.143	0.000
	V63	Visualizador de PDF	NÃO	0	0.143	0.000
	V64	Visualizador de arquivos JSON	NÃO	0	0.143	0.000
<b>Valor Indicador Visualização</b>						<b>0.000</b>
Estado Rio Grande do Sul	V58	Visualizador de conteúdo HTML	NÃO	0	0.143	0.000
	V59	Visualizador de imagens	SIM	1	0.143	0.143
	V60	Visualizador de texto	SIM	1	0.143	0.143
	V61	Visualizador com uso de mapas	NÃO	0	0.143	0.000
	V62	Visualizador de gráficos	SIM	1	0.143	0.143
	V63	Visualizador de PDF	NÃO	0	0.143	0.000
	V64	Visualizador de arquivos JSON	NÃO	0	0.143	0.000
<b>Valor Indicador Visualização</b>						<b>0.429</b>
Município São Paulo	V58	Visualizador de conteúdo HTML	SIM	1	0.143	0.143
	V59	Visualizador de imagens	SIM	1	0.143	0.143
	V60	Visualizador de texto	SIM	1	0.143	0.143
	V61	Visualizador com uso de mapas	SIM	1		0.143
	V62	Visualizador de gráficos	SIM	1	0.143	0.143
	V63	Visualizador de PDF	SIM	1	0.143	0.143
	V64	Visualizador de arquivos JSON	NÃO	0	0.143	0.000
<b>Valor Indicador Visualização</b>						<b>0.857</b>

Fonte: do Autor

### C.5.7 Indicador Temática

Este indicador verifica a existência de categorias temáticas para melhor organizar os conjuntos de dados. A literatura aponta o uso de categorias para melhor organização dos conjuntos de dados, sendo que todos os conjuntos de dados devem estar vinculados a um grupo. Isso facilita sua descoberta, entendimento e a sua organização junto ao portal de dados. A Tabela 32 a seguir apresenta os dados coletados

oriundos do portal selecionado.

### Recomendações

- O portal deve disponibilizar categorias ou grupos temáticos;
- Todos os conjuntos de dados devem estar vinculado a um ou mais grupos;

Tabela 32 – Variáveis Indicador Temática

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V65	Conjunto de dados com pelo menos um grupo vinculado	179	0.016	1	0.016
						<b>Valor Indicador Temática</b>
Estado Rio Grande do Sul	V65	Conjunto de dados com pelo menos um grupo vinculado	1095	0.951	1	0.951
						<b>Valor Indicador Temática</b>
Município São Paulo	V65	Conjunto de dados com pelo menos um grupo vinculado	152	0.779	1	0.779
						<b>Valor Indicador Temática</b>

Fonte: do Autor

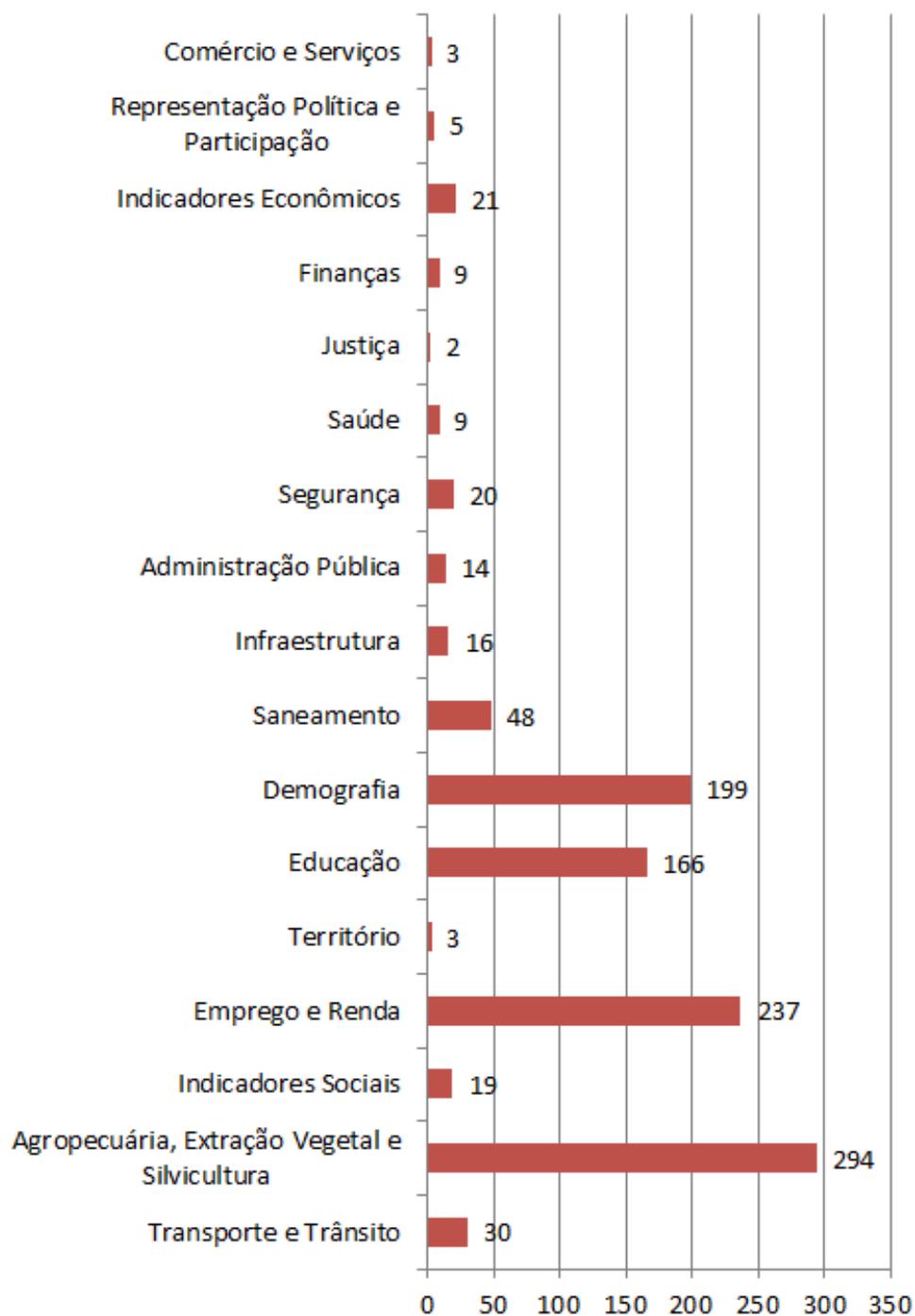
Os Gráficos 33, 34 e 35 apresentam os temas vinculados aos conjuntos de dados. É importante destacar que os valores ressaltados nos gráficos são correspondentes à quantidade de conjunto de dados atribuídos ao tema, sendo que um conjunto de dado pode possuir mais de um tema e este pode estar vinculado a muitos conjuntos de dados.

Gráfico 33 – Temas x Conjuntos de dados - Portal governo federal



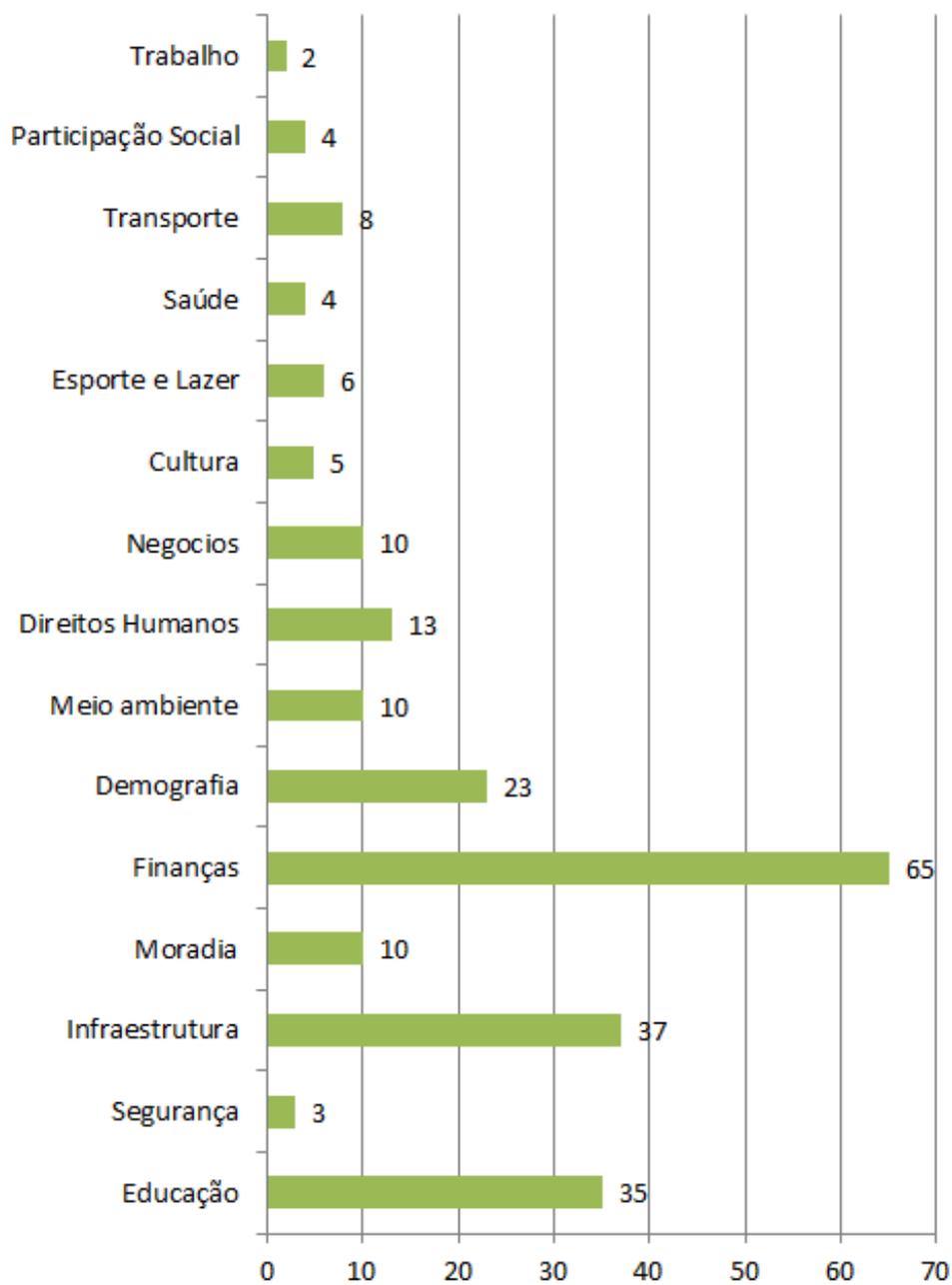
Fonte: O Autor

Gráfico 34 – Temas x Conjuntos de dados - Portal estado do Rio Grande do Sul



Fonte: O Autor

Gráfico 35 – Temas x Conjuntos de dados - Portal município de São Paulo



Fonte: O Autor

### C.5.8 Indicador API|SPARQL

Esta métrica verifica se existe o acesso por API ou SPARQL na plataforma. A disponibilização de API é importante, pois simplifica o acesso aos dados e informações através de requisições em HTTP. O SPARQL também possui sua importância por ser uma linguagem de consulta de grafos em RDF complexos, sendo uma das tecnologias base para emprego da web semântica. A Tabela 33 a seguir apresenta os dados coletados oriundos do portal selecionado.

#### Recomendações

- O portal deve disponibilizar o acesso API ou SPARQL (sendo estes não excluídos)
- Disponibilização da documentação para uso do API;

Tabela 33 – Variáveis Indicador API|SPARQL

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V66	Acesso via API	SIM	1	0.333	0.333
	V67	Acesso via SPARQL	NÃO	0	0.333	0.000
	V68	Informações sobre API	NÃO	0	0.333	0.000
	<b>Valor Indicador API   SPARQL</b>					<b>0.333</b>
Estado Rio Grande do Sul	V66	Acesso via API	SIM	1	0.333	0.333
	V67	Acesso via SPARQL	NÃO	0	0.333	0.000
	V68	Informações sobre API	NÃO	0	0.333	0.000
	<b>Valor Indicador API   SPARQL</b>					<b>0.333</b>
Município São Paulo	V66	Acesso via API	SIM	1	0.333	0.333
	V67	Acesso via SPARQL	NÃO	0	0.333	0.000
	V68	Informações sobre API	SIM	1	0.333	0.333
	<b>Valor Indicador API   SPARQL</b>					<b>0.667</b>

Fonte: do Autor

### C.5.9 Indicador Interatividade

Este indicador verifica aspectos que possibilitem a interação entre *stakeholders*, com o portal e seus responsáveis. Para isso, é verificada a existência de um meio (formulários, por exemplo) para solicitar aos órgãos conjuntos de dados que não estejam publicados ou acessíveis, um espaço para fórum, comentários ou mesmo a possibilidade de efetuar avaliações sobre o portal, ou os conjuntos de dados e seus recursos.

Outro ponto verificado é o compartilhamento através das mídias sociais isto possibilita aos *stakeholders* externalizar os dados publicados para além de seu ecossistema, amplificando assim suas ideias e refletindo também em quesitos de responsabilização e transparência do órgão a qual o portal pertence. Por fim, é verificada a possibilidade de utilizar citações em relação aos conjuntos de dados publicados isso possibilita a pesquisadores, jornalistas, dentre outros referenciarem os dados utilizados e por consequência aumentando as chances de descoberta do portal e seus dados. A Tabela 34 a seguir apresenta os dados coletados oriundos do portal selecionado.

### Recomendações

- Disponibilizar meios para a solicitação de conjuntos de dados;
- Disponibilizar meios para o compartilhamento nas redes sociais;
- Disponibilizar espaços para comentários, avaliações ou mesmo fóruns para discussão;
- Disponibilizar formatos de citação dos conjuntos de dados ou recurso;

Tabela 34 – Variáveis Indicador Interatividade

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V69	Formulários ou informações sobre a solicitação de conjuntos de dados	SIM	1	0.125	0.125
	V70	Disponibilidade de citações	NÃO	0	0.125	0.000
	V71	Disponibilidade de fórum	NÃO	0	0.125	0.000
	V72	Disponibilidade de comentário	NÃO	0	0.125	0.000
	V73	Disponibilidade de avaliação	SIM	1	0.125	0.125
	V74	Compartilhamento no Facebook	SIM	1	0.125	0.125
	V75	Compartilhamento no Google	SIM	1	0.125	0.125
	V76	Compartilhamento no Twitter	SIM	1	0.125	0.125
<b>Valor Indicador Interatividade</b>						<b>0.625</b>
	V69	Formulários ou informações sobre a solicitação de conjuntos de dados	SIM	1	0.125	0.125
	V70	Disponibilidade de citações	NÃO	0	0.125	0.000
	V71	Disponibilidade de fórum	NÃO	0	0.125	0.000

Estado Rio Grande do Sul	V72	Disponibilidade de comentário	NÃO	0	0.125	0.000
	V73	Disponibilidade de avaliação	NÃO	0	0.125	0.000
	V74	Compartilhamento no Facebook	SIM	1	0.125	0.125
	V75	Compartilhamento no Google	SIM	1	0.125	0.125
	V76	Compartilhamento no Twitter	SIM	1	0.125	0.125
	<b>Valor Indicador Interatividade</b>					
Município São Paulo	V69	Formulários ou informações sobre a solicitação de conjuntos de dados	NÃO	0	0.125	0
	V70	Disponibilidade de citações	NÃO	0	0.125	0.000
	V71	Disponibilidade de fórum	NÃO	0	0.125	0.000
	V72	Disponibilidade de comentário	NÃO	0	0.125	0.000
	V73	Disponibilidade de avaliação	NÃO	0	0.125	0.000
	V74	Compartilhamento no Facebook	SIM	1	0.125	0.125
	V75	Compartilhamento no Google	SIM	1	0.125	0.125
	V76	Compartilhamento no Twitter	SIM	1	0.125	0.125
<b>Valor Indicador Interatividade</b>						<b>0.375</b>

Fonte: do Autor

## C.6 DIMENSÃO APRESENTAÇÃO

### C.6.1 Indicador Compreensão

Este indicador verifica a existência de uma ou mais páginas Web que tragam informações sobre o portal, como exemplo: o seu objetivo, periodicidade de atualização, responsáveis, a disponibilização das perguntas frequentes com suas respectivas respostas. Outro ponto que pode ser disponibilizado, são os projetos concluídos (por exemplo, aplicativos), oriundos dos dados publicados. Tais projetos, em sua maioria, são instigados ou mesmo patrocinados pelas organizações a fim de promover o uso dos dados, aumentar a interação com a sociedade e *stakeholders*, por exemplo. Estas informações são importantes, pois facilitam a compreensão e entendimento dos objetivos e a importância do portal para o seu ecossistema ou mesmo para a sociedade no que pertence. A Tabela 35 a seguir apresenta os dados coletados oriundos do portal selecionado.

### Recomendações

- Disponibilizar um espaço no portal (página) com a apresentação do portal, objetivos, periodicidade, dentre outros;
- Disponibilizar um espaço no portal (página) com as perguntas e respostas mais frequentes;

Tabela 35 – Variáveis Indicador Compreensão

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V77	Introdução sobre o portal de dados	SIM	1	0.333	0.333
	V78	Disponibilidade de perguntas mais frequentes (FAQ)	SIM	1	0.333	0.333
	V79	Disponibilização de projetos	SIM	1	0.333	0.333
	<b>Valor Indicador Compreensão</b>					<b>1.000</b>
Estado Rio Grande do Sul	V77	Introdução sobre o portal de dados	SIM	1	0.333	0.333
	V78	Disponibilidade de perguntas mais frequentes (FAQ)	NÃO	0	0.333	0
	V79	Disponibilização de projetos	SIM	1	0.333	0.333
	<b>Valor Indicador Compreensão</b>					<b>0,667</b>
Município São Paulo	V77	Introdução sobre o portal de dados	SIM	1	0.333	0.333
	V78	Disponibilidade de perguntas mais frequentes (FAQ)	NÃO	0	0.333	0
	V79	Disponibilização de projetos	SIM	1	0.333	0.333
	<b>Valor Indicador Compreensão</b>					<b>0,667</b>

Fonte: do Autor

### C.6.2 Indicador Quantitativos

Verifica se existem informações gerais e quantitativas sobre os conjuntos de dados e sobre aspectos do portal de dados abertos, como, por exemplo, número de visitantes, número de downloads, entre outros. Este resumo dos dados quantitativos permite aos *stakeholders* perceber se o portal é explorado, se possui uma quantidade razoável de conjuntos de dados ou mesmo uma maneira simples de perceber o cresci-

mento do portal. A Tabela 36 a seguir apresenta os dados coletados oriundos do portal selecionado.

### Recomendações

- Disponibilizar o maior número de informações quantitativas possíveis, por exemplo:
  - Número de downloads;
  - Total de conjunto de dados;
  - Número de projetos desenvolvidos;
  - Total de recursos;
  - Total geral em MB de dados;
  - Total de editores/parceiros;

Tabela 36 – Variáveis Indicador Quantitativos

Portal	ID	Variáveis	Valor	Valor Normalizado	Peso Normalizado	Valor Ponderado
Governo Federal	V80	Quadro com o resumo das atividades do portal	NÃO	0	0.111	0.000
	V81	Disponibiliza número de downloads	NÃO	0	0.111	0.000
	V82	Disponibiliza total de conjunto de dados	SIM	1	0.111	0.111
	V83	Disponibiliza número de projetos desenvolvidos	SIM	1	0.111	0.111
	V84	Disponibiliza total de recursos	SIM	1	0.111	0.111
	V85	Disponibiliza total em MB de dados	NÃO	0	0.111	0.000
	V86	Disponibiliza total de editores	SIM	1	0.111	0.111
	V87	Disponibiliza total de catálogos coletados	NÃO	0	0.111	0.000
	V88	Disponibiliza número de visitas	NÃO	0	0.111	0.000
	<b>Valor Indicador Quantitativos</b>					
	V80	Quadro com o resumo das atividades do portal	NÃO	0	0.111	0
	V81	Disponibiliza número de downloads	NÃO	0	0.111	0.000

Estado Rio Grande do Sul	V82	Disponibiliza total de conjunto de dados	SIM	1	0.111	0.111
	V83	Disponibiliza número de projetos desenvolvidos	SIM	1	0.111	0.111
	V84	Disponibiliza total de recursos	NÃO	0	0.111	0.000
	V85	Disponibiliza total em MB de dados	NÃO	0	0.111	0.000
	V86	Disponibiliza total de editores	SIM	1	0.111	0.111
	V87	Disponibiliza total de catálogos coletados	NÃO	0	0.111	0.000
	V88	Disponibiliza número de visitas	NÃO	0	0.111	0.000
<b>Valor Indicador Quantitativos</b>						<b>0.333</b>
Município São Paulo	V80	Quadro com o resumo das atividades do portal	SIM	1	0.111	0.111
	V81	Disponibiliza número de downloads	NÃO	0	0.111	0.000
	V82	Disponibiliza total de conjunto de dados	SIM	1	0.111	0.111
	V83	Disponibiliza número de projetos desenvolvidos	SIM	1	0.111	0.111
	V84	Disponibiliza total de recursos	NÃO	0	0.111	0.000
	V85	Disponibiliza total em MB de dados	NÃO	0	0.111	0.000
	V86	Disponibiliza total de editores	SIM	1	0.111	0.111
	V87	Disponibiliza total de catálogos coletados	NÃO	0	0.111	0.000
	V88	Disponibiliza número de visitas	NÃO	0	0.111	0.000
<b>Valor Indicador Quantitativos</b>						<b>0.444</b>

Fonte: do Autor

## APÊNDICE D – QUESTIONÁRIO APLICADO NO GRUPO DE PROFISSIONAIS DA ÁREA DE DADOS ABERTOS

Análise do modelo de referência para a avaliação de portais de dados abertos  
e do Sistema Avaliação de Portais de Dados Abertos (SiAPDA)

Quadro 49 – Questionário utilizado na pesquisa de campo

Informação do avaliador				
Data:				
Nome (Opcional):				
Organização/Instituição:				
Cargo:				
E-mail:				
Critérios de avaliação	Discordo Fortemente	Discordo	Concordo	Concordo Fortemente
<b>Modelo de Referência</b>				
O modelo proposto é adequado para avaliar a aplicação das boas práticas de disponibilização em portais de dados abertos.				
O modelo proposto possibilita identificar deficiências relacionadas às boas práticas de disponibilização em portais de dados abertos.				
O modelo proposto possui uma estrutura (dimensões, indicadores e variáveis) consistente e compreensível.				
O modelo proposto pode ser aplicado em portais de dados abertos no âmbito, federal, estadual, municipal e outras organizações.				
O modelo proposto auxilia na qualidade dos portais de dados abertos.				
O modelo proposto requer conhecimento especializado para sua aplicação.				
<b>Sistema Computacional</b>				
A criação de um sistema computacional para a automatização da coleta, avaliação e análise contribui para o uso do modelo proposto.				

O uso dos padrões de ponderação de avaliação dos portais atende as particularidades dos diferentes portais de dados abertos.				
O sistema computacional apresenta interação adequada com o usuário para avaliar portais de dados.				
Concorda em utilizar o sistema computacional para avaliar o portal de dados de sua organização.				

Caso tenha optado por "Discordo Fortemente" em alguma das afirmativas acima, quais suas sugestões para a melhoria?

Deixe sua sugestão, crítica ou elogio para o autor.