

UNIVERSIDADE FEDERAL DE SANTA CATARINA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**PROPOSTA PARA DOCUMENTAÇÃO DE REQUISITOS EM
PROJETOS DE DATA WAREHOUSE**

Florianópolis

2004

Hegler Correa Tissot

**PROPOSTA PARA DOCUMENTAÇÃO DE REQUISITOS EM
PROJETOS DE DATA WAREHOUSE**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção do grau de Mestre em Engenharia de Produção.

Orientador: Aran Bey Tcholakian Morales

Florianópolis

2004

Hegler Correa Tissot

**PROPOSTA PARA DOCUMENTAÇÃO DE REQUISITOS EM
PROJETOS DE DATA WAREHOUSE**

Esta dissertação foi julgada e aprovada para a obtenção do grau de Mestre em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

Florianópolis, 12 de novembro de 2004.

Edson Pacheco Paladini, Dr.
Coordenador do Programa

BANCA EXAMINADORA

Prof. Aran Bey Tcholakian Morales, Dr.
ORIENTADOR

Prof. Vinícius Medina Kern

Prof. José Leomar Todesco, Dr.

À minha esposa, Laura, pelo constante apoio.

A meu filho, Felipe, pela fundamental ajuda prestada.

RESUMO

TISSOT, Hegler Correa. **Proposta para documentação de requisitos em projetos de Data Warehouse**. 2004. 148f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

O uso estratégico da informação torna-se uma necessidade cada vez mais presente nas organizações e a construção de ferramentas que possibilitem o uso das informações para essa finalidade tem evoluído para acompanhar a própria evolução da tecnologia, com a aplicação de processos de desenvolvimento de software e o uso de metodologia baseada em data warehousing.

Processos de desenvolvimento de software envolvem tanto profissionais de sistemas quanto usuários finais, que devem definir as necessidades e os requisitos a serem seguidos na construção do sistema. Falhas surgem na análise de requisitos geralmente quando a equipe técnica não possui o discernimento adequado para compreender as necessidades dos usuários ou quando é utilizada uma linguagem extremamente técnica para documentar tais necessidades; os usuários têm dificuldade de entender essa linguagem técnica e podem ficar sem argumentação contra a documentação apresentada.

Este trabalho apresenta uma proposta para que o processo de análise de requisitos de um data warehouse utilize um conjunto de artefatos que direcione o levantamento desses requisitos e que utilize uma linguagem mais próxima daquela utilizada no cotidiano dos usuários, diminuindo o risco inerente a esse processo.

Baseado nessa proposta, o trabalho apresenta os resultados obtidos no levantamento de requisitos de um sistema de apoio à decisão, orientado à análise de informações do Diretório de Grupos de Pesquisa da Plataforma Lattes.

Palavras-chaves: data warehouse, análise de requisitos, apoio à tomada de decisão

ABSTRACT

TISSOT, Hegler Correa. **Proposta para documentação de requisitos em projetos de Data Warehouse**. 2004. 148f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

The strategical use of the information becomes a necessity each more present in the organizations. The construction of tools that make possible the use of the information for this purpose has evolved to follow the proper evolution of the technology, with the application software development processes and the use of data warehousing based development methodology.

Software development processes in such a way involve systems analysts and also end users, that must define the needs and the requirements to be followed in the construction of the system. Imperfections appear generally in the requirements analysis when the technical team does not possess the adjusted discernment to understand the end users needs or when a extremely technical language is used language to register such necessities; the users have difficulty to understand this language and can be without argument against the presented documentation.

This work presents a proposal so that the process of data warehouse requirements analysis uses a set of devices that drives the survey of these requirements and uses a language next to that one used in the daily one of the system end users, diminishing the inherent risk to this process.

Based in this proposal, the work presents the results gotten in the survey of decision support system requirements, guided to the analysis of information contained in the “Diretório de Grupos de Pesquisa” of the “Plataforma Lattes” systems.

Key-words: data warehouse, requirements analysis, decisions taking support

SUMÁRIO

1	INTRODUÇÃO	1
1.1	APRESENTAÇÃO	1
1.2	OBJETIVOS	3
1.3	JUSTIFICATIVA	3
1.4	METODOLOGIA DE PESQUISA	4
1.5	ESTRUTURA DA DISSERTAÇÃO	6
2	TECNOLOGIA DA INFORMAÇÃO PARA SUPORTE À DECISÃO	7
2.1	SISTEMAS DE INFORMAÇÃO	9
2.2	<i>DATA WAREHOUSE</i>	13
2.3	RATIONAL UNIFIED PROCESS (RUP)	22
2.4	ANÁLISE DE REQUISITOS PARA SISTEMAS DE INFORMAÇÃO	28
2.5	UML	38
2.6	CONCLUSÕES	51
3	PROPOSTA PARA DOCUMENTAÇÃO DE REQUISITOS EM PROJETOS DE DATA WAREHOUSES	53
3.1	ARTEFATOS PROPOSTOS PARA REQUISITOS DE DW	53
3.2	CLASSIFICAÇÃO DOS REQUISITOS	63
3.3	CONCLUSÕES	77
4	ESTUDO DE CASO	79
4.1	INTRODUÇÃO	79
4.2	ESTUDO DE CASO: DIRETÓRIO LATTES DE GRUPOS DE PESQUISA	79
4.3	CONCLUSÕES	140
5	CONCLUSÕES E TRABALHOS FUTUROS	141
5.1	CONTEXTO	141
5.2	LINGUAGEM E COMUNICAÇÃO	144
5.3	TRABALHOS FUTUROS	145
5.4	CONSIDERAÇÕES FINAIS	145
	REFERÊNCIAS	146

LISTA DE FIGURAS

Figura 1 - Metodologia de pesquisa utilizada.....	6
Figura 2 - Papéis fundamentais dos Sistemas de Informação.....	10
Figura 3 - Arquitetura de um sistema de <i>Business Intelligence</i>	13
Figura 4 - Exemplo de um modelo dimensional com tabelas de fato e dimensões.....	19
Figura 5 - Atributos podem ser adicionados a uma consulta para detalhar o resultado.....	21
Figura 6 - Algumas disciplinas, fases e iterações do processo RUP.....	23
Figura 7 - Impacto de requisitos em um Data Warehouse.....	30
Figura 8 - Níveis de abstração de requisitos em um Data Warehouse.....	34
Figura 9 - Exemplo de um diagrama de casos de uso.....	42
Figura 10 - Exemplo de especificação de um caso de uso.....	43
Figura 11 - Exemplo de um modelo de casos de uso para um sistema analisado.....	45
Figura 12 - Exemplo de um diagrama de estados.....	47
Figura 13 - Exemplo de um diagrama de atividades.....	49
Figura 14 - Exemplo de um diagrama de classes.....	51
Figura 15 - Exemplo de uma dimensão tempo com diferentes granularidades para uma data.....	65
Figura 16 - Exemplo de uma dimensão tempo com granularidades para horas e minutos.....	66
Figura 17 - Exemplo de uma dimensão para representar a distribuição geográfica.....	67
Figura 18 - Modelagem de dimensões que tratam de hierarquias recursivas.....	69
Figura 19 - Origens de dados e funcionalidades do DM do Diretório de Grupos de Pesquisa.....	86
Figura 20 - Diagrama de atores do Data Mart do Diretório de Grupos de Pesquisa.....	87
Figura 21 - Diagrama de casos de uso do Sistema Diretório de Grupos de Pesquisa.....	89
Figura 22 - Diagrama de casos de uso do Data Mart do Diretório de Grupos de Pesquisa.....	89

LISTA DE QUADROS

Quadro 1 - Atividades, tarefas e artefatos da disciplina de modelagem de negócios.....	24
Quadro 2 - Atividades, tarefas e artefatos da disciplina de análise de requisitos	26
Quadro 3 - Atividades, tarefas e artefatos da disciplina de análise e projeto	27
Quadro 4 - Atividades, tarefas e artefatos da disciplina de implementação	28
Quadro 5 - Estrutura básica para a documentação de requisitos para DW	54
Quadro 6 - Relação entre a nomenclatura conceitual e técnica para requisitos de informação.....	73
Quadro 7 - Especificação do caso de uso “Súmula Estatística”	92
Quadro 8 - Especificação do caso de uso “Consulta de indicadores”	93
Quadro 9 - Especificação do caso de uso “Séries Históricas”	94
Quadro 10 - Especificação do caso de uso “Plano Tabular”	95
Quadro 11 - Especificação do caso de uso “Busca Textual”	96
Quadro 12 - Especificação do caso de uso “Estratificação de Grupos”	97
Quadro 13 - Unidades de Análise e de Conteúdo do Data Mart de Grupos de Pesquisa.....	99
Quadro 14 - Propriedades adicionais das Unidades de Análise	99
Quadro 15 - Unidades de Conteúdo do site “Estratificação de Grupos”	100
Quadro 16 - Dimensões utilizadas por cada unidade de análise	106

LISTA DE TABELAS

Tabela 1 - Unidades de distribuição da dimensão “Distribuição Geográfica”	101
Tabela 2 - Unidades de Distribuição da dimensão “Área do conhecimento”	101
Tabela 3 - Unidades de Distribuição da dimensão “Censo”.....	102
Tabela 4 - Unidades de Distribuição da dimensão “Características Pessoais”	103
Tabela 5 - Unidades de Distribuição da dimensão “Tipo de produção”.....	103
Tabela 6 - Unidades de Distribuição da dimensão “Atividade técnica”	104
Tabela 7 - Unidades de Distribuição da dimensão “Empresa”.....	104
Tabela 8 - Unidades de Distribuição da dimensão “Tipo de relacionamento”	105
Tabela 9 - Unidades de Distribuição da dimensão “Forma de remuneração”	105
Tabela 10 - Unidades de Distribuição da dimensão “Setor da atividade econômica”	105
Tabela 11 - Indicadores do site “Séries Históricas” (séries básicas).....	108
Tabela 12 - Indicadores do site “Séries Históricas” disponíveis para “Grupos de Pesquisa”	109
Tabela 13 - Indicadores do site “Séries Históricas” disponíveis para “Linhas de pesquisa”	110
Tabela 14 - Indicadores do site “Séries Históricas” disponíveis para “Pesquisadores”	112
Tabela 15 - Indicadores do site “Séries Históricas” disponíveis para “Produção C&T”	112
Tabela 16 - Indicadores do site “Súmula Estatística” disponíveis para “Grupos de pesquisa”.....	115
Tabela 17 - Indicadores do site “Súmula Estatística” disponíveis para “Linhas de pesquisa”.....	117
Tabela 18 - Indicadores do site “Súmula Estatística” disponíveis para “Pesquisadores”.....	120
Tabela 19 - Indicadores do site “Súmula Estatística” disponíveis para “Estudantes”	121
Tabela 20 - Indicadores do site “Súmula Estatística” disponíveis para “Técnicos”.....	122
Tabela 21 - Indicadores do site “Súmula Estatística” disponíveis para “Produção C&T”.....	123
Tabela 22 - Indicadores do site “Súmula Estatística” disponíveis para “Empresas”	125
Tabela 23 - Conteúdo apresentado como resultado do site “Busca Textual”	128
Tabela 24 - Unidades utilizadas para a geração da descrição de uma consulta do “Plano Tabular” .	133
Tabela 25 - Regras de formação de notas de rodapé	134
Tabela 26 - Fontes de dados das Unidades de Análise e Conteúdo	134
Tabela 27 - Fontes de dados das Unidades de Distribuição (Dimensões).....	134
Tabela 28 - Tipos de relacionamento possíveis entre empresas e grupos de pesquisa.....	135
Tabela 29 - Formas de remuneração possíveis entre empresas e grupos de pesquisa.....	135
Tabela 30 - Operações OLAP disponíveis no “Plano Tabular”	136
Tabela 31 - Unidades de Análise mapeadas como tabela de dimensão.....	144

ABREVIATURAS

BI	<i>Business Intelligence</i>
DGP	Diretório de Grupos de Pesquisa no Brasil
DM	Data Mart
DW	Data Warehouse
EIS	Sistemas de Informação executivos
OLAP	Processamento Analítico On-line
RUP	<i>Rational Unified Process</i>
SAD	Sistemas de apoio à decisão
UML	<i>Unified Modeling Language</i>

1 INTRODUÇÃO

1.1 Apresentação

Na atualidade, as empresas estão evoluindo a forma de tomar decisões. Nessa evolução passam a utilizar a informação existente em seus sistemas como auxílio ao processo decisório para permitir que diretores, gerentes e administradores tomem decisões mais rápidas, reagindo às necessidades de negócio e tornando-se mais competitivas no mercado.

O uso da informação de forma estratégica e para o apoio ao processo decisório torna-se uma realidade cada vez mais presente nas empresas, à medida que se verificam o constante avanço tecnológico e o contínuo aumento da capacidade de tratamento de dados decorrente dessa evolução, marcada principalmente por três épocas:

- a década de 80 foi dedicada à construção e organização de bancos de dados operacionais, com foco principalmente no armazenamento de informações geradas pelos processos transacionais;
- a década de 90 destacou-se pela automação de processos de negócio, o que motivou a criação de grandes sistemas de gerenciamento empresarial;
- na atualidade, as empresas motivam-se a encontrar formas de extrair conhecimento dos bancos de dados para melhor entender os negócios e para alcançar um bom desempenho.

Como reflexo dessa evolução natural dos sistemas de informação, nas primeiras iniciativas de informatização de um ambiente de trabalho, os esforços estão relacionados à construção de sistemas de apoio às atividades operacionais e do dia-

a-dia. À medida que o acesso às informações é disponibilizado, os usuários desses sistemas passam a ter novas necessidades para apoiar suas atividades analíticas, exigindo uma melhor organização dos dados. Essas novas necessidades motivam a construção de um ambiente computacional específico, com o objetivo de atender às atividades analíticas e de tomada de decisão (GARCIA et al, 2001). Para satisfazer essas necessidades, ferramentas e técnicas de análise de dados foram criadas com o objetivo de transformar esses dados em informação útil que possa ser utilizada nos processos de tomada de decisão.

Um projeto de um sistema de informação passa por um processo de desenvolvimento de software que envolve tanto os profissionais de sistemas de informação quanto os usuários finais das aplicações, e esses perfis de profissionais devem interagir e comunicar-se de forma a estabelecer as necessidades e os requisitos a serem seguidos na construção do sistema. Diferentes dos sistemas de informação convencionais, os sistemas de informação que apóiam o processo de tomada de decisão possuem características técnicas de construção a serem consideradas no seu projeto, as quais devem estar de acordo com as necessidades analíticas de seus usuários.

Da mesma forma que os sistemas de informação evoluíram, também evoluíram as técnicas empregadas para projetar esses sistemas; um problema constante observado nos processos de desenvolvimento de sistemas, entretanto, é o fato de os profissionais envolvidos nesse processo utilizarem diferentes linguagens: por um lado os usuários referem-se às suas necessidades de forma conceitual e utilizam uma linguagem mais voltada para as áreas de negócio em que atuam; de outro lado, os profissionais de sistemas de informação fazem uso de um vocabulário técnico, na maioria das vezes incompreendido pelos usuários das aplicações, que comumente

gera conflitos entre o que é especificado como necessidade pelos usuários e o que é projetado e apresentado como uma proposta de solução.

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo geral deste trabalho é apresentar uma proposta para a documentação de requisitos de projetos de sistemas de informação que serão utilizados como apoio ao processo decisório e estratégico, através do uso de uma linguagem acessível, de fácil compreensão e que permita que a documentação gerada possa ser analisada e validada pelos principais usuários do sistema.

1.2.2 Objetivos específicos

- Definir uma taxonomia referente aos principais conceitos técnicos relacionados a sistemas de apoio à tomada de decisões, permitindo que esses conceitos possam ser melhor compreendidos pelos usuários desses sistemas.
- Estruturar um conjunto de artefatos que conduzam a uma documentação de requisitos de fácil compreensão, através do uso de uma linguagem compatível com o domínio de negócio e não focada na tecnologia.
- Identificar artefatos que possam ser utilizados na documentação de requisitos para o tipo de sistema de informação em questão e que possuam características mais voltadas à descrição conceitual do que à descrição técnica do projeto desses sistemas.

1.3 Justificativa

A construção de sistemas de informação falha no aspecto de análise de requisitos geralmente quando a equipe técnica não possui o discernimento

adequado para compreender as necessidades dos usuários ou quando é utilizada uma linguagem extremamente técnica para documentar tais necessidades. Usuários finais têm dificuldade de entender essa linguagem técnica e podem ficar sem argumentação contra a documentação apresentada, freqüentemente validando algo que nem sequer compreendem. Esse problema, não menos grave em sistemas de informação convencionais, pode tomar maiores proporções quando observado no projeto de um sistema de informação que será utilizado para apoiar decisões estratégicas de uma empresa.

Como principal consequência disso, são projetados e criados sistemas de informação que não condizem com a real necessidade dos usuários, e, na maioria desses casos, esses sistemas caem em desuso quando não são imediatamente reprovados logo após a sua implantação.

Por esse motivo, utilizar no processo de análise de requisitos de um sistema de informação um conjunto de artefatos que direcione o levantamento desses requisitos e que utilize uma linguagem mais próxima daquela utilizada no cotidiano dos usuários poderá diminuir o risco inerente a esse processo, ao contrário de quando é utilizada uma linguagem de especificação extremamente técnica.

1.4 Metodologia de pesquisa

Este trabalho está baseado na apresentação de um estudo de caso, e fundamentado nas etapas descritas a seguir:

A primeira etapa consistiu em avaliar através da literatura as principais técnicas utilizadas para a documentação de requisitos que são empregadas no projeto de sistemas de informação convencionais e aquelas específicas para projetos de sistemas de apoio à decisão.

Na segunda etapa foi feita uma crítica sobre os artefatos encontrados, e cada um deles foi classificado conforme a possibilidade de ser legível a um usuário (linguagem de apresentação conceitual) ou somente a um analista de requisitos (linguagem de apresentação técnica).

A terceira etapa consistiu em criar uma nova nomenclatura para os conceitos técnicos envolvidos no projeto e na construção de um sistema de apoio à decisão para permitir que esses conceitos possam ser melhor compreendidos pelos usuários.

Com isso foi possível a criação de uma proposta para a documentação de requisitos que fosse adequada às necessidades exigidas por um projeto de um sistema de apoio à decisão e que utilizasse linguagem e notação conceituais acessíveis aos usuários, permitindo assim uma melhor compreensão da proposta de solução que será apresentada pela equipe técnica.

Essa proposta foi aplicada em um estudo de caso para o projeto do Diretório de Grupos de Pesquisa no Brasil da Plataforma Lattes, que serviu para testar, validar e refinar o artefato de documentação de requisitos proposto.

A Figura 1 exemplifica a metodologia utilizada.

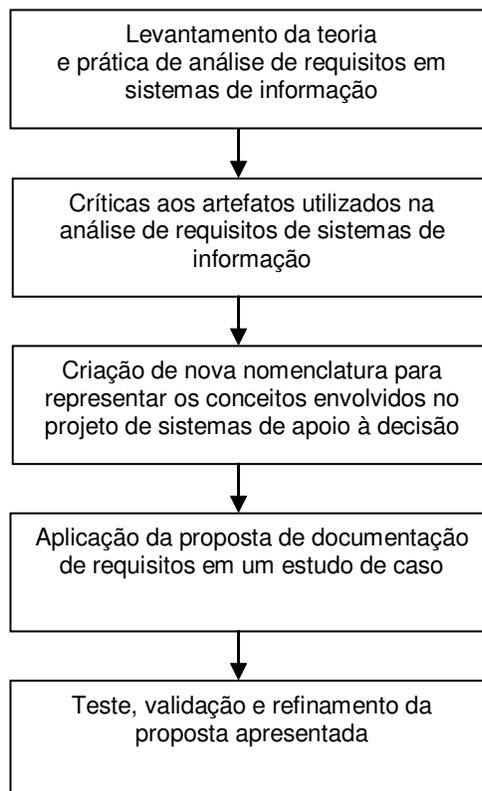


Figura 1 - Metodologia de pesquisa utilizada

1.5 Estrutura da dissertação

Este trabalho está organizado em cinco capítulos. No Capítulo 2 são revisados os principais conceitos relacionados com a construção e os projetos de sistemas de informação. O Capítulo 3 apresenta a proposta de documentação de requisitos; o Capítulo 4, o estudo de caso em que foi aplicada essa proposta; e no Capítulo 5, as conclusões.

2 TECNOLOGIA DA INFORMAÇÃO PARA SUPORTE À DECISÃO

A evolução dos sistemas de informação reflete a forma como a informação passou a ser utilizada dentro de cada um dos tipos de sistema, desde o surgimento dos sistemas de armazenamento transacional até a criação de grandes bases de informação utilizadas como suporte ao processo decisório.

Um sistema de informação representa uma solução para as necessidades ou para os problemas que se apresentam no mundo real. Foram criados processos para o desenvolvimento de cada um desses tipos de sistemas que integram um conjunto de técnicas e artefatos utilizados em cada uma das fases desses processos (RATIONAL, 2002). Tais técnicas são utilizadas para mapear os problemas em requisitos de software, os quais são empregados no projeto e na conseqüente construção de uma solução computacional.

Os sistemas utilizados como ferramentas de apoio ao processo decisório possuem características técnicas e funcionais que os diferenciam dos sistemas de processamento transacional. Essas diferenças interferem diretamente nos métodos utilizados para análise e projeto desses sistemas.

As técnicas de Data Warehousing são aplicadas para a criação de modelos de informação e sistemas de apoio ao processo decisório. Para esse tipo de solução, são utilizados processos de desenvolvimento de sistemas, como RUP.

Nesse enfoque, tipos específicos de sistemas, com características conceituais particulares, podem exigir adaptações nos processos de desenvolvimento utilizados. Essas adaptações podem ser facilmente aplicáveis quando o processo utiliza-se de uma linguagem de modelagem com características de extensibilidade, como é o

caso da UML, comumente utilizada na representação gráfica e visual de projetos de software.

Por fim, a análise de requisitos, como fase inicial de um processo de desenvolvimento de software também pode sofrer adaptações e extensões para ajustar-se ao tipo de sistema de informação a ser desenvolvido.

Buscar um modelo de representação adequado às características de sistemas de apoio à decisão representa diminuir as chances de que esses sistemas sejam projetados com base em requisitos que não reproduzam as reais necessidades dos usuários.

Este capítulo tem como objetivo, portanto, introduzir os principais conceitos envolvidos na construção de sistemas de informação que serão utilizados como ferramentas de apoio ao processo decisório. São relacionadas algumas das características técnicas desse tipo de sistema com as etapas de um processo de desenvolvimento de software e destacados os artefatos utilizados nessas etapas que melhor se adaptam à representação dos requisitos para a construção desses sistemas, abordando as seguintes tecnologias:

- Sistemas de Informação;
- Data Warehouse;
- Processo de Desenvolvimento de Software (RUP – Rational Unified Process);
- Análise de Requisitos para Sistemas de Informação
- UML

2.1 Sistemas de Informação

2.1.1 Conceitos

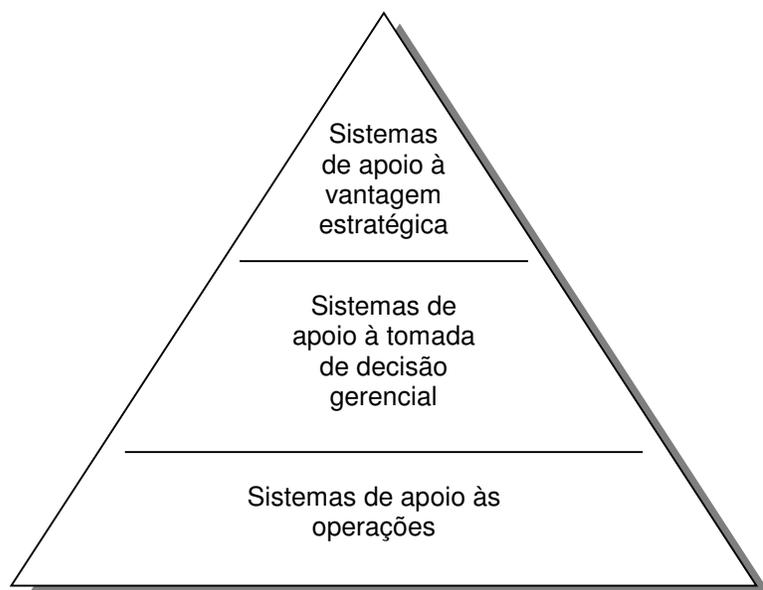
Laudon e Laudon (1996) definem um sistema de informação como um conjunto de componentes inter-relacionados que coletam (ou recuperam), processam, armazenam e distribuem informações para suportar o controle e a tomada de decisão nas organizações.

Campos (1994) afirma que um sistema de informação é uma combinação estruturada de informação, recursos humanos, tecnologia de informação e práticas de trabalho, organizados de forma a assegurar o cumprimento dos objetivos de uma organização.

Para Bio (1985), um sistema de informação é conceituado como sendo um conjunto de partes interdependentes no seu todo que podem ser partes de um todo maior.

Segundo O'Brian (2001), os sistemas de informação desempenham três papéis vitais em qualquer tipo de organização, conforme ilustra a Figura 2.

- Na nível operacional, como suporte de seus processos e operações
- No nível tático, como suporte na tomada de decisões de seus funcionários e gerentes.
- No nível estratégico, como suporte em suas estratégias em busca de vantagem competitiva.



Fonte: O'Brian, 2001.

Figura 2 - Papéis fundamentais dos Sistemas de Informação

2.1.2 Classificação

Os Sistemas de Informação podem ser classificados de acordo com o nível organizacional ao qual atendem:

- Sistemas de Processamento Transacional
- Sistemas de Apoio à Decisão (SAD)
- Sistemas de Informação Executiva (EIS)
- Business Intelligence

Sistemas de Processamento Transacional

Para Sell (2001), os sistemas de informação operacional são utilizados para auxiliar no controle do fluxo das atividades básicas de uma organização, tais como vendas, fluxo de caixa, controle de materiais, entre outros; tais sistemas servem como principal fonte de dados para a construção dos sistemas utilizados nas camadas gerencial e estratégica.

Sistemas de Apoio à Decisão (SAD)

Segundo Polloni (2000), os sistemas de apoio à decisão tratam de assuntos específicos, realizando operações estatísticas, comparativas ou de projeção com os dados da empresa, a fim de identificar novas ações a serem tomadas dentro do negócio da empresa.

O'Brian (2001) afirma que os sistemas de apoio à decisão fornecem apoio aos gerentes durante o processo de tomada de decisão.

De acordo com Rezende e Abreu (2000), os SAD permitem a integração e a coordenação de dados das mais variadas partes, fornecendo informações que sirvam como apoio no processo de tomada de decisões empresariais.

Sistemas de Informação Executiva (EIS)

Um EIS é considerado como um tipo especial de SAD que se destina a apoiar a tomada de decisão da alta administração e são utilizados por executivos de alto nível administrativo como apoio à tomada de decisões estratégicas, fornecendo acesso fácil e imediato a informações sobre os fatores críticos ao sucesso de uma empresa, utilizando recursos gráficos de alta qualidade que podem ser personalizados (LAGO, 2002).

Business Intelligence

Os sistemas de *Business Intelligence* (BI) são a evolução dos EIS e dos SAD. A principal vantagem desses sistemas sobre seus antepassados é a capacidade de cruzar informações de diferentes bancos de dados e analisá-las. Isso possibilita a interpretação da informação de forma contextualizada. Sistemas de BI utilizam hipóteses e procuram relações de causa e efeito, transformando os dados em

informações estratégicas que permitem o direcionamento da atuação da empresa (LAGO, 2002).

A arquitetura de um sistema típico de BI é composta das partes descritas na seqüência, as quais são ilustradas na Figura 3 (WG SYSTEMS, 2003):

- Módulo de ETL (Extração, Transformação e Carga dos Dados): componente dedicado à extração, carga e transformação de dados, responsável pela coleta das informações nas diversas fontes de dados, tais como sistemas de gestão empresarial, arquivos no formato texto (TXT) ou planilhas eletrônicas.
- Data Mart ou Data Warehouse: local onde ficam concentrados todos os dados extraídos dos sistemas operacionais; com um repositório de dados à parte, é possível armazenar informações históricas e agregadas.
- Front end: é a parte visível ao usuário de um projeto de BI, podendo apresentar-se na forma de relatórios padronizados, como um portal de intranet, internet ou extranet, ou como uma ferramenta de análise, de mineração de dados ou de projeções de cenários futuros.



Fonte: WG Systems, 2003

Figura 3 - Arquitetura de um sistema de *Business Intelligence*

2.2 Data Warehouse

2.2.1 Conceito

Ao processo de preparar os dados de um sistema de informação operacional de forma a se ter uma fonte de informações que possam dar suporte à tomada de decisões deu-se o nome de Data Warehousing. Nesse processo, os dados sofrem uma série de transformações, entre elas, a limpeza, a verificação de consistência e a totalização em níveis hierárquicos desejados na análise (KIMBALL et al., 1998).

Em um Data Warehouse (DW) os dados são organizados de maneira a permitir o acesso com um nível de detalhe e uma apresentação mais adequada aos interesses da análise dos dados gerados pelas operações diárias (INMON, 1997).

O conceito de Data Warehouse não se limita a apenas um produto ou um conjunto de produtos, mas também a múltiplos elementos de processamento de dados suportados por diversas tecnologias. Basicamente, ele é composto de:

- bases de dados operacionais espalhadas pela empresa;
- processo de extração, transformação e limpeza dos dados;
- integração dos dados em um modelo lógico, empregando normalmente uma técnica de modelagem denominada *Modelagem Dimensional* e uma representação gráfica conhecida por *cube*;
- espaço de armazenamento que pode ser administrado através de banco de dados relacionais ou multidimensionais;
- ferramentas geradoras de relatórios e de consultas.

Tem-se como resultado de um projeto de Data Warehouse um conjunto de informações que auxiliam no processo de tomada de decisão, permitindo uma melhor visão das curvas de comportamento, uma maior abrangência da visão dos indicadores para análise dos negócios e mais segurança na tomada de decisões (MACHADO, 2000).

Inmon (1997) define o Data Warehouse como sendo um conjunto de dados integrado, não volátil e variável em relação ao tempo. Esses dados são organizados em assuntos e a análise desses dados apóia a tomada de decisões gerenciais.

No passado as empresas utilizavam o próprio ambiente de dados transacionais em seus processos de apoio decisório. O uso concorrente, entretanto, do ambiente de processamento de transações para a contabilização de grandes volumes de dados gerava conflitos entre os sistemas analíticos e transacionais. Sistemas transacionais são focados em informações operacionais e não gerenciais, possuem um limitado foco de requisitos funcionais e não são otimizados para a totalização de dados. Nos últimos anos tem havido um crescimento no número de aplicações práticas das tecnologias de Data Warehouse e alterações constantes nessa

arquitetura têm permitido o rápido aumento do volume de dados acessados (INMON, 2000).

2.2.2 Objetivos

De acordo com Kimball (2002), os principais objetivos de um Data Warehouse são:

- permitir fácil acesso à informação, com conteúdo intuitivo e significado óbvio das informações para os usuários utilizando-se de correta nomenclatura e total legibilidade para as informações armazenadas;
- apresentar informações consistentes e confiáveis, que, mesmo tendo sido originadas de diversas fontes de dados, tenham passado por um processo de limpeza e verificação de qualidade, sendo disponibilizadas para os usuários somente quando estiverem apropriadas para o consumo;
- ser adaptável a mudanças, originadas de alterações de necessidades de usuários, de condições de negócio, de dados ou da própria tecnologia;
- garantir a segurança no que se refere ao acesso às informações armazenadas no Data Warehouse, que muitas vezes são consideradas confidenciais;
- armazenar dados que servirão de base para o processo de tomada de decisões;
- ser aceitável pela comunidade de usuários, pois, uma vez que difere de um sistema de informação operacionais cujo uso torna-se praticamente obrigatório, o DW é uma ferramenta opcional.

2.2.3 Composição

Kimball e Ross (2002) descrevem um Data Warehouse como uma estrutura formada pelos seguintes componentes:

- Fontes de Dados Operacionais (Operational Source Systems);
- Área de Preparo de Dados (Data Staging Area);
- Área de Apresentação de Dados (Data Presentation Area);
- Ferramentas de Acesso a Dados (Data Access Tools).

Fontes de Dados Operacionais (*Operational Source Systems*)

São os sistemas que capturam as transações de negócio, e as consultas nesses sistemas são normalmente direcionadas para um único registro de informação. Usualmente esses sistemas mantêm um pequeno conjunto de dados históricos e são considerados como uma porção externa ao DW, uma vez que não se tem nenhum ou quase nenhum controle sobre o conteúdo e a formatação dos dados nesses sistemas legados operacionais.

Área de Preparo de Dados (*Data Staging Area*)

É a área composta de dados originados das fontes externas ao DW e de um conjunto de processos usualmente denominados ETL (*extract + transformation + load* – extração + transformação + carga). Um dos requisitos de arquitetura da área de preparo de dados é que esta esteja fora dos limites de uso dos usuários de negócio e que não proveja consultas e serviços de apresentação de dados. Durante o processo de extração os dados são lidos das fontes externas e compreendidos. Os dados necessários para o DW são então copiados para a área de preparo para futuras manipulações. Após a extração, o processo de transformação executa algumas operações, tais como limpeza dos dados, correções de escrita, conflitos de domínio dos dados, tratamento de elementos ausentes e transformações para formatos padronizados. A carga do DW usualmente utiliza-se de uma estrutura de

dados dimensional. Somente após a carga dos dados, a indexação, a criação de agregados apropriados e a qualidade garantida desses dados é que estes são disponibilizados para a comunidade de usuários, a qual é informada sobre todas as alterações ocorridas nas dimensões ou sobre os novos conceitos introduzidos nos fatos medidos ou calculados.

Área de Apresentação de Dados (*Data Presentation Area*)

É a área na qual os dados ficam organizados, armazenados e disponíveis diretamente para as consultas dos usuários. Esta é a área que realmente estará visível aos usuários de negócio. A área de apresentação usualmente fica dividida em diversos Data Marts (subconjuntos de um Data Warehouse), sendo cada um designado a tratar dos diferentes processos de negócio da empresa.

Ferramentas de Acesso a Dados (*Data Access Tools*)

São as ferramentas e funcionalidades disponibilizadas aos usuários para consultarem os dados armazenados na área de apresentação, podendo variar desde ferramentas de uso mais simples até mais sofisticado, como as utilizadas para mineração de dados, com o intuito de encontrar relações entre os dados que levam a informações úteis, não triviais, implícitas e previamente desconhecidas. Segundo Thomsen (2002), quando aplicadas a bases de dados dimensionais, essas ferramentas executam operações de agregação nos dados com o objetivo de gerar somatórios e contagens sobre uma determinada combinação de dimensões disponíveis na base de dados.

Atualmente não é comum utilizar o mesmo ambiente para suportar ambos os tipos de processamento quando grandes volumes de dados são tratados. Mello (2002), entretanto, descreve uma proposta alternativa de um modelo de dados

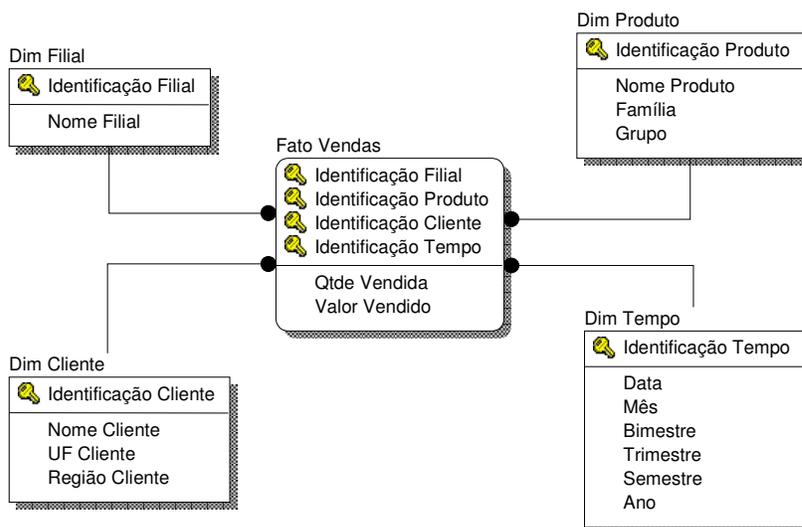
capaz de suportar tanto o processamento transacional quanto o processamento informacional através da eliminação da camada de preparo de dados (*Data Staging Area*) e da unificação dos ambientes transacional e informacional.

2.2.4 Modelagem dimensional

Usualmente, os sistemas de informação utilizados no cotidiano de uma empresa são modelados através de técnicas que permitem respostas rápidas a consultas que requerem um pequeno volume de dados. Com isso, esses sistemas armazenam dados em um formato relacional normalizado, sendo esse modelo constituído basicamente por um conjunto de tabelas e pela definição de relacionamentos entre elas.

O modelo relacional normalizado, entretanto, não é totalmente adequado para atender a consultas que acessem um grande volume de dados a serem sumarizados. Neste caso, torna-se necessário que esses dados passem por um processo de transformação.

Kimball e Ross (2002) definem a modelagem dimensional como uma técnica estruturada desenvolvida para a obtenção de um modelo de dados dimensional. Em sua composição típica esse modelo possui uma grande entidade central (fato) que reflete a evolução dos negócios do dia-a-dia de uma organização e um conjunto de entidades menores (dimensões) arranjadas ao redor da entidade central e utilizadas de forma combinada como variáveis de análise do fato, como ilustrado na Figura 4.



Fonte: do autor.

Figura 4 - Exemplo de um modelo dimensional com tabelas de fato e dimensões

Fatos

São tabelas que armazenam medidas numéricas totalizáveis relacionadas a um determinado assunto ou processo de negócio. Cada registro em uma tabela de fato está relacionado a um conjunto de dimensões que determinam a granularidade dos fatos armazenados e definem qual o escopo dessas medidas. Quanto menor a granularidade de um fato, maior será o nível de detalhe armazenado (KIMBALL; ROSS, 2002).

Dimensões

Tabelas de dimensões armazenam descrições textuais de um processo de negócio, podendo cada uma ser formada por um conjunto de atributos, denominados hierarquias, que servem como base para determinar regras de agrupamento, quebras e filtros em consultas a uma tabela de fato (KIMBALL; ROSS, 2002).

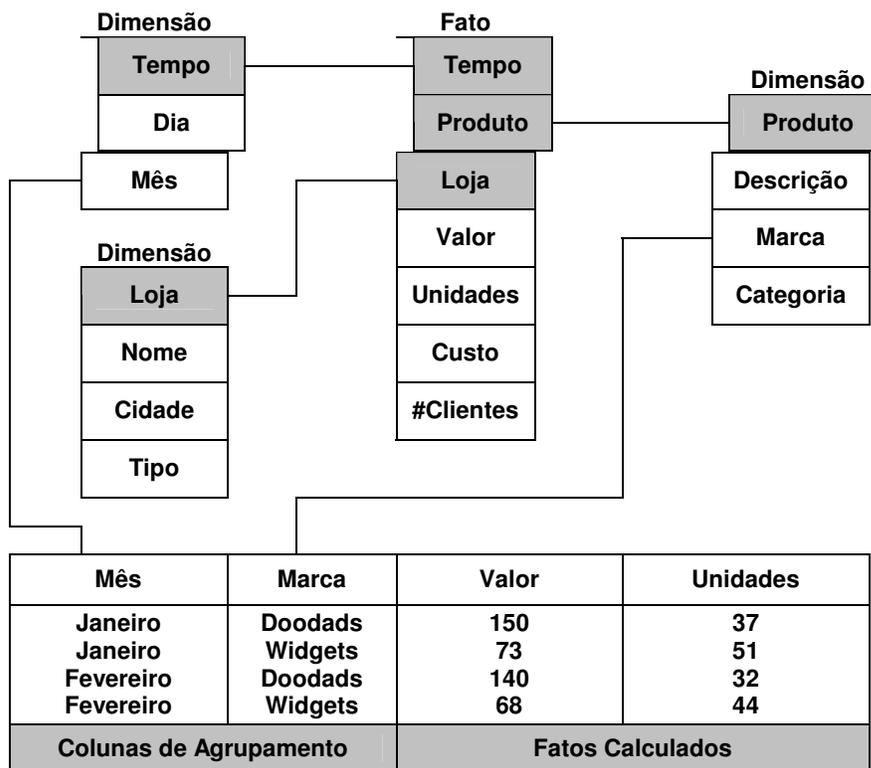
2.2.5 Operações OLAP

Sob o aspecto funcional, as ferramentas OLAP (*On-Line Analytical Processing*), utilizadas para acessar dados em uma base de dados dimensional, disponibilizam operações para seus usuários, sendo as principais (KIMBALL, 1996):

- *Drill Down*
- *Drill Up*
- *Drill Across*

Drill Down

Esta operação pode ser descrita como a maneira de "obter mais detalhes". Quem atua como principal ator neste tipo de operação são os atributos das dimensões, que servem como agrupadores para a totalização dos resultados de uma consulta. Dessa forma pode-se descrever a operação *drill down* como sendo uma operação em que uma nova coluna de agrupamento é adicionada a uma consulta, como ilustrado na Figura 5. Essa operação pode ocorrer de forma automática quando os atributos de detalhamento e os níveis hierárquicos de detalhamento são pré-definidos na aplicação, ou de forma configurável, em que o próprio usuário seleciona qual o atributo deseja adicionar à sua consulta para obter resultados mais detalhados.



Fonte: adaptado de Kimball (1996).

Figura 5 - Atributos podem ser adicionados a uma consulta para detalhar o resultado

Drill Up

Enquanto a operação *drill down* adiciona atributos das dimensões para o detalhamento de resultados, a operação *drill up* retira do resultado essas colunas de agrupamento, não necessariamente na mesma ordem em que estas foram adicionadas, gerando sumarizações com menor nível de detalhamento.

Drill Across

Enquanto uma operação *drill down* obtém dados mais detalhados de uma mesma tabela de fato, a operação *drill across* realiza uma mudança na tabela de fatos a ser analisada, mantendo o mesmo nível de granularidade e detalhamento da consulta. Diversas tabelas de fato, relacionadas em uma seqüência de processos de negócio interligados, podem compartilhar das mesmas dimensões, permitindo a execução da operação *drill across* entre elas.

Nem todos os processos de negócio podem ser modelados em uma única tabela de fatos, principalmente por dois motivos: a) esses fatos podem não compartilhar exatamente as mesmas dimensões; e b) cada tabela de fatos processa fatos distintos, que podem ser gerados em diferentes momentos durante todo o processo.

Outras operações OLAP

Operações OLAP podem alterar o conteúdo ou a forma como as informações são utilizadas na montagem de um resultado - como *filter* para determinar condições a serem aplicadas, ou *sort* para definir como será feita a ordenação dessas informações.

2.3 Rational Unified Process (RUP)

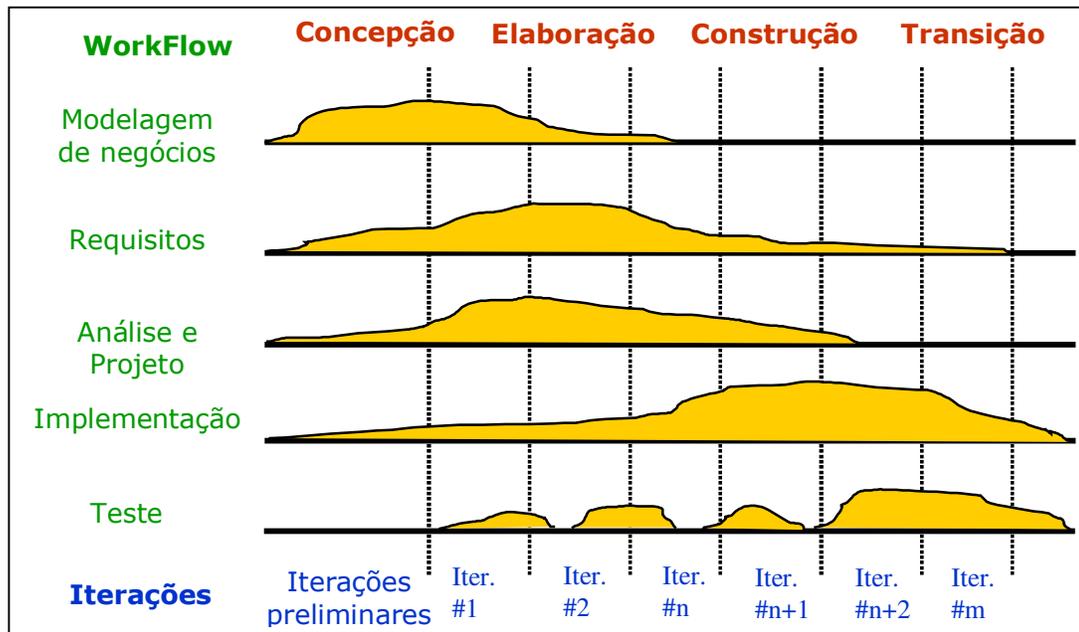
2.3.1 Conceito

O *Rational Unified Process*® (RUP) é um processo de engenharia de *software* baseado no *Unified Process*, que provê uma forma disciplinada de designar tarefas e responsabilidades dentro de uma empresa ou um departamento responsável por desenvolvimento de *software*, com o objetivo de garantir a produção de *software* de alta qualidade que vá ao encontro das necessidades dos usuários finais do produto dentro de um prazo e orçamento previstos (RATIONAL, 2002).

A Figura 6 ilustra a completa arquitetura do RUP, mostrando as diferentes ênfases ao longo do tempo e descreve duas dimensões principais:

- o eixo horizontal representa o tempo, exibindo aspectos dinâmicos do ciclo de vida do processo de desenvolvimento, expressos em termos de fases, iterações e pontos de checagem; e

- o eixo vertical representa disciplinas que agrupam atividades de forma lógica por natureza, retratando o aspecto estático do processo, descrito em termos de componentes, disciplinas, atividades, fluxos de trabalho, artefatos e papéis (funções).



Fonte: Rational (2002).

Figura 6 - Algumas disciplinas, fases e iterações do processo RUP

2.3.2 Modelagem de negócio

Os objetivos da modelagem de negócio no processo RUP são (RATIONAL, 2002):

- compreender a estrutura e as dinâmicas da organização para a qual um sistema será desenvolvido;
- identificar os atuais problemas dessa organização e possíveis melhorias nos processos;

- garantir que todos os envolvidos no processo de desenvolvimento de um sistema, tanto clientes, usuários finais e os próprios desenvolvedores, tenham o mesmo entendimento sobre a organização;
- derivar da modelagem de negócios os requisitos de sistema que devem ser suportados para a organização.

A modelagem de negócios provê um guia de como um analista pode compreender e decifrar visualmente um negócio ou uma área de negócio, que será útil para o desenvolvimento de um modelo de casos de uso de negócio, um modelo de objetos de negócio e um modelo de domínio para o contexto organizacional.

O Quadro 1 mostra as principais tarefas e artefatos gerados com as atividades desempenhadas durante a disciplina de modelagem de negócios.

Atividades	Tarefas	Artefatos gerados
<ul style="list-style-type: none"> • Avaliar e descrever a situação atual do negócio • Identificar, definir e projetar a realização de processos de negócio • Identificar atribuições e responsabilidades • Explorar a automação de processos • Desenvolver um modelo de domínio 	<ul style="list-style-type: none"> • Capturar o vocabulário de negócio comum • Identificar e manter regras de negócio • Definir e ajustar objetivos • Desenvolver um direcionamento para o modelo de negócio • Identificar atores e casos de uso • Identificar entidades de negócio • Detalhar atores, casos de uso e entidades de negócio • Definir requisitos de automação 	<ul style="list-style-type: none"> • Visão de negócio • Glossário de negócio • Regras de negócio • Modelo de casos de uso de negócio • Casos de uso de negócio • Especificação de negócio complementar • Modelo de objetos e entidades de negócio • Unidades organizacionais • Modelo de análise

Fonte: Rational (2002).

Quadro 1 - Atividades, tarefas e artefatos da disciplina de modelagem de negócios

2.3.3 Análise de requisitos

Dentro do processo RUP, Rational (2002) estabelece que a análise de requisitos possui os seguintes propósitos:

- estabelecer e manter um acordo com os usuários e patrocinadores do sistema com relação ao que o sistema deve fazer;
- prover os desenvolvedores do sistema com uma melhor compreensão dos seus requisitos;
- definir os limites e as responsabilidades do sistema;
- prover uma base para o planejamento das etapas e iterações subseqüentes, bem como para a estimativa de custos e tempo de desenvolvimento do sistema.

A disciplina de análise de requisitos envolve atividades como encontrar, manter e gerenciar requisitos para um sistema computacional. Para descrever completamente um sistema, identificando o que este deve realizar, em um esforço que posiciona os seus usuários como importantes fontes de informação, Rational (2002) propõe a construção de:

- modelo de casos de uso: modelo que descreve as funções desejáveis do sistema e de seu ambiente, servindo como entrada essencial para as atividades de análise, projeto e teste do sistema;
- especificação complementar: captura os requisitos do sistema que não são diretamente identificados nos casos de uso do modelo, como requisitos legais e regulamentadores, padrões de aplicação, atributos de qualidade do sistema, ambiente e sistema operacional, requisitos de compatibilidade e regras de projeto.

O modelo de casos de uso consiste de um conjunto de atores e casos de uso, sendo estes descritos em detalhes suficientes para mostrar passo a passo como o sistema deve interagir com os atores.

A especificação complementar é importante ao modelo de casos de uso, pois juntos capturam todos os requisitos de software (funcionais e não funcionais) que precisam ser descritos para uma completa especificação de requisitos do sistema.

Requisitos funcionais detalham ações que o sistema deve estar apto a realizar, sem levar em consideração restrições físicas, especificando, dessa forma, entradas e saídas de um sistema. Requisitos que não são funcionais descrevem atributos do sistema e do seu respectivo ambiente.

De forma complementar aos artefatos mencionados, é também desenvolvido um glossário para definir a terminologia comum a ser usada de forma consistente durante o projeto.

A análise de requisitos está relacionada com a fase de análise e projeto, que usa o modelo de casos de uso e o glossário como sua entrada principal.

O Quadro 2 mostra as principais tarefas e artefatos gerados com as atividades desempenhadas durante a disciplina de análise de requisitos.

Atividades	Tarefas	Artefatos gerados
<ul style="list-style-type: none"> • Analisar o problema • Compreender as necessidades • Definir e gerenciar o escopo do sistema • Gerenciar as modificações nos requisitos 	<ul style="list-style-type: none"> • Capturar o vocabulário comum • Desenvolver uma visão • Encontrar atores e casos de uso • Desenvolver um plano de gerenciamento de requisitos • Definir e gerenciar solicitações dos usuários • Gerenciar dependências • Priorizar e detalhar casos de uso • Detalhar requisitos de software • Modelar e prototipar a interface do usuário • Estruturar modelo de casos de uso 	<ul style="list-style-type: none"> • Glossário • Visão geral do sistema • Modelo de casos de uso • Plano de gerenciamento de requisitos • Características dos requisitos • Solicitações dos usuários • Especificação complementar • Especificação de requisitos de software • Enredo de casos de uso e casos de uso detalhados • Protótipo da interface

Fonte: Rational (2002).

Quadro 2 - Atividades, tarefas e artefatos da disciplina de análise de requisitos

2.3.4 Análise e projeto

Os principais propósitos da disciplina de análise e projeto são:

- transformar os requisitos em um projeto do sistema a ser construído;
- desenvolver uma arquitetura para o sistema;
- adaptar o projeto ao ambiente de implementação, adequando-o para atender a requisitos de performance.

O Quadro 3 mostra as principais tarefas e artefatos gerados com as atividades desempenhadas durante a disciplina de análise e projeto.

Atividades	Tarefas	Artefatos gerados
<ul style="list-style-type: none"> • Definir e avaliar a viabilidade da arquitetura • Analisar comportamento • Projetar componentes • Projetar a base de dados 	<ul style="list-style-type: none"> • Analisar e descrever a arquitetura de execução • Análisar casos de uso • Identificar elementos de projeto • Projetar e priorizar casos de uso • Projetar classes e pacotes • Projetar subsistemas e base de dados • Implementar componentes • Identificar mecanismos e elementos de projeto, incorporando elementos já existentes • Estruturar o modelo de implementação 	<ul style="list-style-type: none"> • Documento de arquitetura de software • Modelo do projeto • Realização de casos de uso • Lista de riscos • Visão do projeto • Pacotes de componentes • Especificação de teste • Especificação complementar • Projeto de classes e dos subsistemas • Interface • Modelo de dados

Fonte: Rational (2002).

Quadro 3 - Atividades, tarefas e artefatos da disciplina de análise e projeto

2.3.5 Implementação

O propósito da disciplina de implementação é:

- definir a organização do código em termos de subsistemas organizados em camadas; e

- implementar classes e objetos em relação a artefatos de implementação (código-fonte, executável, binário, etc.).

O Quadro 4 mostra as principais tarefas e artefatos gerados com as atividades desempenhadas durante a disciplina de implementação.

Atividades	Tarefas	Artefatos gerados
<ul style="list-style-type: none"> • Estruturar o modelo de implementação • Planejar a integração • Implementar componentes • Integrar os subsistemas e o sistema 	<ul style="list-style-type: none"> • Estruturar o modelo de implementação • Planejar a integração do sistema e subsistemas • Corrigir defeitos • Implementar componentes • Revisar código • Realizar testes de unidades • Integrar os subsistemas e o sistema 	<ul style="list-style-type: none"> • Documento de arquitetura de software • Modelo de implementação do sistema e subsistemas • Plano de construção e integração • Componentes • Testes de componentes • Software implementado

Fonte: Rational (2002).

Quadro 4 - Atividades, tarefas e artefatos da disciplina de implementação

2.4 Análise de requisitos para Sistemas de Informação

2.4.1 Conceitos

No contexto de software, uma especificação pode ser representada por um documento escrito, um modelo gráfico, um modelo matemático formal, uma coleção de casos de uso, um protótipo ou qualquer combinação dos itens citados. A especificação do sistema é o produto final produzido pelos engenheiros de requisitos, sendo usada como a base para as engenharias de hardware, software e banco de dados, pois descreve funções, regras e performance requeridas de um sistema, bem como a informação (dados e controle) que será usada como entrada e saída do sistema (KOTONYA; SOMMERVILLE, 1998).

Segundo Jacobson, Booch e Rumbaugh (1999), a análise de requisitos é um processo que permite descobrir o que deve ser exatamente projetado e construído para um sistema de informação.

A análise de requisitos está inserida dentro do contexto de um processo de desenvolvimento de software como uma das principais etapas, pois é a etapa que irá direcionar todos os demais esforços a serem despendidos na construção dentro do processo de desenvolvimento de software, determinando as necessidades e especificando funcionalidades às quais o sistema deve atender.

O *Unified Process* (Processo Unificado) é descrito por Jacobson, Booch e Rumbaugh (1999) como uma proposta de um processo para o desenvolvimento de sistemas de informação. Processos de desenvolvimento de software podem variar de acordo com diferentes contextos, tipos de sistemas, podendo ser adaptados às necessidades de um projeto ou de uma empresa. Tendo sido projetado para ser especializado, o *Unified Process* pode adequar-se a diferentes necessidades apresentadas em um processo de construção de um Data Warehouse. Alguns dos principais fatores que influenciam na maneira com que os processos de desenvolvimento podem variar são descritos a seguir.

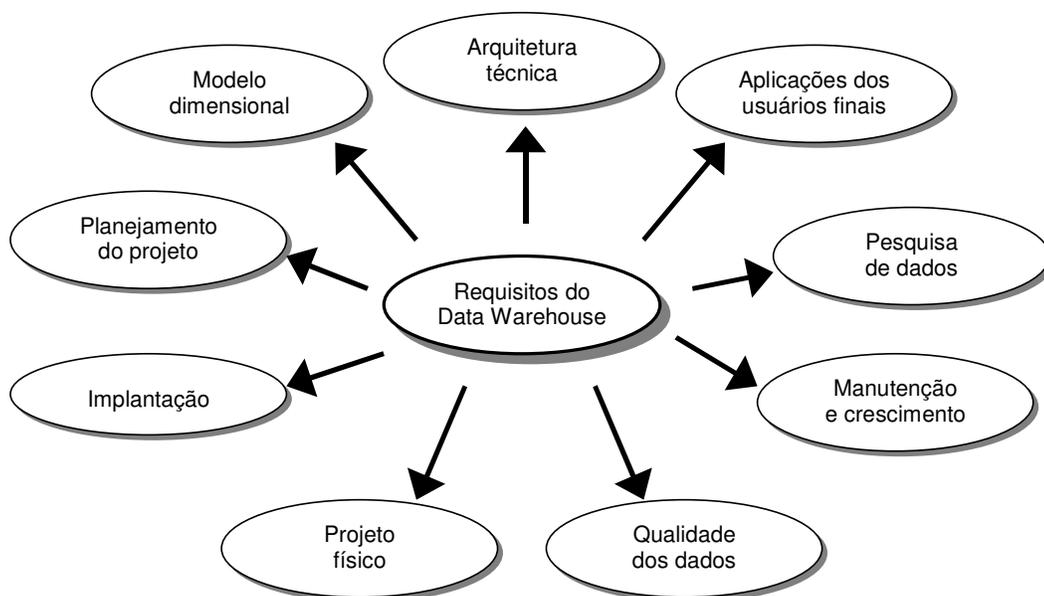
- Fatores organizacionais, como a estrutura e a cultura organizacional, a forma de gerenciamento de projetos, as diferentes habilidades e competências disponíveis, experiências anteriores e até mesmo os sistemas computacionais já existentes.
- O domínio do problema, como o escopo da aplicação, os diferentes processos de negócio a serem suportados e a própria comunidade de usuários.
- Fatores relacionados ao ciclo de vida, incluindo a tecnologia existente, a experiência no desenvolvimento de software da equipe técnica e o planejamento de futuras versões do sistema.

- Fatores técnicos, entre os quais linguagens de programação, ferramentas de desenvolvimento, bases de dados e arquiteturas computacionais.

Como efeito desses fatores, um processo de desenvolvimento pode ser adaptado removendo, estendendo ou criando uma nova conotação ou estrutura para os artefatos.

2.4.2 Análise de requisitos para projetos de Data Warehouse

Segundo Kimball et al. (1998), o objetivo de apurar os requisitos de um Data Warehouse é prover uma especificação compreensível dos principais aspectos técnicos e permitir que a implementação do sistema se complete com sucesso. Os aspectos técnicos do sistema diretamente afetados pelos requisitos do DW estão ilustrados na Figura 7.



Fonte: adaptado de Kimball et al., 1998.

Figura 7 - Impacto de requisitos em um Data Warehouse

Por envolver usualmente mais de uma das unidades organizacionais de uma empresa, a construção de um DW pode tornar-se uma tarefa desafiadora. Além de servir como suporte para o processo de tomada de decisões, o DW é também utilizado para armazenar dados históricos, limpos, validados, e sumarizados, que permitem explorar e analisar um processo de negócio de forma compreensível e flexível (BRUCKNER; LIST; SCHIEFER, 2001).

Um DW deve prover um sistema de suporte à decisão, direcionando os usuários de todos os níveis hierárquicos da organização que necessitam de diferentes níveis de granularidade ou detalhamento para contemplarem suas necessidades específicas. Numa estrutura organizacional, diferentes requisitos de processamento e análise de dados podem ser identificados em níveis hierárquicos distintos, como, por exemplo, nos níveis hierárquicos mais altos, em que são necessários dados com menor detalhamento e maior granularidade do que nos níveis hierárquicos mais baixos. Sistemas de suporte à decisão, isolados e localizados em cada nível hierárquico ou em cada departamento, não são uma solução desejável. Ao serem apurados e combinados os requisitos de usuários de todos os níveis hierárquicos, será possível modelar um sistema de suporte à decisão para toda a organização e criar um ambiente que disponibilize dados coerentes e a integração funcional das aplicações que acessam esses dados. As necessidades decorrentes desse processo de análise são contempladas com visões dimensionais dos dados, e o projeto de um sistema de DW é muito dependente de requisitos que contemplem a geração dessas visões. Projetos de DW podem freqüentemente falhar quando são construídos a partir de uma má compreensão dos requisitos de análise (LIST; SCHIEFER; TJOA, 2001).

Durante o processo de definição de requisitos, os analistas de sistemas trabalham junto com os patrocinadores do projeto e com os usuários para descrever esses requisitos. Não é pouco freqüente, entretanto, que a equipe de desenvolvimento tenha dificuldades de compreender a terminologia de negócio utilizada, encontrando descrições pouco formais para utilizar na implementação do sistema. Ao mesmo tempo, ao apresentar uma especificação técnica do sistema para os usuários, estes podem ter semelhantes dificuldades de compreensão, sendo forçados a aceitar o projeto por receio de impor obstáculos ao desenvolvimento. Pelo fato de os diversos perfis de profissionais envolvidos (analistas, desenvolvedores e usuários) não utilizarem a mesma linguagem, essa dificuldade de comunicação pode fazer com que um sistema DW não alcance os requisitos inicialmente definidos. Quando as especificações técnicas mal-elaboradas dificultam ou tornam o uso do DW impraticável, este passa a não gerar o efeito esperado pela empresa, e os departamentos passam a criar soluções isoladas para atender às suas próprias necessidades. Nessa situação, tornam-se impossíveis análises no escopo da organização (BRUCKNER; LIST; SCHIEFER, 2001).

Modelar um DW de modo que seu projeto seja simultaneamente preciso e amigável é um desafio, visto que tal projeto deve utilizar uma simbologia com uma especificação clara de cada símbolo componente empregado. Por outro lado, cada um desses símbolos deve ter uma semântica definida e compreensão intuitiva para os usuários. Para satisfazer todas as partes envolvidas em um projeto de DW, uma especificação compreensível e intuitiva é necessária e pode ser alcançada, por exemplo, através da utilização de casos de uso. Além de fazerem parte da UML, que é uma notação padrão e geralmente aceita como tal para especificação de software,

os casos de uso podem ser utilizados em um escopo mais genérico, em que são suprimidos detalhes técnicos de implementação (LIST; SCHIEFER; TJOA, 2001).

De acordo com Bruckner, List e Schiefer (2001), ao mesmo tempo em que um conjunto de requisitos pode determinar comportamentos funcionais – como definir quais os dados que podem ser acessados e com que nível de segurança, como os dados são transformados e organizados ou como são calculados e agregados –, requisitos podem estabelecer um direcionamento e definir expectativas do patrocinador do projeto, os quais permitirão melhorar seus negócios. Por esse motivo, requisitos para um DW podem ser divididos em diferentes níveis de abstração, como os ilustrados na Figura 8.

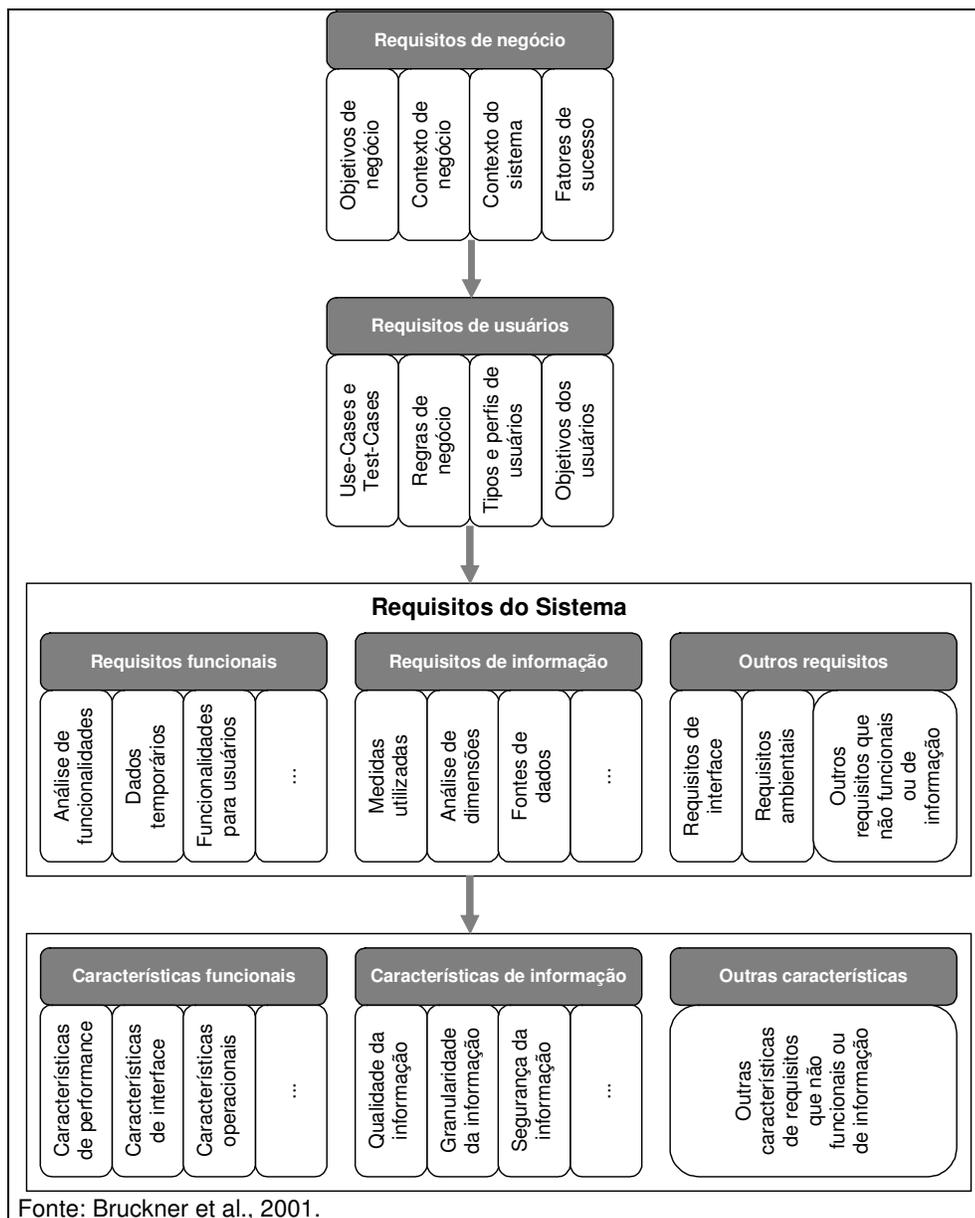


Figura 8 - Níveis de abstração de requisitos em um Data Warehouse

Requisitos de negócio

Requisitos de negócio estão no mais alto nível de abstração da cadeia de requisitos, expressam os mais significativos objetivos de uma organização em relação ao projeto do DW e identificam os principais benefícios que o sistema irá prover para a empresa e para os usuários, podendo expressar até mesmo novas oportunidades de negócio. São usualmente capturados em um documento que

descreve a visão e o escopo do projeto e podem conter ainda um diagrama contextual do sistema, além de outras informações como:

- objetivos de negócio;
- contexto de negócio;
- contexto do sistema; e
- fatores de sucesso para o projeto.

Requisitos de usuários

Este conjunto de requisitos descreve as tarefas que os usuários estarão aptos a realizar com o auxílio do DW, sendo coletados de pessoas que realmente irão trabalhar e utilizar o sistema, e que poderão descrever tanto as tarefas que necessitam realizar com requisitos não funcionais importantes para a aceitação do DW.

Requisitos de usuários são capturados através de casos de uso ou da descrição de cenários e confrontados com o contexto e os objetivos estabelecidos pelos requisitos de negócio, descrevendo o que os usuários necessitam realizar com o DW, ao contrário do enfoque tradicional utilizado para avaliar o que os usuários esperam que um sistema de informação faça.

A descrição de requisitos de usuários pode conter:

- casos de uso e casos de teste;
- regras de negócio;
- tipos e perfis de usuários; e
- objetivos dos usuários.

Requisitos do sistema

Devem estar de acordo com os requisitos de negócio e com os requisitos de usuários, representando uma importante fonte de informação para a equipe de desenvolvimento e sendo exibidos em um nível bastante expressivo de detalhamento.

Requisitos funcionais do sistema definem funcionalidades que devem ser criadas pela equipe de desenvolvimento para permitir que usuários cumpram suas tarefas ou para satisfazer requisitos de negócio, definindo o comportamento do sistema, expresso na forma de serviços ou funções que o sistema deve realizar, as funcionalidades a serem disponibilizadas aos usuários e a especificação de *front end*.

Requisitos de informação do sistema definem e descrevem as informações e dados necessários para a empresa, os quais devem ser disponibilizados pelo DW. A especificação desses dados ainda descrever a qualidade dos dados, as fontes de origem, como eles devem ser processados e em que métodos de análise serão utilizados.

Outros requisitos, que não sejam funcionais ou de informação, podem ser especificados para descrever outros aspectos relevantes do sistema, tais como interface entre diferentes sistemas ou requisitos ambientais, cultura, políticas ou aspectos legais.

Características dos requisitos

São propriedades ou qualidades do DW que complementam as definições dos requisitos funcionais e de informação, descrevendo características sob vários aspectos, importantes para os usuários ou para a equipe de desenvolvimento.

Incluem definições de padrões, condições, interfaces externas, requisitos de performance, regras de implementação e projeto, atributos de qualidade e granularidade dos dados.

2.4.3 Gerenciamento de requisitos

É necessário persistir nas alterações de requisitos através de todo o ciclo de vida de um sistema. Nesse sentido, o gerenciamento de requisitos é o processo que gerencia mudanças nos requisitos de um sistema, que evoluem devido a mudanças no seu ambiente e conforme os clientes ou usuários desenvolvem um melhor entendimento de suas reais necessidades. Novos requisitos podem surgir, ou os requisitos já apurados podem sofrer alterações em todos os estágios do processo de desenvolvimento do sistema, devido a erros durante o processo de engenharia de requisitos, erros de projeto ou problemas de implementação. As principais preocupações durante o processo de gerenciamento de requisitos são (KOTONYA; SOMMERVILLE, 1998):

- gerenciar mudanças nos requisitos acordados;
- gerenciar os relacionamentos entre os requisitos;
- gerenciar as dependências entre o documento de requisitos e outros documentos produzidos ao longo do sistema e do processo de engenharia de software.

Kotonya e Sommerville (1998) destacam ainda a importância da rastreabilidade no processo de gerenciamento de requisitos. Um requisito é rastreável se for possível identificar quem o solicitou, porque existe, quais os requisitos relacionados e como os requisitos se relacionam a outras informações, tais como a partes do

projeto do sistema, implementações e documentos do usuário, que podem ser utilizados para identificar todos os requisitos afetados por mudanças propostas.

2.5 UML

Processos de desenvolvimento de software devem ser suportados por ferramentas automatizadas. Essa relação entre ferramentas e processo é determinante de maneira que a escolha das ferramentas gera impacto direto no processo, ao mesmo tempo que a definição de um processo define o tipo de ferramenta a ser utilizada. Uma ferramenta de modelagem não implica um processo de desenvolvimento, mas quando esse processo é utilizado, a ferramenta deve suportá-lo. A UML (*Unified Modeling Language*), como uma linguagem visual padrão, pode ser empregada como uma das ferramentas em um processo de desenvolvimento; entretanto, é apenas uma linguagem de modelagem e não define uma forma específica de como deve ser utilizada para a construção de um sistema computacional (JACOBSON; BOOCH; RUMBAUGH, 1999).

A UML é uma linguagem de modelagem criada com o objetivo de se tornar um padrão para desenhar diagramas de projetos, principalmente de sistemas orientados a objetos. Sua notação gráfica tornou-a popular a partir do final dos anos oitenta e início dos anos noventa (FOWLER; SCOTT, 2000).

Segundo Alhir (1998), a UML é uma linguagem unificada de modelagem que reuniu as melhores práticas para modelagem de sistemas; ao definir uma notação para projetar diagramas, a UML tornou-se robusta e flexível para capturar conhecimentos sobre determinado assunto e para expressá-los em um formato padronizado.

Originando-se da unificação dos métodos de análise orientada a objetos desenvolvidos por Booch, Rumbaugh (OMT) e Jacobson (OOSE), a UML foi

estabelecida em 1997 como um padrão pela *Object Management Group* (OMG) (FOWLER; SCOTT, 2000).

De acordo com Furlan (1998), a UML não representa uma ruptura radical de Booch, OMT ou OOSE, e pode ser considerada como um passo na escala de evolução, cujos objetivos são:

- fornecer aos usuários uma linguagem de modelagem visual expressiva, pronta para o uso e que permita o desenvolvimento de modelos de negócio;
- fornecer mecanismos de extensão e especialização para apoiar os conceitos essenciais, permitindo somar novos conceitos e notações para os assuntos não cobertos e especializar conceitos, notações e restrições para domínios de aplicação particulares;
- ser independente de uma linguagem de programação ou de um processo de desenvolvimento;
- prover uma base formal para entender a linguagem de modelagem, sem a necessidade de níveis sofisticados de notação especializada, como a notação matemática de baixo nível ou as definições operacionais que sejam equivalentes à programação;
- encorajar o crescimento do número de ferramentas orientadas a objetos;
- suportar conceitos de desenvolvimento de nível mais elevado, como colaboração, estrutura de trabalho, padrões e componentes;
- integrar as melhores práticas do mercado, incluindo visões amplamente variadas baseadas em níveis de abstração, domínios, arquiteturas, estágios de ciclo de vida, tecnologias de implementação, entre outros, para ajudar a desmistificar o ciclo de modelagem.

Através de um conjunto de diagramas, a UML permite especificar, visualizar e documentar sistemas e processos, independentemente de linguagem de programação ou processo de desenvolvimento que venha a ser utilizado. Um diagrama é a representação gráfica de um conjunto de elementos, geralmente representados como gráficos de vértices (itens) e arcos (relacionamentos). Os diagramas representam uma visão parcial dos elementos que compõem o sistema e podem conter qualquer combinação de itens e de relacionamentos (BOOCH; RUMBAUGH; JACOBSON, 2000).

Segundo Furlan, (1998), a UML é formada por um conjunto de tipos distintos de diagramas, apresentados a seguir.

- Diagramas de casos de uso: descrevem funcionalidades percebidas por atores externos (usuário, um dispositivo ou outro sistema) quando interagem com um sistema.
- Diagramas de classes: denotam a estrutura estática de um sistema e as classes que representam a estrutura das informações por ele manipuladas.
- Diagramas de objetos: são variações dos diagramas de classe que mostram o número de instâncias dessas classes, sendo úteis para exemplificar diagramas de classe complexos.
- Diagramas de estados: mostram as possíveis seqüências de estados que um objeto assume em sua vida em resposta a estímulos recebidos; são considerados complementares aos diagramas de classes, pois relacionam os possíveis estados que objetos da classe podem ter e quais eventos podem causar a transição de um estado para outro.
- Diagramas de atividades: utilizados para descrever os fluxos dirigidos por processamento interno e as atividades realizadas em uma operação.

- Diagramas de seqüência: representam a interação de tempo dos objetos, mostrando a seqüência de mensagens enviadas entre eles através de duas dimensões: uma vertical, para representar o tempo, e uma horizontal, para representar os objetos.
- Diagramas de colaboração: desenhados na forma de um diagrama de objetos, mostram, através de números para evidenciar a seqüência de mensagens, a interação dinâmica de um caso de uso organizada em torno de objetos e seus vínculos.
- Diagramas de componentes: mostram dependências entre componentes de software, inclusive códigos-fonte, binário ou executável.
- Diagramas de implantação: ilustram os elementos de configuração de processamento e seus componentes de software, processos e objetos, incluindo o uso físico do sistema, como computadores, dispositivos e suas conexões.

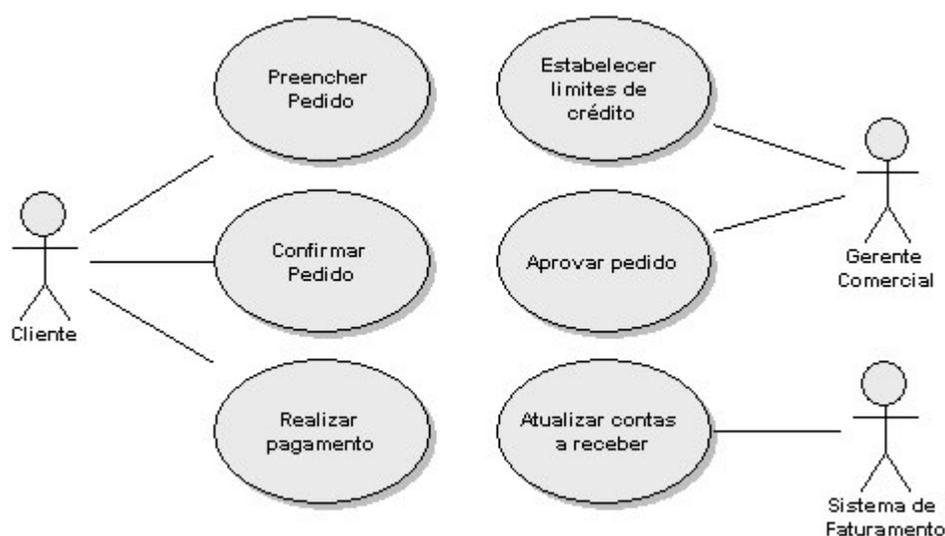
Por estarem mais relacionados com a análise de requisitos, alguns desses diagramas são descritos a seguir (FOWLER; SCOTT, 2000).

2.5.1 Diagramas de Casos de Uso

Caso de uso é um artefato da UML utilizado para documentação e especificação de um ou mais cenários. Cada cenário representa uma seqüência de passos que descrevem a interação entre um usuário (atores) e um sistema.

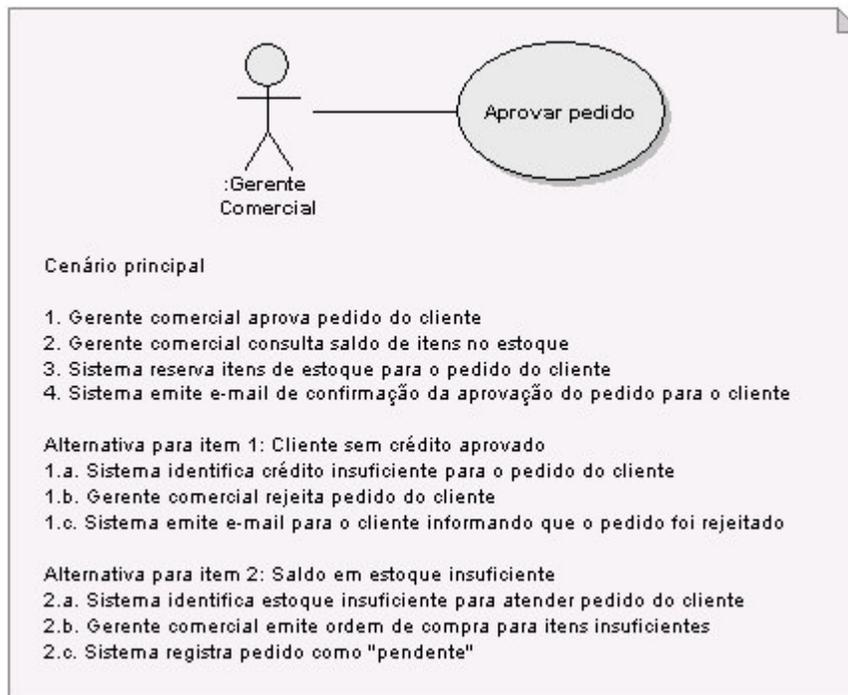
Em um diagrama de casos de uso, os casos que correspondem a um mesmo assunto ou processo são relacionados com os atores que os desempenham. Um único ator pode desempenhar um ou mais casos de uso, e estes podem ser desempenhados por um ou vários atores.

Nesse diagrama, os atores são graficamente representados pela figura humana, apesar de poderem simbolizar também outros elementos de interação, como um outro sistema que necessite de informações do sistema que está sendo projetado. Os casos de uso são representados no diagrama por uma elipse que contém uma breve descrição; e as ligações (linhas) representam quais atores desempenham cada um dos casos de uso. A Figura 9 ilustra um exemplo de diagrama de caso de uso para um sistema de comércio eletrônico, e a Figura 10 mostra como esse caso pode ser especificado.



Fonte: do autor.

Figura 9 - Exemplo de um diagrama de casos de uso



Fonte: do autor.

Figura 10 - Exemplo de especificação de um caso de uso

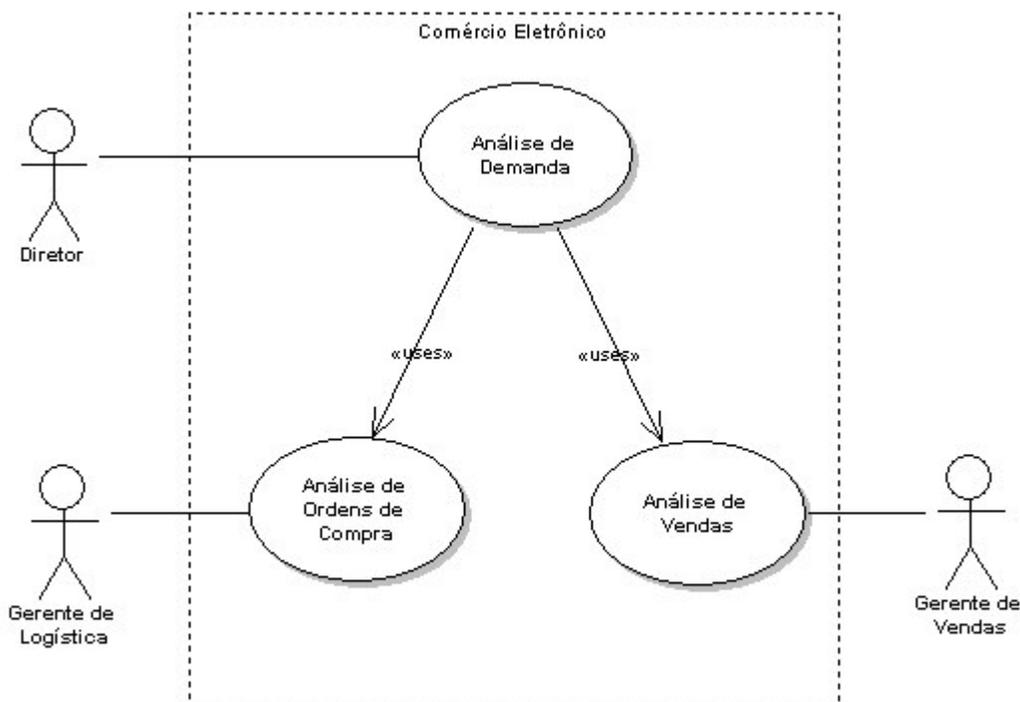
Casos de uso são utilizados para descrever o modelo de uma aplicação e do software a ser desenvolvido para atendê-la. O modelo de aplicação que será usado como base para a especificação do software integra diversos tipos de informações dos usuários, e as características dos diagramas e descrições de casos de uso estão de acordo com os atributos que esse modelo de aplicação deve possuir, tais como (IEEE, 1984):

- alto nível de abstração: apresentação do nível de visão dos usuários do sistema e captura de seus conceitos diretamente, não misturando informação de alto nível com aquela que é relevante apenas para os níveis mais baixos do processo de desenvolvimento;
- legibilidade: apresentação da especificação para os usuários em sua visão do conteúdo, evitando-se problemas conhecidos de não entendimento de linguagens formais de especificação de requisitos;

- precisão: descrição dos requisitos de alto nível necessários para definir o escopo do projeto, determinar os limites do sistema e estabelecer os principais objetos, regras e processos envolvidos; detalhes adicionais podem ser incluídos ou deixados para etapas subseqüentes, evitando poluir o modelo da aplicação e dificultar o entendimento dos usuários; e
- rastreabilidade: o estágio de definição de requisitos é usualmente seguido pelas fases de análise detalhada e projeto, sendo necessário que o modelo da aplicação possua uma estrutura que permita seu mapeamento ou rastreabilidade em etapas posteriores.

Um modelo de casos de uso da aplicação deve descrever na organização processos de negócio que serão analisados e deve também prover informação para o usuário do DW, identificando como os analistas utilizam-no e especificando papéis para os usuários e processos de negócio que cada papel analisa, contendo os componentes conceituais (além dos que tradicionalmente compõem um diagrama de casos de uso) ilustrados na Figura 11 e explicados a seguir.

- Sistema analisado: representado por um retângulo de cantos arredondados, é um conceito de modelagem que simboliza a área ou o negócio que deve ser analisado.
- Analista: representado por um ator com um papel que pode ser definido e que pode interagir com o sistema de análise.
- Caso de uso: é a construção de modelagem para um processo de negócio a ser analisado.
- Associações: identificam casos de uso que agregam funcionalidades de outros casos de uso.



Fonte: List et al., 2001.

Figura 11 - Exemplo de um modelo de casos de uso para um sistema analisado

Como a descrição dos casos de uso usualmente não provê a equipe de desenvolvimento de um DW com informações suficientemente detalhadas que devam ser disponibilizadas e sobre as funcionalidades a serem construídas, a tentativa de concentrar todos os requisitos de um DW nos casos de uso pode torná-los extremamente complexos. Por outro lado, se apenas os requisitos de usuários constarem nos casos de uso, a equipe de desenvolvimento estará desprovida de detalhes técnicos e de informações necessárias para a construção do sistema. Para contornar esse problema, cada caso de uso deve ser elaborado considerando-se o contexto dos requisitos de sistema, quando este não for muito pequeno e sua descrição não for suficiente para detalhá-lo. Um requisito de sistema pode ser utilizado por mais de um caso de uso, devendo ser associado a cada um deles e não detalhado repetidas vezes dentro de cada caso de uso (BRUCKNER; LIST; SCHIEFER, 2001).

2.5.2 Diagramas de estados

Um estado é uma situação que satisfaz alguma condição durante a vida de um objeto, o qual irá executar alguma atividade em resposta a um evento ou esperar que este ocorra (FURLAN, 1998).

Diagramas de estado são utilizados para descrever o comportamento de um sistema, identificando todos os estados possíveis que um objeto pode ter e determinando como o estado desse objeto muda a partir de eventos e condições que o afetem através de vários casos de uso.

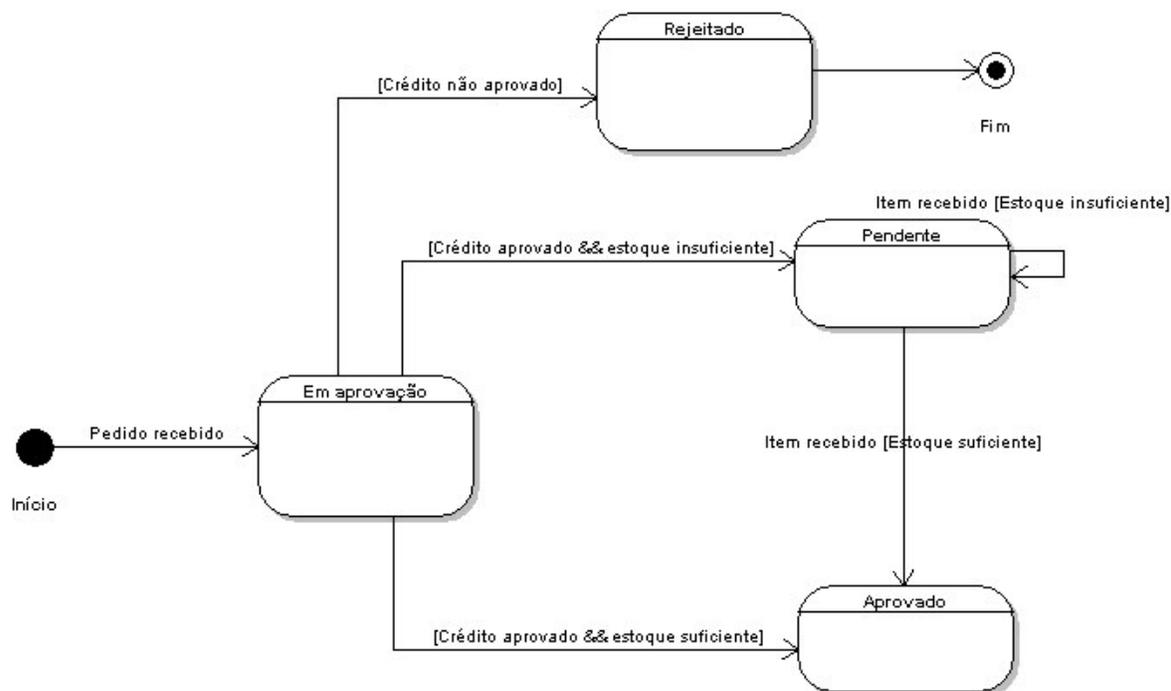
Sendo os casos de uso um dos principais artefatos gerados durante a análise de requisitos, os diagramas de estados podem ser úteis também nessa fase para um melhor detalhamento do comportamento dos objetos envolvidos nos casos de uso.

Abaixo são apresentados os principais componentes de um diagrama de estado.

- Início: representado por um círculo, identifica os possíveis estados iniciais para um objeto.
- Transição: representada por setas que mostram quais estados, a partir do atual, um objeto pode assumir.
- Rótulo: texto associado a uma transição que descreve os eventos que geraram a transição, as condições para que esta ocorra ou as ações a serem tomadas quando a transição ocorre; usualmente um rótulo é expresso na forma “evento [condição] / ação”.
- Estado: representado por um retângulo de bordas arredondadas com uma descrição do estado e, opcionalmente, uma atividade que será executada enquanto o objeto permanecer nesse estado.

A principal diferença entre a ação de um rótulo e a atividade de um estado é que uma ação é um processo de curta duração que não pode ser interrompido; já uma atividade é um processo que pode ter maior duração e ser interrompido por um evento.

A Figura 12 mostra um exemplo de como o estado de um pedido se altera durante todo o processo de compra em um sistema de comércio eletrônico.



Fonte: do autor.

Figura 12 - Exemplo de um diagrama de estados

2.5.3 Diagramas de atividades

Atividades são ações executadas por qualquer um dos atores envolvidos em um sistema ou pelo próprio sistema (FURLAN, 1998).

Um diagrama de atividades permite descrever uma seqüência de atividades, dando suporte à modelagem de comportamentos condicionais ou paralelos em um sistema. Esses diagramas são úteis para a compreensão de fluxos de trabalho

(workflow), descrição de um algoritmo seqüencial, projeção de aplicações de processamento paralelo e, principalmente, análise de casos de uso.

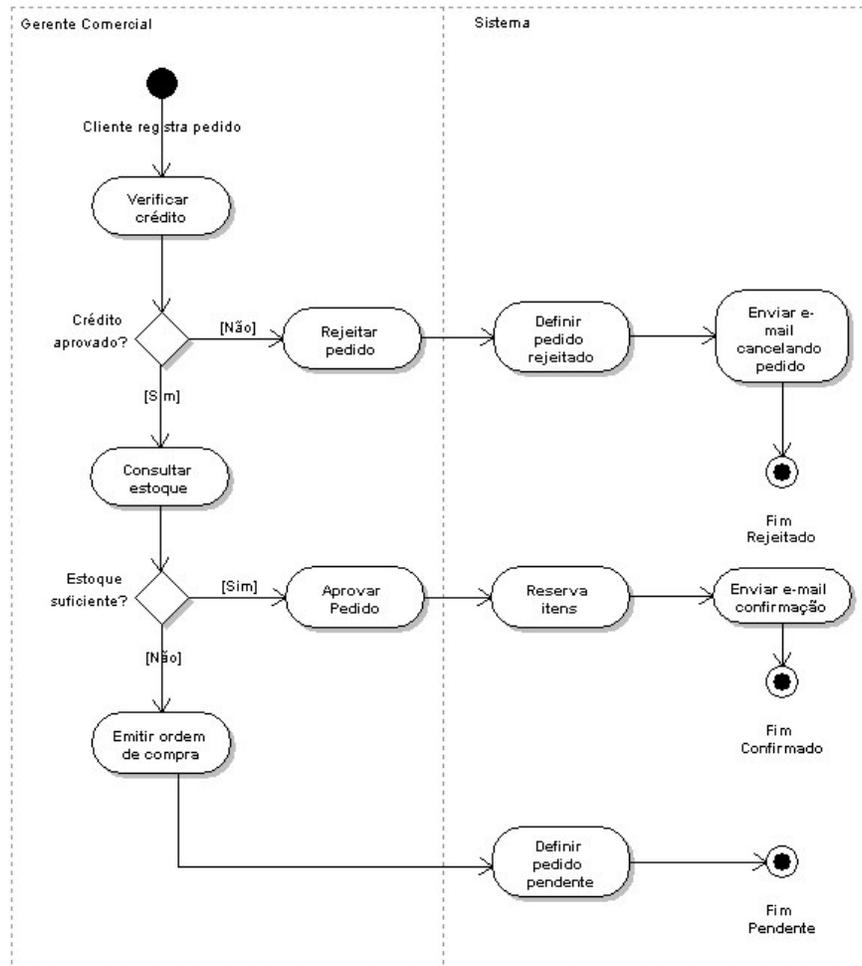
Com uma apresentação na forma de diagrama de componentes visuais, diagramas de atividades podem ser úteis também no detalhamento de casos de uso, facilitando a compreensão dos usuários de um sistema durante a fase de levantamento de requisitos.

A seguir, são apresentados os principais componentes de um diagrama de atividades.

- Início e fim: representados por um círculo e por um círculo com uma circunferência, identificam o ponto de partida para a execução das atividades e as situações em que a seqüência de atividades é finalizada.
- Atividade: representada por um retângulo de laterais arredondadas, contém uma descrição da atividade que será realizada.
- Transição: representada por setas que mostram a seqüência lógica de execução de atividades.
- Desvio condicional e intercalação: representado por um losango, identifica caminhos alternativos a serem seguidos para a execução das atividades dependendo de condições estabelecidas – as diferentes condições ficam associadas a cada transição possível a partir do ponto de desvio e são descritas no formato “[condição]”; a intercalação representa o término dos caminhos alternativos.
- Separação e junção: são representadas por segmentos de reta de maior espessura que indicam os caminhos que serão seguidos de forma paralela e onde eles se encontram.

- Raias: representadas por linhas paralelas tracejadas, permitem identificar os atores que executam cada uma das atividades descritas no diagrama.

A Figura 13 mostra um exemplo de diagrama de atividades que descreve o processo de compra em um site de comércio eletrônico.



Fonte: do autor.

Figura 13 - Exemplo de um diagrama de atividades

2.5.4 Diagramas de classes

Segundo Furlan (1998), uma classe “é a representação de um conjunto de coisas reais ou abstratas reconhecidas como sendo do mesmo tipo por compartilhar as mesmas características de atributos, operações, relações e semântica”. Classes

descrevem um tipo que relaciona um conjunto de objetos, ou seja, modelam a estrutura e as capacidades de uma instância de objeto.

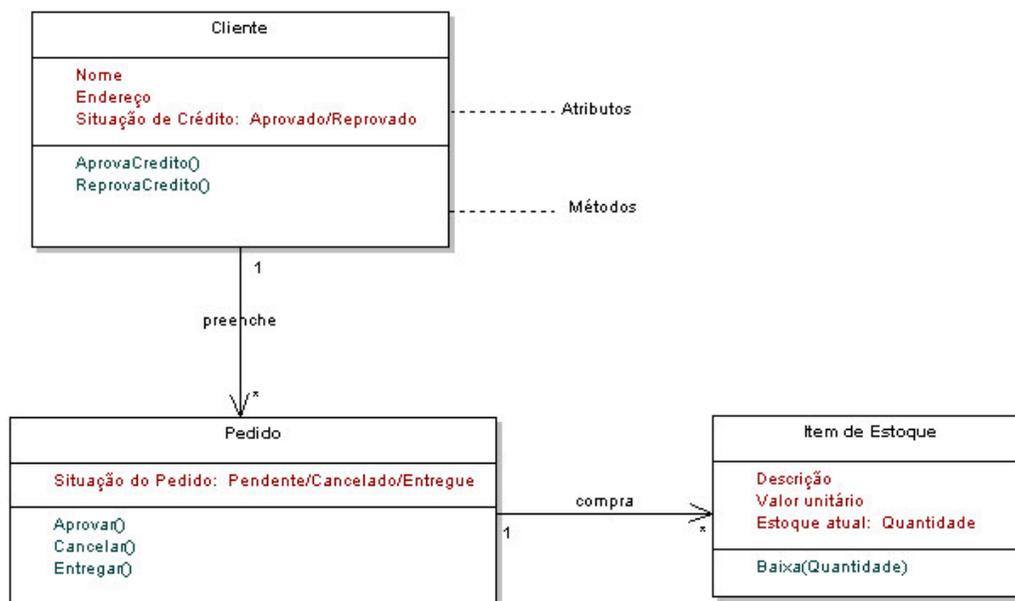
Diagramas de classes descrevem os tipos de objetos utilizados em um sistema e os relacionamentos que existem entre eles, podendo ser construídos sob três perspectivas, listadas a seguir.

- Conceitual: representa os conceitos do domínio que está sendo estudado. Normalmente projetado com pouca ou nenhuma preocupação com o software, é independente de qualquer linguagem de desenvolvimento.
- Especificação: representa os conceitos de interface do software, mas não sua implementação.
- Implementação: representa um modelo de classes que descreve a implementação do software.

Além das perspectivas, têm-se também os principais componentes de um diagrama de classes, exibidos abaixo.

- Classes: representadas por um retângulo, são identificadas por um nome e possuem uma descrição de atributos e métodos aplicáveis.
- Associações: representadas por linhas, indicam a maneira como as classes se relacionam.

A Figura 14 mostra um exemplo de diagrama de classes definido sob a perspectiva conceitual:



Fonte: do autor.

Figura 14 - Exemplo de um diagrama de classes

Os objetos que surgem na definição de requisitos destacam características destes ou dos processos de negócio analisados, mas também representam dimensões clássicas ou fatos usualmente utilizados em um DW.

2.6 Conclusões

Neste capítulo foram apresentados os principais conceitos relacionados à análise de requisitos em projetos de DW, entre eles:

- a classificação dos sistemas de informação, que contextualiza um projeto de DW dentro da evolução dos diferentes tipos e modelos de sistemas;
- os conceitos relacionados a um projeto de DW, como sua estrutura, as operações aplicadas na análise de dados, a forma como os dados são modelados e os conceitos dimensionais, Esses conceitos são utilizados como pontos focais na elaboração de uma documentação de requisitos;
- o modo como um DW passa por um processo de desenvolvimento de software composto de diversas disciplinas, entre elas a análise de requisitos, que é

praticamente a última etapa na qual há grande interação com os usuários finais na compreensão do problema;

- a maneira com que os requisitos, na análise de requisitos, devem ser classificados e documentados em uma estrutura que possa ser compreendida e validada não somente pela equipe técnica, mas também pelos principais usuários do sistema que será projetado. A documentação desses requisitos deve utilizar artefatos que transmitam o aspecto conceitual dos requisitos – como os diagramas de casos de uso, de atividade ou de estados da UML –, que permitem descrever características funcionais e estruturais de um sistema em uma linguagem gráfica de fácil compreensão.

Os conceitos acima foram utilizados como base para a elaboração da proposta de documentação de requisitos apresentada no próximo capítulo.

3 PROPOSTA PARA DOCUMENTAÇÃO DE REQUISITOS EM PROJETOS DE DATA WAREHOUSES

Este capítulo apresenta uma proposta de conteúdo para a documentação de requisitos em projetos de Data Warehouse. Nesta proposta, procura-se determinar qual seria o conteúdo mais adaptado à documentação de requisitos, permitindo a um analista preparar e adequar as técnicas de avaliação e levantamento de necessidades para melhor explorar o conhecimento dos usuários e mais facilmente alcançar o objetivo de determinação de requisitos para um DW.

Não é objetivo deste trabalho estabelecer uma forma de executar a disciplina de análise de requisitos em seu contexto mais amplo, tampouco descrever técnicas a serem utilizadas na coleta de informações, como entrevistas, análise de relatórios ou modelagem de processos.

Nas seções seguintes são apresentados os artefatos propostos para a elaboração da documentação de requisitos para um DW, e como os requisitos podem ser classificados para facilitar a documentação dos mesmos.

3.1 Artefatos propostos para requisitos de DW

Apesar de o principal foco da disciplina de análise de requisitos ser a especificação destes para um sistema de informação ou para um DW, outros documentos ou artefatos podem ser produzidos durante esse processo, os quais serão utilizados tanto como auxílio na identificação quanto na documentação complementar dos requisitos apurados.

Baseada em modelos apresentados pelo processo RUP, a documentação de requisitos gerada para os estudos de caso apresentado está representada no Quadro 5 em uma estrutura de artefatos e seções.

Artefatos	Seções
Visão geral do projeto	<ul style="list-style-type: none"> • Apresentação • Descrição do problema • Objetivos do projeto • Descrição dos principais usuários • Descrição da solução • Descrição das principais funcionalidades • Outras informações • Histórico de revisões
Modelo de casos de uso	<ul style="list-style-type: none"> • Diagrama de atores • Diagrama de casos de uso • Especificação de casos de uso
Especificação dos requisitos	<ul style="list-style-type: none"> • Requisitos de informação • Requisitos funcionais • Requisitos não funcionais
Especificação complementar	
Glossário	

Fonte: do autor.

Quadro 5 - Estrutura básica para a documentação de requisitos para DW

3.1.1 Visão geral do projeto

Este artefato tem como objetivos coletar, analisar e definir as necessidades dos usuários e as funcionalidades do DW de forma mais genérica e introdutória. A visão geral do projeto descreve as necessidades dos usuários, o porquê de elas existirem e como o DW vai contemplá-las, definindo o contexto e um escopo para o projeto.

Em princípio, deve-se tentar compor a visão geral do projeto com os itens a seguir:

Apresentação

Nesta seção, deve-se apresentar o contexto geral de negócio que envolve o projeto do DW, descrevendo itens como os principais conceitos envolvidos, o histórico evolutivo e a situação atual dentro de uma organização.

Descrição do problema

O problema que será solucionado pela implementação de um projeto de DW pode ser descrito com informações como:

- uma breve descrição geral do problema;
- referência dos principais usuários (ou tipo de usuários) relacionados ao problema;
- o impacto atual do problema;
- os principais benefícios que seriam gerados com a implementação de uma solução adequada.

Objetivos do projeto

Ao descrever os objetivos de um projeto de DW deve-se relacionar o que se espera alcançar com a construção desse DW. Nesta seção da documentação, podem ser incluídos os requisitos de negócio envolvidos com o problema a ser solucionado.

É importante salientar que requisitos de negócio não são mencionados no item da documentação referente aos requisitos do sistema, justamente por não representarem requisito do sistema em si. Requisitos de negócio e requisitos de um sistema de informação não devem ser confundidos. A palavra “requisito” denota algo que deve, obrigatoriamente, ser apresentado ou alcançado. Um exemplo de requisito de negócio poderia ser “aumentar em 15% o volume de vendas de uma

empresa”. Não é possível transferir uma responsabilidade como essa para um sistema de informação, seria como dizer que, se a empresa não atingir a meta de 15% de aumento de vendas, esse fracasso ocorreu devido ao sistema de informação. Requisitos de negócio devem aparecer na documentação de um sistema como objetivos a serem alcançados ou apoiados por um sistema de informação.

Descrição dos principais usuários

A correta identificação dos principais usuários ou de seus perfis envolvidos com o problema é uma das etapas do processo de modelagem de requisitos que permitirá o desenvolvimento de uma solução de DW que efetivamente contemple as necessidades desses usuários. Os objetivos são descrever e contemplar não somente as necessidades dos usuários mas também relatar as suas responsabilidades e os problemas-chave que eles acreditam que serão resolvidos com a implementação de um DW, podendo surgir daí justificativas para a modelagem de determinados requisitos.

Ao descrever um usuário do DW ou o seu perfil, podem-se identificar os itens descritos a seguir.

- Nome do usuário ou de seu perfil.
- Descrição das principais responsabilidades desse usuário ou perfil com relação ao problema descrito.
- Descrição dos problemas relacionados ao usuário ou perfil, incluindo tarefas complexas que serão simplificadas com a implementação do DW, o tempo gasto com a análise de dados no ambiente operacional atual e as aplicações utilizadas.
- Descrição das principais necessidades do usuário ou perfil, identificando possíveis soluções que são atualmente utilizadas, como a implementação do

DW, que poderá ajudar o usuário a chegar a essas soluções, e as melhorias esperadas com a implementação do DW em relação às soluções atuais.

- Usuários dos sistemas transacionais que serão utilizados como fonte de dados na construção do DW e quais serão efetivamente os usuários do DW.

Descrição da solução

A visão geral do projeto deve descrever, quando possível, as principais capacidades, as relações com outros sistemas ou aplicações e as configurações e dependências da solução a ser implementada. Sob a perspectiva de relacionamento com outros sistemas, deve-se relacionar a interação entre tais sistemas e o DW e identificar as necessidades de integrações entre eles. Quando houver necessidade, podem ser destacadas características específicas do ambiente de hardware e software no qual a solução de DW será implementada, desde que efetivamente sejam determinantes na modelagem de um ou mais requisitos do sistema.

Descrição das principais funcionalidades

Apresenta uma descrição de forma mais genérica das principais funcionalidades da solução a ser implementada, ou seja, o que deve ser suportado pelo sistema para contemplar as necessidades dos usuários.

Em um projeto de Data Warehouse essas funcionalidades refletem diretamente em todos os tipos de requisitos do sistema e devem ser descritas suficientemente para que sejam detalhadas com a construção do modelo de casos de uso e com a elaboração da especificação complementar.

As funcionalidades devem ser descritas sob o aspecto do “por que” devem ser implementadas, e não do aspecto do “como” – responsabilidade da etapa de análise

e projeto –, e priorizadas sob diferentes perspectivas, tais como benefício gerado, esforço de implementação, riscos ou estabilidade.

Outras informações

A descrição de outras informações pode ser necessária para que o contexto do problema e o escopo da solução sejam compreendidos. Esta seção pode mencionar requisitos da plataforma de hardware e software essenciais à implementação da solução, requisitos de desempenho, níveis de tolerância a falhas aceitáveis e outras características que não foram citadas na relação de funcionalidades.

Histórico de revisões

Este histórico mostra como a elaboração da documentação especifica que os requisitos ocorre de forma dinâmica, sendo que o acréscimo de novas informações ou de correções pode ocorrer durante qualquer fase do projeto.

O histórico de revisões deve identificar quando o documento foi feito ou se houve alterações, apontando o autor e uma breve descrição da modificação. A primeira linha do histórico mostra quando a primeira versão do documento foi elaborada.

Nas demais seções da documentação que descrevem a visão geral do projeto, podem surgir referências a uma versão específica do documento, a partir da qual, por exemplo, for inserido ou alterado um parágrafo.

3.1.2 Modelo de casos de uso

O modelo de casos de uso deve descrever o relacionamento entre os usuários (atores) e cada uma das funcionalidades do sistema proposto. Para um Data Warehouse este modelo pode ser baseado na identificação dos diferentes perfis de usuários descritos na visão geral do projeto. Cada caso de uso pode ser detalhado

para mostrar como as funcionalidades do sistema serão utilizadas pelos usuários de cada perfil.

A proposta de conteúdo para o modelo de casos de uso está estruturada no conjunto de seções a seguir.

Diagrama de atores

O diagrama de atores deve mostrar o relacionamento e, principalmente, a hierarquia existente entre os diferentes perfis de usuários identificados na visão geral do projeto. Essa hierarquia poderá interferir, principalmente, no levantamento de requisitos de segurança quando for necessário estabelecer para cada um dos perfis limites de acesso às informações disponíveis no modelo dimensional.

Diagramas de casos de uso

Os diagramas de casos de uso podem ser compostos agrupando-se as funcionalidades de uso de um determinado perfil de usuários. As funcionalidades que devem ser consideradas neste momento são aquelas relacionadas a especificações de necessidades de negócio e não a visões de detalhes de implementação. “Analisar o histórico anual de vendas” é um bom exemplo de funcionalidade a ser inserida nesse contexto; ao contrário, “executar detalhamento na consulta de indicadores através da operação *drill down*”, apesar de ser uma provável funcionalidade a ser implementada no sistema, será descrita como um requisito funcional dentro das operações OLAP e não faz parte do elenco de funcionalidades de negócio que possam ser compreendidas diretamente pelos usuários finais do sistema.

Especificação dos casos de uso

A especificação dos casos de uso deve apresentar uma descrição de como ocorre a interação entre o usuário e a funcionalidade do sistema que está sendo modelada. Essa especificação pode incluir informações como:

- nome do caso de uso;
- breve descrição que permita identificar o propósito do caso de uso (não mais do que um parágrafo);
- histórico de revisões;
- fluxo básico de eventos que descreve o que sistema faz para cada ação do usuário (ator), sem a preocupação de descrever “como” o sistema executa cada uma dessas operações;
- fluxos alternativos que descrevam com mais detalhes uma ou mais das ações mencionadas no fluxo básico do caso de uso;
- condições necessárias que devem ser verificadas antes da realização do caso de uso e também pós-condições que devem ser confirmadas após a sua realização;
- informações e diagramas complementares essenciais à correta compreensão da funcionalidade descrita em um caso de uso, como diagramas de atividades utilizados para detalhar alguma ação específica ou simplesmente para ilustrar a estrutura do fluxo de eventos do caso de uso.

É sempre importante destacar que os diagramas e a especificação dos casos de uso devem ser documentados em um nível de abstração que permita a total compreensão pelos usuários finais do sistema do tipo de funcionalidade que será disponibilizada.

3.1.3 Especificação de requisitos

A especificação de requisitos é o objetivo principal da fase de análise de requisitos, e como resultado dessa especificação espera-se um artefato que liste os requisitos necessários para o Data Warehouse que está sendo proposto. Requisitos em um sistema de Data Warehouse possuem algumas características distintas das de um sistema de informação transacional em virtude das próprias características intrínsecas daquele tipo de sistema.

- Em um Data Warehouse os usuários têm apenas acesso à consulta das informações, não precisam especificar validações usualmente necessárias para rotinas transacionais, como níveis de permissão para inclusão, alteração ou exclusão de dados.
- A manutenção da base de dados de um Data Warehouse é feita através de um conjunto de rotinas de carga e, quando necessário, da modelagem e utilização de uma área de preparo de dados (*Data Staging Area*).
- As operações realizadas por um usuário resumem-se basicamente às de consulta OLAP, as quais podem ser restritas em virtude do uso de uma ferramenta de consulta OLAP específica ou podem ser definidas como requisitos de construção quando uma ferramenta de visualização específica para o Data Warehouse proposto for construída.

O item 3.2 descreve como os requisitos de um Data Warehouse podem ser classificados para facilitar o seu levantamento e a sua documentação.

3.1.4 Especificação complementar

A especificação complementar deve ser utilizada sempre que for necessário detalhar um requisito, por exemplo, para descrever um processo de cálculo, especificar regras de negócio ou algoritmos a serem utilizados em um processo de carga de dados, ou sempre que for preciso documentar especificações que estejam relacionadas com mais de um requisito do sistema simultaneamente.

Não há padronizações estipuladas para a documentação de especificação complementar, sendo usualmente utilizados os formatos descritivo, tabular ou até mesmo gráfico. O importante é a especificação complementar não se prender a detalhes técnicos que não possam ser compreendidos pelos usuários finais do Data Warehouse.

3.1.5 Glossário

Termos e siglas utilizados dentro do escopo do sistema devem ser incluídos em um glossário que permita a rápida consulta através da busca indexada de seu conteúdo.

Referências cruzadas podem ser definidas entre os termos do glossário para identificar sinônimos. Em alguns casos, pode ser útil organizar os termos em grupos, que devem ser brevemente descritos para facilitar a compreensão do que cada grupo representa.

3.1.6 Rastreabilidade

Para garantir que a característica de rastreabilidade esteja disponível no modelo de requisitos de um sistema, devem ser identificados individualmente todos os componentes de cada documento gerado nessa fase e, sempre que possível,

relacionados esses componentes. Assim, deve ser criado um conjunto de relacionamentos que permite, a qualquer tempo, a identificação de quais componentes foram utilizados para a especificação de outros. Esse reconhecimento pode ser feito com uma simples numeração seqüencial ou através da combinação da numeração seqüencial com um conjunto de prefixos que identifiquem o tipo de componente.

3.2 Classificação dos requisitos

Durante o processo de análise de requisitos, deve-se procurar uma classificação que permita diferenciar os tipos de requisitos a serem documentados. Não há uma classificação específica a ser utilizada, porém as diferentes características apresentadas por diferentes tipos de requisitos direcionam para uma classificação mínima com as seguintes subdivisões:

- requisitos de informação;
- requisitos funcionais;
- requisitos não funcionais.

Os requisitos de usuário não estão nesta classificação e são descritos de forma implícita nos diagramas e na especificação de casos de uso, contemplando exatamente a visão de negócio que caracteriza as necessidades diretas dos usuários.

3.2.1 Requisitos de informação

Os requisitos de informação devem descrever as necessidades de informação que devem ser modeladas na fase de análise e projeto do sistema. Eles possuem importância fundamental na construção do Data Warehouse, uma vez que são

diretamente responsáveis pelo modelo de dados dimensional do projeto. Neste grupo de requisitos devem ser descritas as necessidades de informações que irão compor as futuras tabelas de fato e de dimensões do DW.

A modelagem dimensional dos dados propriamente dita não deve ser realizada na fase de análise de requisitos. Uma visão conceitual do modelo dimensional, entretanto, pode ser criada para descrever as hierarquias necessárias para as diferentes dimensões e o conteúdo que irá compor cada fato.

A especificação complementar pode ser utilizada para descrever detalhes a serem considerados nas rotinas e nos algoritmos que serão usados para transferir os dados dos sistemas transacionais de origem para o DW através da utilização da área de preparo de dados (*Data Staging Area*). O projeto detalhado das rotinas de carga do DW, entretanto, não é responsabilidade direta da disciplina de análise de requisitos, e sim da disciplina de análise e projeto do sistema.

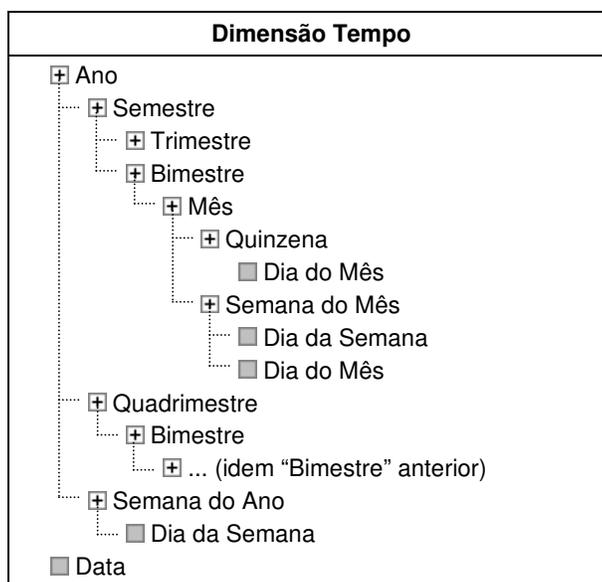
Algumas dimensões, por serem bastante comuns em modelos dimensionais, são discutidas a seguir.

Dimensões “Tempo”

As dimensões de tempo são utilizadas para posicionar, dentro de uma ordem cronológica, os fatos a serem armazenados em uma tabela de fatos do modelo dimensional. Não são usualmente representadas na forma de tabelas em sistemas transacionais, estando essa informação normalmente contida em atributos de registros de dados históricos, como pedidos, compras, vendas, movimentação bancária, etc.

Para representar a distribuição cronológica, têm-se dois tipos de dimensões, divididas por uma questão de otimização e performance do modelo dimensional resultante (KIMBAL et al., 1998):

- a dimensão que representa o dia e suas maiores granularidades, como mês, ano e todas as subdivisões intermediárias de tempo; e
- a dimensão que representa o dia em suas menores granularidades, como horas, minutos e segundos.



Fonte: do autor.

Figura 15 - Exemplo de uma dimensão tempo com diferentes granularidades para uma data

A Figura 15 mostra um exemplo de uma dimensão tempo em que foram definidos dois subconjuntos de atributos. O primeiro trata das diferentes granularidades geradas a partir de uma data em relação ao número de subdivisões distintas de meses dentro de um mesmo ano.

- “Data” representa um dia específico.
- “Ano” compreende o ano em um formato de 4 dígitos, e os valores possíveis dependem do período de tempo necessário a ser utilizado para esta dimensão, que será determinado pelo conteúdo das tabelas de fato do sistema.
- “Semestre” pode conter valores “1” ou “2”.
- “Quadrimestre” pode conter valores de “1” a “3”.
- “Trimestre” pode conter valores de “1” a “4”.
- “Bimestre” pode conter valores de “1” a “6”.

- “Mês” pode conter valores de “01” a “12”.
- “Quinzena” pode conter valores “1” ou “2”.
- “Dia do mês” pode conter valores de “01” a “31”.

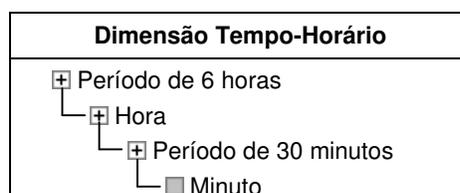
Em uma dimensão de tempo, não há necessariamente uma razão específica para que a hierarquia “dia” seja a de menor granularidade, sendo esta uma decisão de projeto.

O segundo subconjunto trata das granularidades geradas a partir dessa data em relação à distribuição semanal.

- “Dia da semana” pode conter valores de “1” a “7” representando o dia de domingo até o de sábado.
- “Semana do mês” pode conter valores de “1” a “6”.
- “Semana do ano” pode conter valores de “01” a “52”.

É evidente que nem todos os atributos mencionados para a dimensão Tempo serão necessários em todos os modelos dimensionais a serem criados, podendo essa dimensão ser adaptada de acordo com os requisitos essenciais a cada caso.

Dimensões de tempo que representam dias e suas maiores granularidades permitem modelar um sistema dimensional que responda a perguntas do tipo “Qual o mês, trimestre ou semestre de maior volume de vendas?” ou “Em qual ano o primeiro trimestre foi mais produtivo?”.



Fonte: do autor.

Figura 16 - Exemplo de uma dimensão tempo com granularidades para horas e minutos

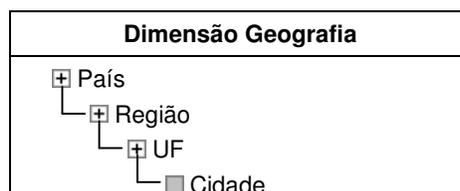
A Figura 16 exibe um exemplo de uma dimensão Tempo em que foram definidos atributos que representam informações relativas ao horário que determinado fato foi realizado:

- “Minuto” pode conter valores entre “00” e “59”.
- “Período de 30 minutos” pode conter os valores “00-29” ou “30-59” indicando se um fato ocorreu na primeira ou segunda metade de uma determinada hora.
- “Hora” pode conter valores entre “00” e “23”.
- “Período de 6 horas” pode conter os valores “00-06”, “06-12”, “12-18” ou “18-24” indicando em qual período do dia um fato ocorreu.

Dimensões de Tempo com granularidades que representam horas e minutos permitem modelar um sistema dimensional de forma a se obterem respostas a perguntas do tipo “Qual o horário ou período do dia em que é realizado o maior número de vendas?”.

Dimensão “Distribuição geográfica”

Uma dimensão que representa a distribuição geográfica, também denominada “Dimensão Geografia”, é utilizada em modelos dimensionais quando os fatos estão associados a dimensões originadas de tabelas transacionais que possuem atributos de localização (endereço, cidade, Estado, país, etc.).



Fonte: do autor.

Figura 17 - Exemplo de uma dimensão para representar a distribuição geográfica

A Figura 17 mostra um exemplo de uma “Dimensão geográfica” onde foram definidos atributos para representar as diferentes granularidades, como cidade e estado. Alguns atributos, como “Região” e “Unidade da Federação” podem ser reproduzidos através de valores codificados (siglas), descritivos ou ambos. Neste último caso seriam necessários dois atributos para representar cada um – codificado e descritivo.

Nos casos em que a maioria das análises referem-se a fatos ocorridos dentro do território nacional, a modelagem dimensional da distribuição geográfica pode não incluir o atributo “País”. Se esse atributo for considerado, a informação referente a cada um dos demais atributos pode ser desconhecida ou ainda sem interesse para análise dentro do modelo dimensional para certos valores de domínio do atributo “País”. Nesses casos um valor do tipo “Exterior” pode ser inserido no domínio de valores possíveis para as hierarquias “Região”, “Unidade da Federação” e “Cidade”.

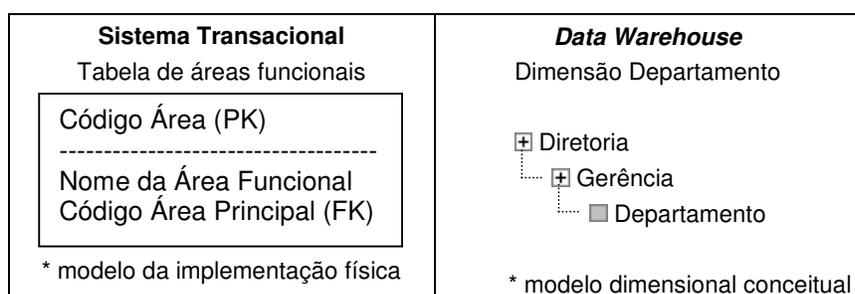
Dimensões que tratam hierarquias de forma recursiva

Dimensões com hierarquias tratadas de forma recursivas são usualmente encontradas em sistemas transacionais com auto-relacionamento em uma das fontes (tabelas) de dados usadas para a geração de uma dimensão. Exemplos comuns desse tipo de estrutura são encontrados nos departamentos de uma empresa, nas hierarquias de centros de custos, nos planos de contas contábeis e nas hierarquias de funcionários.

A principal característica dessa forma de estruturação dos dados de um sistema transacional é a possibilidade de criação de um número indeterminado de níveis hierárquicos. Ao transportar essa estrutura diretamente para um DW, essa característica pode gerar dificuldades na modelagem dimensional dos dados e na implementação dos requisitos funcionais de consulta, principalmente no que se

refere às nomenclaturas a serem atribuídas para cada um dos níveis hierárquicos. Usuários do DW podem também encontrar dificuldades em compreender como essa estrutura está organizada.

Uma solução que pode ser adotada para a implementação das hierarquias recursivas em um DW é o corte da estrutura em um dos níveis, o que possibilitará não somente criar uma nomenclatura auto-explicativa para cada nível hierárquico como também delimitar o nível de complexidade de consulta que poderá ser realizada pelos usuários do DW.



Fonte: do autor.

Figura 18 - Modelagem de dimensões que tratam de hierarquias recursivas

A Figura 18 mostra como uma tabela de áreas funcionais com estrutura hierárquica recursiva foi modelada em um DW como uma dimensão com apenas três níveis hierárquicos.

- Na tabela de áreas funcionais, a chave de identificação de uma área principal (PK – *Primary Key*) é utilizada para relacionar essa área principal com as suas áreas funcionais subordinadas, que possuem o código da área principal como chave de ligação (FK - *Foreign Key* ou Chave estrangeira), e não há um limite no número de níveis que podem ser definidos para essa hierarquia.
- Para modelar no DW a dimensão, apenas os três primeiros níveis hierárquicos de áreas funcionais foram utilizados, os quais foram denominados respectivamente de “Diretoria”, “Gerência” e “Departamento”.

Nomenclatura conceitual

Considerando-se que um dos objetivos de uma etapa de levantamento e análise de requisitos em um projeto de DW é apurar e documentar as necessidades dos usuários, é importante que os conceitos relacionados com os requisitos de informação sejam apresentados com uma nomenclatura não técnica, evitando o uso de termos como “data mart”, “hierarquia”, “tabela de fatos”. Isso permitirá que os usuários do sistema compreendam com mais facilidade o que foi documentado e auxiliará o processo de validação dessa documentação.

Os termos propostos a seguir são utilizados na apresentação do estudo de caso. Esses conceitos foram definidos através de aplicações prática dos termos em projetos de DW e do refinamento progressivo dos mesmo nesses projetos:

- Assunto: corresponde a um conjunto de informações relacionadas à mesma área funcional ou ao mesmo processo de negócio de uma organização, com equivalência ao conceito técnico de um “Data Mart”. São exemplos de assuntos distintos em um Data Warehouse: “Desempenho comercial”, “Produção de equipamentos”, “Controle de qualidade de produtos”, “Satisfação do cliente”. Sob o aspecto conceitual um assunto pode conter um conjunto de fatos e dimensões associadas, e um fato pode estar associado a mais de um assunto simultaneamente.
- Unidade de análise: dentro de um assunto, cada unidade de análise deve reunir indicadores que representem uma determinada característica a ser analisada de cada um dos assuntos do DW. Uma unidade de análise será implementada fisicamente em um banco de dados através de uma ou mais tabelas de fato, e cada tabela de fato pode conter uma ou mais unidades de análise. São exemplos

de unidades de análise para o assunto “Desempenho comercial”: “Propostas apresentadas” e “Vendas realizadas”.

- Unidade de conteúdo: cada unidade de conteúdo representa um indicador distinto da unidade de análise que pode ser observado e analisado sob diferentes aspectos, sendo sua representação física mais comum uma coluna da tabela de fato ou uma fórmula entre colunas distintas, com valores usualmente sumarizados através de operadores de contagem (COUNT), soma (SUM), cálculo de média (AVG) ou percentuais relativos. São exemplos de unidades de conteúdo para a unidade de análise “Vendas realizadas”: “Quantidade (em unidade) vendida”, “Valor (monetário) das vendas”, “Quantidade de vendas realizadas (notas fiscais emitidas)”.
- Unidade de distribuição: uma unidade de distribuição é utilizada para classificar o valor dos indicadores representados pelas unidades de conteúdo e é descrita fisicamente em um banco de dados pelas hierarquias das dimensões associadas à tabela de fato de uma unidade de análise. Uma unidade de conteúdo “Valor de vendas”, por exemplo, pode ser classificada pelas seguintes distribuições: distribuição geográfica das filiais, dos clientes ou dos representantes (“região”, “UF”, “cidade”); distribuição temporal das vendas realizadas (ano, mês, dia); distribuição nas categorias de produtos (“grupo de produto”, “família”, “tipo de produto”, “cor”); e distribuição segundo o tipo de venda realizada (“à vista”, “a prazo”, “financiada”).
- Unidade de filtro: pode ser: a) uma unidade de distribuição utilizada para selecionar um subconjunto de valores de uma ou mais unidades de conteúdo de uma unidade de análise. Por exemplo, em “vendas realizadas pelas filiais da Região Sul”, a unidade de distribuição “região” está sendo utilizada como

unidade de filtro para selecionar informações referentes apenas a uma região específica; ou b) uma faixa de valores de uma unidade de conteúdo utilizada para restringir o domínio de valores dos indicadores que devem ser analisados. Por exemplo, em “filiais com vendas realizadas superiores a R\$ 5.000,00 no mês atual”, a unidade de conteúdo “Valor das vendas” está sendo utilizada como unidade de filtro para o conjunto de filiais que satisfaçam o valor mínimo de venda a ser analisado.

- Unidade de agrupamento: é uma unidade de distribuição utilizada para subdividir os valores dos indicadores representados pelas unidades de conteúdo de uma unidade de análise em linhas (quebra ou agrupamento horizontal) ou colunas (quebra ou agrupamento vertical). Uma consulta com unidade de quebra ou agrupamento é usualmente descrita como “total de <unidade de conteúdo> **por** <unidade de distribuição>”. Quanto maior o nível (quantidade) de quebras horizontais maior será o número de linhas resultantes em uma consulta, e quanto maior o nível de quebras verticais maior será o número de colunas resultantes.
- Unidade de classificação: é uma unidade de distribuição ou de conteúdo utilizada para classificar o resultado de uma consulta de forma ascendente ou descendente – de forma alfabética para valores de unidades de distribuição com conteúdo textual ou de forma crescente ou decrescente para unidades de distribuição ou de conteúdo com valores numéricos.

Os conceitos apresentados podem ser relacionados a conceitos técnicos conhecidos por analistas e projetistas de Data Warehouse, como demonstrado no Quadro 6. Essa relação não é rígida, servindo apenas como referência para uma melhor compreensão dos termos conceituais apresentados.

Nomenclatura conceitual	Nomenclatura técnica
Assunto	Data Mart
Unidade de Análise	Tabela de Fatos
Unidade de Conteúdo	Coluna de uma Tabela de Fatos, ou operações sobre colunas
Unidade de Distribuição	Hierarquia de uma Tabela de Dimensão
Unidade de Filtro	Operação OLAP “ <i>Filter</i> ”
Unidade de Agrupamento	Operações OLAP “ <i>Drill Down</i> ” e “ <i>Drill Up</i> ”
Unidade de Classificação	Operação OLAP “ <i>Sort</i> ”

Quadro 6 - Relação entre a nomenclatura conceitual e técnica para requisitos de informação

3.2.2 Requisitos funcionais

Estes requisitos devem descrever as funcionalidades que estarão disponíveis no sistema. Sendo a consulta (OLAP) à base dimensional a principal funcionalidade em um DW, é necessário que os requisitos funcionais detalhem como essa operação poderá ser executada, considerando-se duas situações distintas:

- quando uma ferramenta de consulta OLAP de mercado for utilizada, devem-se verificar quais operações OLAP (*drill down*, *drill up*, *drill across*, etc.) estão disponíveis nessa ferramenta e se essas operações são suficientes para atender às necessidades de cada um dos usuários ou perfis de usuários do sistema; e
- se uma aplicação for construída especificamente para consultar os dados de DW, devem ser especificadas quais das operações OLAP estarão disponíveis para os usuários e como cada uma delas será executada considerando-se as possíveis diferenças entre os diversos ambientes operacionais de execução da aplicação para o usuário final, como Web e cliente/servidor.

Para determinados usuários ou perfis de usuários pode ser necessário ainda identificar as principais visões a serem disponibilizadas para consulta, as quais devem trazer valores pré-selecionados para a consulta OLAP para um determinado fato a ser analisado. Cada visão pode:

- gerar um resultado de consulta a ser apresentado, considerando valores predeterminados de filtros e cortes para dimensões ou hierarquias de um fato definido, e permitindo que o usuário realize as operações OLAP sobre o resultado apresentado; e
- solicitar um conjunto de valores a serem utilizados como filtro para hierarquias previamente determinadas, apresentando o resultado da consulta para os valores especificados pelo usuário, os quais podem ser informados ou não.

3.2.3 Requisitos não funcionais

Paim e Castro (2002) descrevem um conjunto de requisitos não funcionais próprios para projetos de DW e sugerem a criação de um catálogo deste tipo de requisitos, dividindo-os em quatro grandes categorias, as quais não devem ser necessariamente mencionadas em todo e qualquer projeto de DW.

Requisitos de performance

Correspondem a requisitos que especificam o grau de eficiência de um DW ao responder a solicitações de processamento.

- Tempo de resposta: determina o tempo mínimo ou ideal para o processamento de consultas analíticas à base de dados dimensional e para o processamento da carga de dados.

- Consumo de memória: especifica o nível de otimização do uso de memória principal e secundária necessárias para o armazenamento e o processamento das informações contidas no DW.

Requisitos de segurança

Requisitos que denotam o grau de proteção e o nível de estabilidade do DW e de seus dados.

- Integridade: nível de precisão e validade dos dados dimensionais, armazenados ou sumarizados ao longo de diversos caminhos de agregação, coerência lógica dos dados, nível de redundância aceitável e grau em que o conhecimento fundamental está corretamente implementado no modelo dimensional.
- Disponibilidade: capacidade de alcançar todos os níveis de decisão, percentual de tempo que o DW está disponível considerando aspectos como carga e efetivação de dados, tolerância a falhas e capacidade de recuperação.
- Confidencialidade: capacidade de impedir acesso não autorizado aos dados e de liberar acesso a informações de domínio público.

Implementação de requisitos de confidencialidade

Entre os requisitos de segurança, a implementação de requisitos de confidencialidade é de extrema importância sempre que um DW tratar de informações consideradas sigilosas dentro de uma organização. Este tipo de requisito deve descrever características que permitam identificar as regras a serem utilizadas para determinar se um usuário ou perfil de usuários pode acessar cada uma das informações disponíveis no sistema.

Os requisitos de segurança relacionados à confidencialidade devem descrever:

- a necessidade de restringir o acesso aos dados disponibilizados no DW e como os usuários devem ser identificados e validados no sistema: usualmente essas operações são contempladas pela liberação de uma senha de acesso ao sistema;
- a quais unidades de análise ou de conteúdo um usuário ou perfil de usuário possui acesso: um usuário pode possuir a permissão de acesso ao fato que representa “vendas”, mas, entre o conjunto de atributos desse fato, pode apenas consultar a informação que representa “quantidades vendidas”, e não “valores vendidos”;
- se um usuário ou perfil de usuário pode utilizar todas as dimensões e unidades de distribuição disponíveis para uma unidade de análise com o intuito de realizar as operações de consulta ou se estas estarão restritas de alguma forma, por exemplo, usuários da área comercial não têm restrição de uso das dimensões relacionadas a “vendas”, porém os das demais áreas podem apenas utilizar as dimensões “tempo” e “produto”, não sendo permitida a consulta por “cliente” ou “filial”;
- se um usuário ou perfil de usuário pode utilizar todos os valores de domínio de uma dimensão ou hierarquia: pode ser necessário estabelecer que cada usuário da área comercial poderá consultar informações de “vendas” relativas somente à sua filial;
- se um usuário ou perfil de usuário pode utilizar-se de todos os requisitos funcionais definidos no sistema.

Requisitos de multidimensionalidade

São requisitos que descrevem a habilidade do DW em representar requisitos de suporte à decisão e em prover acesso a dados dimensionais e factuais.

- Conformidade: habilidade de representar, através de sua especificação final, aspectos próprios do DW exatamente da forma como devem ser ou foram implementados.
- Integrabilidade: determina a capacidade do DW em integrar informações operacionais bem como a frequência de atualização que satisfaça aos usuários de negócio.
- Interpretabilidade: requisitos que incluem característica como o nível em que os dados podem ser interpretados eficientemente para modelar o DW, o grau de legibilidade da documentação nas fontes de dados operacionais e o nível de descrição do prazo de validade e estabilidade dos dados.

Requisitos de usabilidade

Requisitos que indicam o nível de facilidade de uso das ferramentas e os softwares utilizados para acessar os dados contidos no DW.

- Operabilidade: facilidade de operar o DW e suas ferramentas.
- Flexibilidade: facilidade das ferramentas do DW em permitir a execução de consultas *ad hoc*.
- Facilidade de aprendizado: relação de habilidades físicas ou intelectuais necessárias para aprender os conceitos envolvidos com as informações que formam a base de dados do DW.

3.3 Conclusões

Neste capítulo foi apresentada uma proposta para a documentação de requisitos de projetos de DW que define um conjunto de artefatos a serem gerados, com base na integração de técnicas como UML e aspectos da modelagem dimensional. Ao ser utilizada por um analista de requisitos, esta proposta pode servir como um roteiro

para auxiliar na elaboração da estratégia mais adequada para a documentação dos requisitos de um DW.

4 ESTUDO DE CASO

4.1 Introdução

Este capítulo apresenta um estudo de caso em que foi aplicada a proposta de documentação de requisitos descrita no capítulo anterior. As seções seguintes apresentam o resultado final dessa documentação. Em algumas seções da documentação de requisitos, houve adaptações e remoção de alguns itens para melhor ilustrar as propostas apresentadas neste trabalho. Os históricos de alterações, por exemplo, foram suprimidos para facilitar a apresentação final do resultado. O estudo de caso apresenta o resultado obtido pela análise de requisitos do projeto “Diretório Lattes de Grupos de Pesquisa”.

4.2 Estudo de caso: Diretório Lattes de Grupos de Pesquisa

4.2.1 Visão geral do projeto

Apresentação

O Diretório de Grupos de Pesquisa no Brasil (DGP), projeto do CNPq, contém uma base de dados mantida desde 1992. Originou-se em 1991, a partir de uma proposta de elaboração de um almanaque de pesquisa no CNPq bem como de um levantamento de grupos de pesquisa realizado pelo Fórum Nacional de Pró-Reitores de Pesquisa, com o intuito de organizar o Programa de Laboratórios Associados encomendado em 1990 pela então Secretaria de Ciência e Tecnologia (atual Ministério de Ciência e Tecnologia) (GUIMARÃES, 1994). O objetivo do projeto era oferecer um suporte informacional atualizado acerca das atividades de pesquisa

através de bases censitárias sobre todos os grupos de pesquisa em atividade no País.

Hoje, o DGP tem o intento de ser uma plataforma de informação sobre o parque científico e tecnológico brasileiro, que em virtude do esforço empreendido pelo Brasil após a Segunda Grande Guerra, hoje é considerado o maior parque de C&T da América Latina. Entretanto, ainda há carência de informação, o que enfraquece e dificulta a tomada de decisão sobre os desígnios da C&T nacional. Tal fato transforma o DGP em um instrumento essencial para a gestão de C&T (MARTINS; GALVÃO, 1994).

Como parte integrante da Plataforma Lattes, o Diretório de Grupos de Pesquisa no Brasil (<<http://lattes.cnpq.br/diretorio>>) é uma base de dados sobre os grupos de pesquisa em atividade que armazena informações referentes aos recursos humanos constituintes dos grupos, às linhas de pesquisa em andamento, às especialidades do conhecimento, aos setores de aplicação envolvidos, à produção científica e tecnológica e aos padrões de interação com o setor produtivo.

Os grupos de pesquisa inventariados estão localizados em universidades, instituições isoladas de ensino superior, institutos de pesquisa científica, institutos tecnológicos, laboratórios de pesquisa e desenvolvimento de empresas estatais e em algumas organizações não-governamentais com atuação em pesquisa científica ou tecnológica.

A base de dados que se apresenta no Diretório de Grupos de Pesquisa possui três finalidades principais:

- no que se refere à sua utilização pela comunidade científica e tecnológica no cotidiano do exercício profissional, é um eficiente instrumento para o intercâmbio

e a troca de informações, sendo capaz de responder “quem é quem”, “onde se encontra”, “o que está fazendo” e “o que produziu recentemente”;

- em relação ao planejamento e à gestão das atividades de C&T, constitui-se uma poderosa ferramenta de análise, seja no nível das instituições, no das sociedades científicas ou, ainda, no das várias instâncias de organização político-administrativas do País, cujo caráter censitário convida ao aprofundamento do conhecimento por meio das inúmeras possibilidades de estudos empíricos que alcancem respostas sobre campos não cobertos pelos dados, como, por exemplo, o financiamento ou a avaliação qualitativa da produção científica e tecnológica;
- finalmente, no que se refere à base de dados, na medida em que pretende ser recorrente, virá a ter cada vez mais um importante papel na preservação da memória da atividade científico-tecnológica no País.

Descrição do problema

A falta de um repositório de informações organizado de forma a subsidiar processos analíticos sobre o Diretório de Grupos de Pesquisa dificulta análises sobre a evolução dos censos de pesquisa e sobre os dados associados aos currículos dos integrantes dos grupos de pesquisa ao longo do tempo.

A implementação de uma base de dados dimensionais com informações extraídas da base de dados operacional do Diretório de Grupos de Pesquisa permitiria atender a duas visões principais de análise: a) censo da pesquisa e evolução das principais unidades de análises e atores do censo (grupos de pesquisa, linhas de pesquisa, pesquisadores, estudantes); e b) evolução das unidades de distribuição, como áreas do conhecimento, das instituições participantes e setores de atividade.

Objetivos do projeto

O Data Mart do Diretório de Grupos de Pesquisa tem como objetivos:

- implementar uma base de dados dimensional que subsidie análises e recuperação de informações sobre os diversos censos realizados para fins de consulta em portais e sistemas de informação ligados à Plataforma Lattes; e
- implementar e disponibilizar para a comunidade científica as diferentes análises a serem realizadas, segundo uma visão histórica do censo do Diretório de Grupos de Pesquisa, através da construção de instrumentos (portais e sistemas de informação) de divulgação e de análise dos censos.

Descrição dos Principais Usuários

Os principais usuários do Sistema Diretório de Grupos de Pesquisa são descritos a seguir. O Sistema Diretório de Grupos de Pesquisa contempla as informações operacionais que são utilizadas como base para a construção do Data Mart do Diretório de Grupos de Pesquisa.

- Dirigentes de pesquisa das instituições participantes são responsáveis pela definição dos líderes e pela certificação dos grupos na base de dados do Diretório. Os dirigentes são também responsáveis pela qualidade da informação que está sendo certificada, já que possuem à sua disposição uma série de instrumentos de análises estatísticas e textuais sobre os grupos e integrantes da instituição.
- O cadastro das informações referentes ao grupo, aos pesquisadores, aos estudantes, ao pessoal de apoio técnico e às linhas de pesquisa associadas ao grupo são de responsabilidade dos líderes dos grupos.

- Os dados pessoais sobre os pesquisadores e estudantes relativos à produção científica, tecnológica e artística são de responsabilidade de cada pesquisador e estudante, que os informam através da atualização do Currículo Lattes.

Já o Diretório de Grupos de Pesquisa possui como principais usuários:

- Usuários em geral (incluindo pesquisadores, estudantes, empresários, entre outros), que utilizarão principalmente a funcionalidade da “Busca Textual” para localizar informações sobre os grupos com maior nível de detalhe ou para intercâmbio de informações entre participantes de diferentes grupos de pesquisa.
- Gestores de C&T do CNPq, de fundações, de universidades, entre outros, que farão uso das funcionalidades que fornecem indicadores (Séries Históricas, Súmula Estatística, Plano Tabular e Estratificação dos Grupos de Pesquisa) com interesse de avaliar a distribuição quantitativa da pesquisa sob determinados aspectos ou através de comparativos institucionais.

Descrição da solução

A construção de um data mart a partir da base de dados do Diretório de Grupos de Pesquisa permitirá não somente a realização de análise das informações sobre os diversos censos realizados como também a disponibilização de visões comparativas históricas entre informações de censos anteriores ou entre os censos anteriores e a posição atual do Diretório de Grupos de Pesquisa.

Evolução censitária

O Data Mart de Grupos de Pesquisa permitirá a construção de instrumentos de análise da evolução dos censos da pesquisa, realizados junto aos grupos brasileiros, contemplando o mapeamento de auxílios, avaliação, censo e estatísticas referentes aos grupos de pesquisa, na ótica de análise dos profissionais do CNPq. Tal

construção se dará através de alternativas de conexão entre os diferentes censos dos grupos, de forma a apresentar os levantamentos censitários segundo evolução no tempo.

Evolução temporal

A evolução temporal levará em conta os períodos registrados nas bases de dados do Diretório de Grupos de Pesquisa e nas bases com as quais ele se comunica, referindo-se às variáveis de tempo associadas a grupos (ex.: data de criação) e a indivíduos integrantes (ex.: data de formação, data de publicação, etc.), de forma a permitir que grupos de pesquisa mapeados em censos posteriores possam ser considerados em análises de evolução que compreendam períodos anteriores aos de sua criação.

Descrição das principais funcionalidades

Data Mart de Grupos de Pesquisa

Composto de uma base de dados dimensional gerada a partir de informações extraídas do Sistema Diretório de Grupos de Pesquisa, subsidiará análises e recuperação de informações sobre os diversos censos realizados, para fins de consultas nos sites de divulgação e de análise desses censos.

Séries Históricas

Site que permitirá analisar a evolução de elementos da pesquisa brasileira, ao longo dos censos realizados. Informações na escala temporal têm interessado não somente aos analistas de C&T. Um exemplo é a participação feminina crescente na pesquisa nacional, que vem ganhando destaque em veículos de imprensa de grande circulação no Brasil.

Súmula Estatística

Deve disponibilizar um conjunto de tabelas e gráficos com informações que sintetizem o conteúdo da base de dados e que forneçam um retrato bastante nítido da capacidade instalada de pesquisa no País. Embora de conteúdo predominantemente descritivo, deve fornecer pistas analíticas sobre determinadas características desse retrato. Este site é formado por tabelas estatísticas com indicadores da atividade de pesquisa para cada uma das unidades de informação disponíveis nas bases de dados do Diretório de Grupos de Pesquisa (*i.e.*, Grupos, Linhas de Pesquisa, Pesquisadores, Estudantes, Técnicos e Produção).

Plano Tabular

Deve permitir a visualização quantitativa do perfil da pesquisa no Brasil. Trata-se de um sistema (site) que oferece a alternativa de cruzamento de variáveis capazes de gerar resultados que podem ser exportados para outros formatos de arquivos para futuras consultas. Considerando-se as consultas a elementos específicos, o Plano Tabular deve disponibilizar um conjunto amplo de indicadores através das combinações entre as variáveis escolhidas de forma dinâmica e configurável pelo o usuário do Data Mart do Diretório de Grupos de Pesquisa.

Busca Textual

O site de Busca Textual deve permitir a recuperação de informações sobre qualquer dado presente na base do Data Mart do Diretório de Grupos de Pesquisa, a partir de três decisões a serem tomadas pelo usuário: (i) qual informação deverá ser recuperada; (ii) onde essa informação deverá ser encontrada; e (iii) segundo qual critério deve ser procurada. Esse site possibilitará conhecer o detalhe da atividade científica e tecnológica do País por diversas formas de consulta direta – livre ou

direcionada com procura específica e quebras para classificação das informações desejadas.

Estratificação dos Grupos de Pesquisa

Site que possibilitará a divulgação dos critérios e dos resultados da avaliação qualitativa dos grupos de pesquisa de Instituições de Ensino Superior, segundo a relação de julgamento das agências e a produção dos integrantes do grupo de pesquisa.

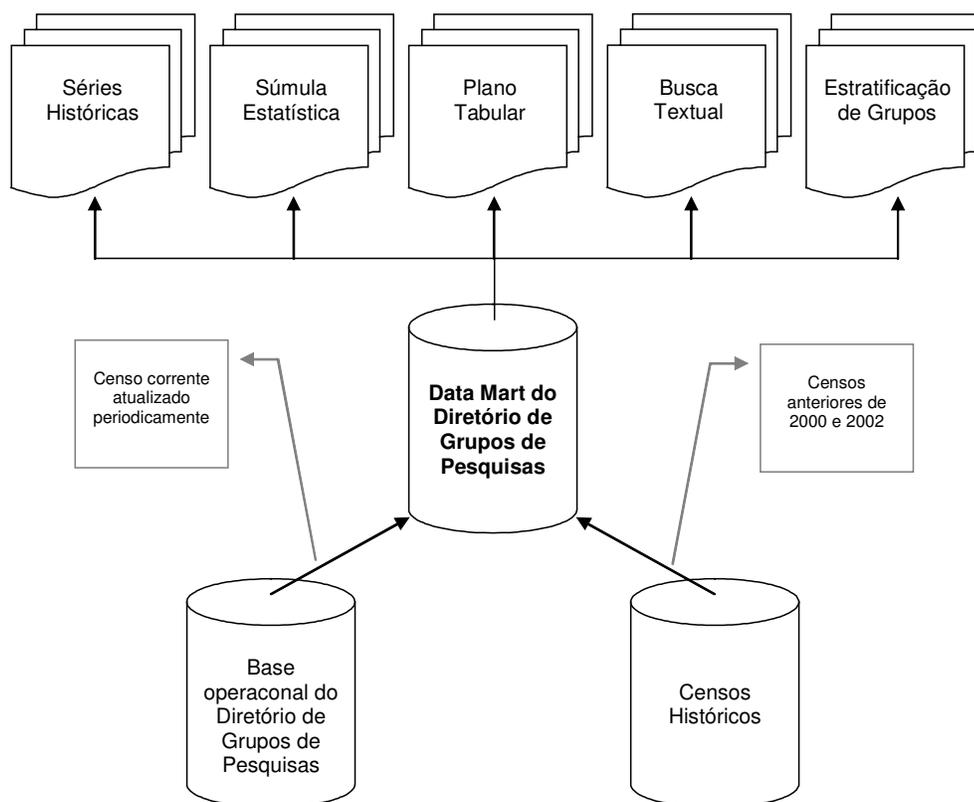


Figura 19 - Origens de dados e funcionalidades do DM do Diretório de Grupos de Pesquisa

Os resultados a serem apresentados nos sites “Séries Históricas”, “Súmula Estatística” e “Estratificação dos Grupos de Pesquisa” serão gerados e armazenados estaticamente a partir das informações dos censos, de forma que as análises sejam rapidamente mostradas para os usuários, dispensando o acesso à base do Data

Mart do Diretório de Grupos de Pesquisa para a construção dinâmica dos resultados sempre que um desses indicadores for solicitado.

4.2.2 Modelo de Casos de Uso

Diagrama de atores

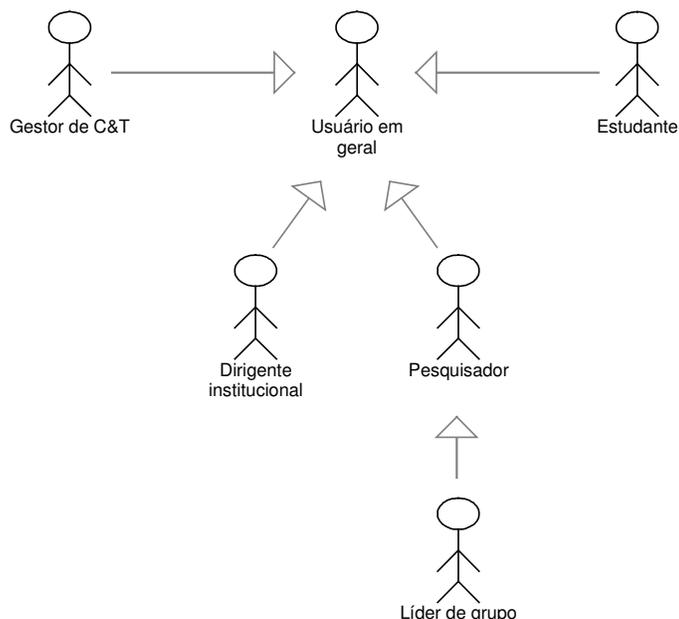


Figura 20 - Diagrama de atores do Data Mart do Diretório de Grupos de Pesquisa

ACT-01: Usuário em geral

Qualquer usuário que possa ter interesse em consultar informações sobre a distribuição quantitativa da pesquisa no Brasil, como empresários e investidores, profissionais de setores públicos e privados, etc.

ACT-02: Gestor de C&T

Gestores de C&T do CNPq, de fundações e universidades que farão uso dos indicadores de Ciência e Tecnologia oferecidos pelos sites que utilizam o Data Mart de Grupos de Pesquisa (Séries Históricas, Súmula Estatística, Plano Tabular e Estratificação dos Grupos de Pesquisa) com interesse de avaliar a distribuição

quantitativa da pesquisa sob determinados aspectos ou através de comparativos institucionais.

ACT-03: Dirigente Institucional

Dirigentes de pesquisa das instituições participantes do Diretório de Grupos de Pesquisa responsáveis pela definição dos líderes e pela certificação de grupos de pesquisa de sua instituição.

ACT-04: Pesquisador

Pessoa com currículo Lattes cadastrado na base do CNPq que é associada por um líder a um ou mais grupos de pesquisa e que pode orientar estudantes desses grupos. O pesquisador utilizará principalmente o site da “Busca Textual” para localizar informações sobre os grupos com maior nível de detalhe ou para intercambiá-las entre participantes de diferentes grupos de pesquisa.

ACT-05: Estudante

Pessoa com currículo Lattes cadastrado na base do CNPq que é associada a um ou mais grupos de pesquisa pelo líder de cada um dos grupos e orientada por pesquisadores desses grupos. O estudante também fará uso principalmente do site da “Busca Textual”.

ACT-06: Líder de Grupo

Pesquisador definido pelo dirigente institucional como líder de um grupo, tornando-se responsável pela atualização de informações do grupo de pesquisa que lidera, como a definição de linhas de pesquisa associadas a cada grupo e a de seus integrantes (pesquisadores, estudantes e pessoal técnico).

Diagramas de Casos de Uso

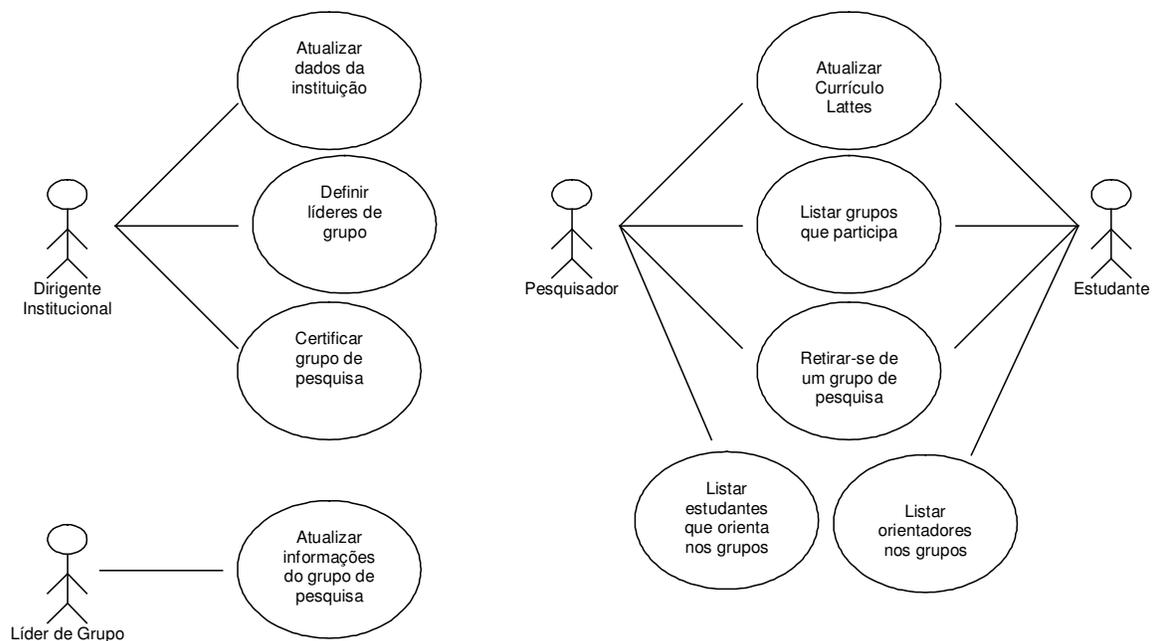


Figura 21 - Diagrama de casos de uso do Sistema Diretório de Grupos de Pesquisa

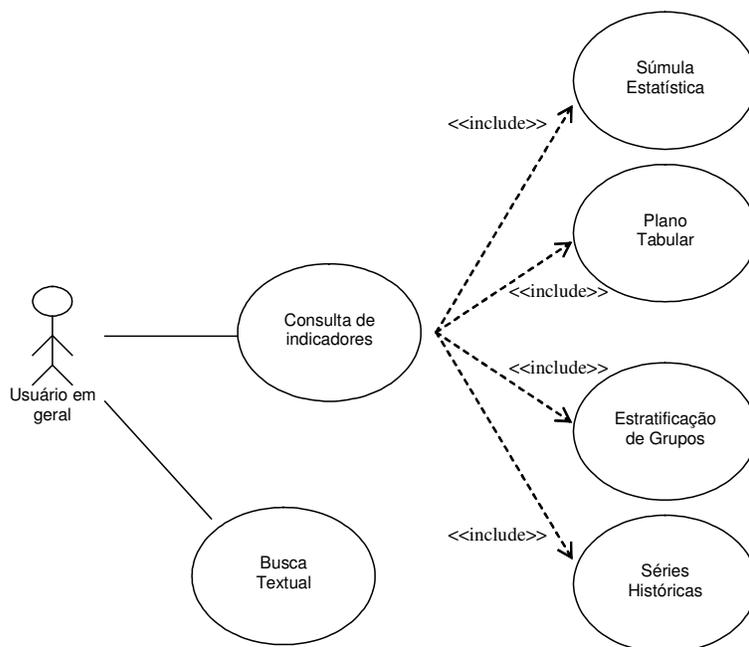


Figura 22 - Diagrama de casos de uso do Data Mart do Diretório de Grupos de Pesquisa

Especificação dos Casos de Uso

Sistema Diretório de Grupos de Pesquisa

Por se tratarem de casos de uso relacionados aos sistemas que manipulam a base de dados operacional do Diretório de Grupos de Pesquisa já implementados e em funcionamento no CNPq, eles são descritos a seguir de forma resumida apenas para que se compreendam as principais funcionalidades e o contexto de cada um.

UC-01: Atualizar dados da instituição

O dirigente cadastra ou atualiza os dados de sua própria instituição, os quais serão utilizados pelo CNPq para que possa entrar em contato com o dirigente institucional.

UC-02: Definir líderes de grupo

Além de visualizar a lista atual de líderes que foram cadastrados pelo dirigente institucional, este pode incluir um novo líder na lista da instituição. Essa inclusão deve ser feita através de uma pesquisa na base do currículo Lattes do CNPq (a presença do currículo do líder nessa base é condição necessária para a sua inclusão). Essa habilitação permitirá aos líderes obter o sistema de cadastro de grupos de pesquisa e enviar o grupo para ser analisado pelo dirigente institucional.

UC-03: Certificar grupo de pesquisa

O dirigente obtém informações sobre os grupos de sua instituição, classificados segundo a grande área e a área do conhecimento predominante nos grupos e de acordo com o órgão e a unidade do grupo. A partir da relação de grupos que estão aguardando certificação pela instituição, onde constam os grupos enviados pelos líderes mas não avaliados pelo dirigente, o dirigente poderá realizar a certificação de

um grupo de pesquisa ou negá-la. Entretanto, somente grupos de pesquisa certificados participam dos censos.

UC-04: Atualizar informações do grupo de pesquisa

O líder de um grupo de pesquisa cadastra e atualiza as informações dos grupos que lidera identificando:

- os pesquisadores e estudantes que possuam currículo Lattes e que devam estar associados ao grupo de pesquisa;
- a relação do pessoal técnico associado ao grupo, indicando o tipo de atividade de cada um;
- a área predominante e as demais áreas de conhecimento às quais o grupo está associado; e
- as linhas de pesquisas do grupo e os setores aos quais elas estão associadas.

UC-05: Atualizar currículo Lattes

Ao atualizarem as informações do currículo Lattes, líderes de grupo, pesquisadores e estudantes associados a um grupo de pesquisa atualizarão também as informações sobre a produção científica e tecnológica relacionada a esse grupo.

UC-06: Listar grupos de que participa

Aos pesquisadores e estudantes é permitido verificar a lista dos grupos de pesquisa de que participam bem como a situação de cada um dos grupos (*status*).

UC-07: Retirar-se de um grupo de pesquisa

Pesquisadores ou estudantes podem ser incorretamente associados a um grupo de pesquisa do qual não participam; nesse caso, ambos podem realizar sua desvinculação desse grupo.

UC-08: Listar estudantes que orienta nos grupos

O pesquisador poderá visualizar uma lista dos estudantes que orienta nos grupos de pesquisa de que participa.

UC-09: Listar orientadores nos grupos

O estudante poderá visualizar seu orientador nos grupos de pesquisa de que participa.

Data Mart do Diretório de Grupos de Pesquisa

Considerando-se que os casos de uso a seguir correspondem a funcionalidades que serão implementadas e suportadas pelos Data Marts do Diretório de Grupos de Pesquisa, esses casos são descritos de forma mais detalhada e associados aos respectivos requisitos, à especificação adicional ou aos termos do glossário, itens necessários para sua rastreabilidade e compreensão.

UC-10: Súmula Estatística

Descrição	O usuário acessa o site para obter indicadores referentes a um determinado censo realizado sobre o Diretório de Grupos de Pesquisa por tratarem-se de indicadores que sintetizam o conteúdo da base de dados e fornecem um retrato bastante nítido da capacidade instalada de pesquisa no País, armazenados de forma estática (HTML sem acesso dinâmico ao banco de dados)	
Cenário principal	<u>Usuário</u> <ul style="list-style-type: none"> • Seleciona o ano referente ao censo • Seleciona uma das unidades de análise disponíveis • Seleciona a opção de unidades de distribuição que deseja visualizar 	<u>Sistema</u> <ul style="list-style-type: none"> • Apresenta os indicadores referentes ao ano do censo e as unidades de análise e de distribuição na forma tabular e gráfica
Precondições	<ul style="list-style-type: none"> • O processo de geração do conteúdo estático dos indicadores que serão disponibilizados no site "Súmula Estatística" deve ter sido realizado previamente 	
Pós-condições	<ul style="list-style-type: none"> • Estatísticas de acesso ao site devem ser atualizadas 	
Itens relacionados	<u>Requisito(s) Funcional(is)</u> <ul style="list-style-type: none"> • Visualização de indicadores no site "Súmula Estatística" 	
Observações	<ul style="list-style-type: none"> • Tendo em vista a natureza estática do conteúdo das páginas apresentadas neste site, poderá haver divergência entre os indicadores apresentados e os mesmos indicadores obtidos no site "Plano Tabular" para o mesmo ano de referência, caso este último tenha sofrido qualquer tipo de atualização e o conteúdo do site "Súmula Estatística" não tenha sido recriado 	

Quadro 7 - Especificação do caso de uso "Súmula Estatística"

UC-11: Consulta de Indicadores

Descrição	O usuário acessa o site de divulgação das informações armazenadas no Data Mart do Diretório de Grupos de Pesquisa para obter indicadores atuais ou históricos ou para realizar busca de informações referentes aos grupos de pesquisas cadastrados	
Cenário principal	<u>Usuário</u> <ul style="list-style-type: none"> Acessa o site principal do Data Mart do Diretório de Grupos de Pesquisa Seleciona a opção de busca ou de consulta de indicadores entre as cinco aplicações disponíveis 	<u>Sistema</u> <ul style="list-style-type: none"> Apresenta as opções de consulta de indicadores possíveis e a busca textual Desvia a navegação do site principal para o site específico correspondente à aplicação selecionada pelo usuário
Precondições	<ul style="list-style-type: none"> O Data Mart do Diretório de Grupos de Pesquisa deve estar atualizado e ativo, ou seja, não pode estar sofrendo um processo de carga durante as consultas para evitar problemas no tempo de resposta ou inconsistência dos indicadores ou da busca 	
Pós-condições	<ul style="list-style-type: none"> “nenhuma” 	
Itens relacionados	<u>Caso(s) de Uso</u> <ul style="list-style-type: none"> Séries Históricas Súmula Estatística Plano Tabular Busca Textual Estratificação de Grupos 	
Observações	<ul style="list-style-type: none"> O site principal de acesso às informações do Data Mart do Diretório de Grupos de Pesquisa serve apenas como um centralizador dos diferentes tipos de informações que podem ser visualizadas ou pesquisadas pelos usuários com as aplicações descritas nos casos de uso mencionados no campo “Referências” 	

Quadro 8 - Especificação do caso de uso “Consulta de indicadores”

UC-12: Séries Históricas

Descrição	O usuário acessa o site para obter indicadores comparativos entre os censos realizados sobre o Diretório de Grupos de Pesquisa por tratarem-se de indicadores que sintetizam o conteúdo da base de dados e fornecem um retrato evolutivo bastante nítido da pesquisa no País, armazenados de forma estática (HTML sem acesso dinâmico ao banco de dados)	
Cenário principal	<u>Usuário</u> <ul style="list-style-type: none"> • Seleciona uma das unidades de análise disponíveis • Seleciona a opção de unidades de distribuição que deseja visualizar 	<u>Sistema</u> <ul style="list-style-type: none"> • Apresenta os indicadores referentes a cada um dos censos realizados na forma tabular, conforme as unidades de análise e de distribuição selecionadas
Precondições	<ul style="list-style-type: none"> • O processo de geração do conteúdo estático dos indicadores que serão disponibilizados no site “Séries Históricas” deve ter sido realizado previamente 	
Pós-condições	<ul style="list-style-type: none"> • As estatísticas de acesso ao site devem ser atualizadas 	
Itens relacionados	<u>Requisito(s) Funcional(is)</u> <ul style="list-style-type: none"> • Visualização de indicadores no site “Séries Históricas” 	
Observações	<ul style="list-style-type: none"> • Tendo em vista a natureza estática do conteúdo das páginas apresentadas neste site, poderá haver divergência entre os indicadores apresentados e os mesmos indicadores obtidos no site “Plano Tabular”, caso este último tenha sofrido qualquer tipo de atualização ou de correção e o conteúdo do site “Séries Históricas” não tenha sido recriado 	

Quadro 9 - Especificação do caso de uso “Séries Históricas”

UC-13: Plano Tabular

Descrição	O usuário acessa o site para obter indicadores referentes a um determinado censo realizado sobre o Diretório de Grupos de Pesquisa, calculados de forma dinâmica (com acesso ao banco de dados para utilizar os dados mais atualizados possíveis)	
Cenário principal	<u>Usuário</u> <ul style="list-style-type: none"> • Seleciona o ano referente ao censo (ou “Todos”, para obter uma série histórica evolutiva dos censos) • Seleciona uma das unidades de análise disponíveis • Seleciona uma ou mais unidades de conteúdo entre as disponíveis para visualização • Seleciona a(s) unidade(s) de distribuição que deseja utilizar como filtro(s) e define os valores ou a faixa de valores a serem filtrados • Seleciona a(s) unidade(s) de distribuição que deseja utilizar como agrupador(es) horizontal(is) ou vertical(is)¹ • Seleciona a unidade de classificação a ser utilizada para ordenar o resultado de forma crescente ou decrescente • Pode realizar uma das operações OLAP disponíveis para ser aplicada no resultado da consulta 	<u>Sistema</u> <ul style="list-style-type: none"> • Apresenta os indicadores referentes ao ano do censo e às unidades de análise, de filtro, de agrupamento e de classificação selecionadas na forma tabular
Precondições	<ul style="list-style-type: none"> • A carga de dados referentes ao ano do censo selecionado deve ter sido realizada • A base do Data Mart não pode estar em processo de carga de dados 	
Pós-condições	<ul style="list-style-type: none"> • As estatísticas de acesso ao site devem ser atualizadas 	
Itens relacionados	<u>Requisito(s) Funcional(is)</u> <ul style="list-style-type: none"> • Visualização de indicadores no site “Plano Tabular” • Visualização do resultado da “Busca Textual” 	
Observações	<ul style="list-style-type: none"> • ¹Em séries históricas a distribuição vertical será fixa de acordo com a unidade de distribuição “Ano do censo” • Tendo em vista a natureza dinâmica do conteúdo das páginas apresentadas neste site, poderá haver divergência entre os indicadores apresentados e os mesmos indicadores obtidos no sites “Séries Históricas” e “Súmula Estatística” para o mesmo ano de referência, caso este último tenha sofrido qualquer tipo de atualização e o conteúdo desses sites não tenha sido recriado • Ao visualizar o resultado de uma consulta, o usuário poderá clicar sobre o valor dos indicadores para ver uma lista referente a grupos, linhas, pesquisadores, estudantes, técnicos ou produções (esta somente no censo mais atual), os quais foram utilizados para compor o valor desse indicador, sendo essa lista apresentada no mesmo formato usado na apresentação do resultado da busca textual 	

Quadro 10 - Especificação do caso de uso “Plano Tabular”

UC-14: Busca Textual

Descrição	O usuário acessa o site para obter informações referentes a uma das unidades de análise possíveis, localizadas através de critérios definidos por ele mesmo	
Cenário principal	<u>Usuário</u> <ul style="list-style-type: none"> • Seleciona o ano referente ao censo ou o censo atual • Seleciona a unidade de análise sobre a qual deseja realizar a busca textual • Informa as palavras-chave e os critérios específicos de seleção • Clica sobre um dos itens resultantes da pesquisa para obter detalhes sobre um determinado grupo, linha de pesquisa, pesquisador, estudante, técnico ou produção 	<u>Sistema</u> <ul style="list-style-type: none"> • Utiliza-se de algoritmos específicos de busca para localizar as informações que satisfaçam os critérios e as palavras-chave especificadas pelo usuário • Apresenta o resultado da busca textual na forma de uma lista de grupos, linhas de pesquisa, pesquisadores, estudantes, técnicos ou produções localizadas • Apresenta um formulário detalhando o resultado da busca textual para um dos registros de grupos, linhas de pesquisa, pesquisadores, estudantes, técnicos ou produções encontradas
Precondições	<ul style="list-style-type: none"> • Verificar as condições a serem estabelecidas pelos algoritmos utilizados para a implementação da busca textual propriamente dita e da visualização do formulário com informações detalhadas sobre grupos, linhas de pesquisa, pesquisadores, estudantes, técnicos e produções 	
Pós-condições	<ul style="list-style-type: none"> • As estatísticas de acesso ao site devem ser atualizadas 	
Itens relacionados	<u>Requisito(s) Funcional(is)</u> <ul style="list-style-type: none"> • Visualização do resultado da “Busca Textual” 	
Observações	<ul style="list-style-type: none"> • A implementação dos algoritmos de busca textual e de visualização dos formulários contendo informações detalhadas sobre grupos, linhas de pesquisa, pesquisadores, estudantes, técnicos ou produções não será suportada diretamente pelo Data Mart do Diretório de Grupos de Pesquisa 	

Quadro 11 - Especificação do caso de uso “Busca Textual”

Apesar de estar citado como um requisito que deve ser atendido pelo Data Mart do Diretório de Grupos de Pesquisa, o site da busca textual não utilizará diretamente as informações sumarizadas do Data Mart para realizar uma busca através do uso de palavras-chave ou de critérios específicos de seleção. As bases de dados dimensionais não são projetadas para dar suporte a esse tipo de pesquisa, e outros mecanismos de buscas devem ser implementados com esse fim. Entretanto, no site de indicadores “Plano Tabular” os usuários terão à disposição uma funcionalidade

através da qual será possível clicar sobre o valor dos indicadores para visualizar uma lista de grupos, linhas de pesquisa, pesquisadores, estudantes, técnicos ou produções (esta somente no censo mais atual) que foram utilizados para compor o valor desse indicador. A lista é mostrada seguindo o mesmo formato utilizado na apresentação do resultado da busca textual.

UC-15: Estratificação de Grupos

Descrição	O usuário acessa o site para obter indicadores referentes à classificação dos grupos de pesquisa vinculados às Instituições de Ensino Superior (IES) segundo os critérios de estratificação definidos pelo CNPq. Esses indicadores são armazenados de forma estática (HTML sem acesso dinâmico ao banco de dados)	
Cenário principal	<u>Usuário</u> <ul style="list-style-type: none"> • Seleciona o ano referente ao censo • Seleciona o tipo de indicador • Seleciona a tabela de indicadores que deseja visualizar 	<u>Sistema</u> <ul style="list-style-type: none"> • Apresenta os indicadores para o censo e tipo de indicador selecionado
Precondições	<ul style="list-style-type: none"> • O processo de geração do conteúdo estático dos indicadores a serem disponibilizados no site “Estratificação de Grupos” deve ter sido realizado previamente 	
Pós-condições	<ul style="list-style-type: none"> • As estatísticas de acesso ao site devem ser atualizadas 	
Itens relacionados	<u>Requisito(s) Funcional(is)</u> <ul style="list-style-type: none"> • Visualização de Indicadores no site “Estratificação de Grupos” 	
Observações	<ul style="list-style-type: none"> • Os algoritmos de estratificação de grupos não serão implementados diretamente pelo Data Mart do Diretório de Grupos de Pesquisa, ficando a base de dados do Data Mart responsável apenas por armazenar e disponibilizar o resultado da estratificação para cada um dos censos 	

Quadro 12 - Especificação do caso de uso “Estratificação de Grupos”

4.2.3 Especificação de requisitos

Requisitos de informação

Os requisitos abaixo descrevem as informações que poderão ser utilizadas como unidade de análise, unidade de conteúdo e unidade de distribuição (filtro, agrupamento ou classificação – ver definição no glossário) na visualização de indicadores e na busca textual.

RQI-01: Unidades de Análise e Unidades de Conteúdo

A seguir são apresentadas as unidades de análise e de conteúdo a serem disponibilizadas no Data Mart de Grupos de Pesquisa e que devem ser consideradas para a construção das tabelas de fato do modelo dimensional.

Unidade de Análise	Unidades de Conteúdo
Grupos de pesquisa	<ul style="list-style-type: none"> • <u>Grupos de pesquisa</u>: quantidade de grupos de pesquisa distintos segundo os critérios considerados • <u>Linhas de pesquisa</u>: quantidade de linhas de pesquisa associadas aos grupos de pesquisa considerados • <u>Pesquisadores</u>: quantidade de pesquisadores associados aos grupos de pesquisa considerados (quantidade distinta ou não dependendo da aplicação, já que o mesmo pesquisador pode participar de mais de um grupo de pesquisa) • <u>Estudantes</u>: quantidade de estudantes associados aos grupos de pesquisa considerados (quantidade distinta ou não dependendo da aplicação, já que o mesmo estudante pode participar de mais de um grupo de pesquisa) • <u>Técnicos</u>: quantidade de técnicos associados aos grupos de pesquisa considerados • <u>Relação L/G</u>: quantidade média de linhas de pesquisa associadas a cada grupo de pesquisa • <u>Relação P/G</u>: quantidade média de pesquisadores associados a cada grupo de pesquisa • <u>Relação E/G</u>: quantidade média de estudantes associados a cada grupo de pesquisa • <u>Relação P/L</u>: quantidade média de pesquisadores associados a cada linha de pesquisa de cada um dos grupos

Unidade de Análise	Unidades de Conteúdo
Linhas de pesquisa	<ul style="list-style-type: none"> • <u>Linhas de pesquisa</u>: quantidade de linhas de pesquisa segundo os critérios considerados • <u>Pesquisadores</u>: quantidade de pesquisadores associados às linhas de pesquisa consideradas (quantidade distinta ou não dependendo da aplicação, já que o mesmo pesquisador pode participar de mais de uma linha de pesquisa do mesmo grupo ou de grupos distintos) • <u>Grupos de pesquisa</u>: quantidade de grupos de pesquisa aos quais as linhas de pesquisa consideradas estão associadas • <u>Relação L/G</u>: quantidade média de linhas de pesquisa associadas a cada grupo de pesquisa • <u>Relação L/P</u>: quantidade média de linhas de pesquisa associadas a cada pesquisador

Unidade de Análise	Unidade de Conteúdo
Pesquisadores	<ul style="list-style-type: none"> • <u>Pesquisadores</u>: quantidade de pesquisadores segundo os critérios considerados

Unidade de Análise	Unidade de Conteúdo
Estudantes	<ul style="list-style-type: none"> • <u>Estudantes</u>: quantidade de estudantes segundo os critérios considerados

Unidade de Análise	Unidade de Conteúdo
Pessoal Técnico	<ul style="list-style-type: none"> • <u>Pessoal técnico</u>: quantidade de pessoal técnico segundo os critérios considerados

Unidade de Análise	Unidade de Conteúdo
Produção C&T	<ul style="list-style-type: none"> <u>Produção C&T</u>: quantidade de produção (considerando-se a produção dos quatro últimos anos anteriores ao censo realizado, incluindo o ano do censo) dos pesquisadores e estudantes associados aos grupos de pesquisa segundo os critérios considerados

Unidade de Análise	Unidade de Conteúdo
Empresas	<ul style="list-style-type: none"> <u>Empresas</u>: quantidade de empresas associadas aos grupos de pesquisa segundo os critérios considerados

Quadro 13 - Unidades de Análise e de Conteúdo do Data Mart de Grupos de Pesquisa

Além das unidades de conteúdo citadas, que serão usadas diretamente na construção dos valores dos indicadores de uma determinada consulta, outras características são de interesse dos usuários para se construírem indicadores. Essas características representam algumas propriedades das unidades de análise e podem ser utilizadas tanto como unidades de conteúdo como unidades de filtro.

Unidades de Análise	Propriedades
Grupos de pesquisa	<ul style="list-style-type: none"> <u>Grupo unitário</u>: grupo de pesquisa com apenas 1 (um) pesquisador e nenhum estudante ou técnico associados <u>Grupo sem estudante</u>: grupo de pesquisa com pesquisadores ou técnicos, mas nenhum estudante associado <u>Grupo sem técnico</u>: grupo de pesquisa com pesquisadores ou estudantes, mas nenhum técnico associado <u>Grupo com mais de 10 pesquisadores</u>: grupo de pesquisa com número total de pesquisadores associados superior a 10 <u>Grupo com mais de 10 linhas</u>: grupo de pesquisa com número total de linhas de pesquisa superior a 10 <u>Grupo sem líder doutor</u>: grupo de pesquisa cuja titulação máxima do 1º líder é inferior a "Doutorado" <u>Grupo sem doutor</u>: grupo de pesquisa que não apresenta pesquisador ou estudante associado com titulação máxima igual ou superior a "Doutorado"
Linhas de pesquisa	<ul style="list-style-type: none"> <u>Linha unitária</u>: linha de pesquisa com apenas 1 (um) pesquisador e nenhum estudante ou técnico associados <u>Linha sem estudante</u>: linha de pesquisa com pesquisadores ou técnicos, mas nenhum estudante associado <u>Linha com mais de 10 pesquisadores</u>: linha de pesquisa com número total de pesquisadores associados superior a 10 <u>Linha sem doutor</u>: linha de pesquisa que não apresenta pesquisador ou estudante associado com titulação máxima igual ou superior a "Doutorado"

Quadro 14 - Propriedades adicionais das Unidades de Análise

Serão também consideradas as seguintes unidades de conteúdo utilizadas na visualização dos indicadores de estratificação:

Unidades de Conteúdo	Descrição	Variações
Grupos IES	grupos vinculados a instituições de ensino superior	<ul style="list-style-type: none"> • quantidade absoluta • % relativo ao total
Doutores	número de pesquisadores com titulação máxima igual a doutorado	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total
BPQ	bolsistas de pesquisa do CNPq nas modalidades PQ	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total • relação BPQ/grupo • relação BPQ/doutor
BPQ (I)	parte de BPQ constituído pelos bolsistas de categoria I (IA,IB,IC)	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total
BPQ (II)	parte de BPQ constituído pelos bolsistas de categoria II (IIA,IIB,IIC)	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total
DOC	docentes com titulação máxima igual a doutorado, vinculados a programas de pós-graduação reconhecidos e avaliados pela CAPES com grau igual ou superior a 3 (avaliação do ano anterior ao do censo)	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total • relação DOC/grupo
DOC(6;7)	subconjunto de DOC, formado por docentes vinculados a programas de pós-graduação avaliados com grau 6 ou 7	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total
DOC(5)	subconjunto de DOC, constituído por docentes vinculados a programas de pós-graduação avaliados com grau 5	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total
DOC(3;4)	subconjunto de DOC, integrado por docentes vinculados a programas de pós-graduação avaliados com grau 3 ou 4	<ul style="list-style-type: none"> • quantidade absoluta • quantidade média • % relativo ao total
Produção C&T	quantidade de cada tipo e subtipo de produção de pesquisadores e estudantes considerando os últimos quatro anos anteriores ao do censo (inclusive o ano do censo atual)	<ul style="list-style-type: none"> • quantidade absoluta
Q	escore de qualificação	<ul style="list-style-type: none"> • média
P	escore de produtividade	<ul style="list-style-type: none"> • média

Quadro 15 - Unidades de Conteúdo do site “Estratificação de Grupos”

RQI-02: Dimensão “Distribuição Geográfica”

Dimensão que provê informações sobre as instituições participantes do Diretório de Grupos de Pesquisa e sobre a distribuição geográfica dessas instituições em regiões e Estados.

Unidades de Distribuição	Descrição	Domínio
+ Região	Nome da região	<ul style="list-style-type: none"> • Centro-Oeste • Norte • Nordeste • Sul • Sudeste
+ UF	Unidade da Federação	<ul style="list-style-type: none"> • Relação das 27 Unidades da Federação (incluindo Distrito Federal)
■ Instituição	Nome da instituição	<ul style="list-style-type: none"> • Instituições participantes do Diretório de Grupos de Pesquisa

Tabela 1 - Unidades de distribuição da dimensão “Distribuição Geográfica”

RQI-03: Dimensão “Árvore do Conhecimento”

Dimensão que provê informações sobre as áreas de conhecimento e os setores de aplicação aos quais estão associados os grupos de pesquisa, os pesquisadores e suas produções. Os setores de aplicação são considerados apenas para a análise de linhas de pesquisa, uma vez que os grupos de pesquisa não possuem relacionamento com essa hierarquia.

Unidades de Distribuição	Descrição	Domínio
+ Grupo de grandes áreas	Agrupamento de grandes áreas de conhecimento	<ul style="list-style-type: none"> • Ciências da Natureza • Ciências da Vida • Humanidades
+ Grande área	Nome da grande área de conhecimento predominante do grupo de pesquisa	<ul style="list-style-type: none"> • Ciências Agrárias, Ciências Biológicas, Ciências Exatas e da Terra, Ciências Humanas, Ciências Sociais Aplicadas, Ciências da Saúde, Engenharias, Linguística, Letras e Artes
■ Área	Nome da grande área de conhecimento predominante do grupo de pesquisa	<ul style="list-style-type: none"> • Relação das áreas de conhecimento associadas a cada grande área
■ Setor de aplicação	Descrição do setor de aplicação	<ul style="list-style-type: none"> • Relação dos setores de aplicação (em torno de 45 ocorrências)

Tabela 2 - Unidades de Distribuição da dimensão “Área do conhecimento”

RQI-04: Dimensão “Censo”

Pelo fato de a realização dos censos ter uma periodicidade bianual, a identificação deles será feita através do ano de sua realização. Entretanto, a dimensão deve também considerar o censo atual, que, ao contrário de ser uma visão estática de um ano específico, é a visão dinâmica do momento presente, atualizada periodicamente.

Unidades de Distribuição	Descrição	Domínio
■ Ano do censo	Ano de realização do censo	Para o ano do censo os valores a serem considerados inicialmente são: <ul style="list-style-type: none"> • 2000 • 2002 • 2004 • Atual

Tabela 3 - Unidades de Distribuição da dimensão “Censo”

RQI-05: Dimensão “Características Pessoais”

Dimensão que provê informações sobre as características dos pesquisadores e dos estudantes associados aos grupos de pesquisa. O pessoal técnico não é considerado nesta dimensão por não haver a exigência de possuírem “Currículo Lattes” cadastrado na base de dados do CNPq. As informações referentes às unidades de distribuição da dimensão “Características Pessoais” devem ser mantidas de forma histórica para que não sejam perdidas com a evolução temporal dos grupos de pesquisa e dos próprios integrantes de cada grupo.

Unidades de Distribuição	Descrição	Domínio
<input type="checkbox"/> Faixa etária	Faixa etária dos pesquisadores e estudantes associados a um grupo de pesquisa	<ul style="list-style-type: none"> • Até 19” • De [20-24] a [60-64] com intervalos fixos de 5 em 5 anos • [65 ou mais] • Não informado
<input type="checkbox"/> Nível de treinamento	Nível da formação em andamento dos estudantes ou titulação máxima dos pesquisadores	<ul style="list-style-type: none"> • Graduação • Especialização • Mestrado • Doutorado • Não informado
<input type="checkbox"/> Sexo	Sexo dos pesquisadores e estudantes de um grupo de pesquisa	<ul style="list-style-type: none"> • Masculino • Feminino • Não informado
<input checked="" type="checkbox"/> Nacionalidade	Nacionalidade dos pesquisadores e estudantes de um grupo de pesquisa	<ul style="list-style-type: none"> • Brasileira • Estrangeira
<input type="checkbox"/> País de origem	País de origem dos pesquisadores e estudantes de um grupo de pesquisa	<ul style="list-style-type: none"> • Brasil para a nacionalidade Brasileira • Nome do país para os de nacionalidade Estrangeira
<input type="checkbox"/> Liderança	Status de liderança de grupo de pesquisa dos pesquisadores	<ul style="list-style-type: none"> • Líder • Não líder

Tabela 4 - Unidades de Distribuição da dimensão “Características Pessoais”

RQI-06: Dimensão “Tipo de produção”

Classifica a produção dos pesquisadores e estudantes associados aos grupos de pesquisa em tipos e subtipos, conforme a tabela de domínios abaixo.

Unidades de Distribuição	Descrição	Domínio
<input checked="" type="checkbox"/> Tipo	Tipo de produção	<ul style="list-style-type: none"> • Produção bibliográfica • Produção técnica • Orientações concluídas • Orientações em andamento • Produção artística, cultural e demais
<input type="checkbox"/> Subtipo	Subtipo de produção	<ul style="list-style-type: none"> • Subtipos de produção para cada um dos tipos de produção citados

Tabela 5 - Unidades de Distribuição da dimensão “Tipo de produção”

RQI-07: Dimensão “Atividade Técnica”

Representa uma unidade de distribuição que determina o tipo de atividade técnica do pessoal técnico (sem currículo Lattes) associado a um grupo de pesquisa pelo líder do grupo.

Unidade de Distribuição	Descrição	Domínio
■ Atividade técnica	Tipo de atividade técnica desempenhada por cada indivíduo da relação de pessoal técnico associada a um grupo de pesquisa	<ul style="list-style-type: none"> Relação dos tipos de atividade técnica definidos no Sistema Diretório de Grupos de Pesquisa (em torno de 45 ocorrências)

Tabela 6 - Unidades de Distribuição da dimensão “Atividade técnica”

RQI-08: Dimensão: “Empresa”

Representa características referentes a empresas que possuem algum tipo de relacionamento com o grupo de pesquisa. Essa associação entre grupos e empresas foi introduzida a partir do Censo 2002 e é informada pelo líder do grupo.

Unidades de Distribuição	Descrição	Domínio
☐ Natureza jurídica	Natureza jurídica da empresa	<ul style="list-style-type: none"> Administração Pública Empresas Públicas e de Economia Mista Empresas Privadas Entidades sem Fins Lucrativos Situação Jurídica Ignorada
■ Empresa	Nome da empresa	<ul style="list-style-type: none"> Razão Social das empresas que possuem relacionamento com um grupo de pesquisa
■ Número de funcionários	Faixa de número de funcionários atuantes na empresa	<ul style="list-style-type: none"> Até 19 20 a 49 50 a 99 100 a 499 500 ou mais Ignorado

Tabela 7 - Unidades de Distribuição da dimensão “Empresa”

RQI-09: Dimensão “Tipo de relacionamento”

Representa o tipo de relacionamento existente entre grupos de pesquisa e empresas, e cada empresa associada a um grupo de pesquisa pode ter até três tipos de relacionamento indicados pelo líder do grupo.

Unidade de Distribuição	Descrição	Domínio
■ Tipo de relacionamento	Tipo de relacionamento entre empresas e grupos de pesquisa	<ul style="list-style-type: none"> Relação dos 14 tipos de relacionamento possíveis entre empresas e grupos de pesquisa

Tabela 8 - Unidades de Distribuição da dimensão “Tipo de relacionamento”

RQI-10: Dimensão “Forma de remuneração”

Representa a forma de remuneração existente entre grupos de pesquisa e empresas, e cada empresa associada a um grupo de pesquisa pode ter até três formas de remuneração indicadas pelo líder do grupo.

Unidade de Distribuição	Descrição	Domínio
■ Forma de remuneração	Forma de remuneração entre empresas e grupos de pesquisa	<ul style="list-style-type: none"> Relação das 10 formas de remuneração possíveis entre empresas e grupos de pesquisa

Tabela 9 - Unidades de Distribuição da dimensão “Forma de remuneração”

RQI-11: Dimensão “Setor da atividade econômica”

Representa os setores econômicos aos quais uma empresa pode estar associada. Cada empresa associada a um grupo de pesquisa pode ter até três setores econômicos indicados pelo líder do grupo.

Unidades de Distribuição	Descrição	Domínio
<ul style="list-style-type: none"> ⊕ Atividade econômica <ul style="list-style-type: none"> ⊕ Divisão <ul style="list-style-type: none"> ■ Grupo 	<ul style="list-style-type: none"> Nível 1 do setor da atividade econômica das empresas associadas a grupos de pesquisa Nível 2 do setor da atividade econômica das empresas associadas a grupos de pesquisa Nível 3 do setor da atividade econômica das empresas associadas a grupos de pesquisa 	<ul style="list-style-type: none"> Hierarquia dos setores de atividade econômicos considerados

Tabela 10 - Unidades de Distribuição da dimensão “Setor da atividade econômica”

RQI-12: Relação entre as unidades de análise e dimensões

As dimensões relacionadas com cada uma das unidades de análise determinam as possíveis unidades de distribuição que poderão ser utilizadas como filtro ou variável de agrupamento na visualização de indicadores ou na busca textual. Essas relações estão especificadas na tabela abaixo.

Dimensões	Unidades de Análise						
	Grupo	Linha	Pesquisador	Estudante	Pessoal Técnico	Produção C&T	Empresa
Censo	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Empresa							Sim
Distribuição Geográfica	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Árvore do Conhecimento	Sim	Sim ¹	Sim ²	Sim ²	Sim ³	Sim ²	Sim ³
Características Pessoais			Sim	Sim			
Atividade Técnica					Sim		
Tipo de Produção						Sim	
Tipo de Relacionamento							Sim
Forma de Remuneração							Sim
Setor Econômico							Sim

Quadro 16 - Dimensões utilizadas por cada unidade de análise

¹ Para a unidade de análise “Linhas de Pesquisa” a dimensão “Árvore do Conhecimento” poderá ser utilizada sob dois aspectos: um referente à área predominante do grupo de pesquisa ao qual a linha de pesquisa está associada e outro referente à área da própria linha de pesquisa.

² Para as unidade de análise “Pesquisadores”, “Estudantes” e “Produção C&T” a dimensão “Árvore do Conhecimento” poderá ser utilizada sob dois aspectos: um referente à área predominante do grupo de pesquisa ao qual o pesquisador ou o estudante está associado e outro referente à área das linhas de pesquisa dos grupos às quais possam estar associados – neste último caso, a hierarquia “Setor de Aplicação” poderá também ser considerada.

³ Para as unidades de análise “Pessoal Técnico” e “Empresas” a dimensão “Árvore do Conhecimento” poderá ser utilizada somente no tocante à área predominante do grupo de pesquisa ao qual o técnico está associado.

Requisitos Funcionais

Os requisitos abaixo descrevem funcionalidades visíveis ao usuário final das aplicações que utilizarão o Data Mart do Diretório de Grupos de Pesquisa como fonte de dados para a visualização de indicadores.

RQF-01: Visualização de indicadores no site “Séries Históricas”

No site “Séries Históricas” serão publicados alguns indicadores gerados estaticamente a partir das informações disponíveis no Data Mart do Diretório de Grupos de Pesquisa contendo informações comparativas de todos os censos anteriores, de forma que as análises sejam rapidamente apresentadas para os usuários. Nas definições de indicadores para o site “Séries Históricas” deve-se considerar o ano do censo como unidade de distribuição horizontal. As demais unidades de distribuição mencionadas devem ser consideradas como unidades de distribuição verticais, desde que não haja qualquer observação ao contrário. Os indicadores disponíveis para cada uma das unidades de análise são descritos a seguir.

Séries Básicas

Consulta	Descrição do resultado
Principais dimensões por censo	<u>Unidade de Distribuição</u> <ul style="list-style-type: none"> Cada uma das unidades de conteúdo <u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> Quantidade de instituições Quantidade de grupos Quantidade de pesquisadores Quantidade de pesquisadores doutores % da quantidade de pesquisadores doutores em relação à quantidade de pesquisadores <u>Unidade de Classificação</u> <ul style="list-style-type: none"> Nenhuma <u>Apresentação do resultado</u> <ul style="list-style-type: none"> Tabular

Pesquisadores doutores por grande área	<u>Unidade de Distribuição</u> <ul style="list-style-type: none"> Grande área do conhecimento predominante do grupo <u>Unidade de Conteúdo</u> <ul style="list-style-type: none"> Quantidade de pesquisadores doutores <u>Unidade de Classificação</u> <ul style="list-style-type: none"> Quantidade de pesquisadores doutores do último censo (decrecente) <u>Apresentação do resultado</u> <ul style="list-style-type: none"> Tabular
Pesquisadores por sexo (%)	<u>Unidade de Distribuição</u> <ul style="list-style-type: none"> Sexo do pesquisador <u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> % da quantidade de pesquisadores em relação ao total destes Idade mediana dos pesquisadores segundo o sexo <u>Unidade de Classificação</u> <ul style="list-style-type: none"> Sexo do pesquisador <u>Apresentação do resultado</u> <ul style="list-style-type: none"> Tabular para cada uma das unidades de conteúdo <u>Totalizações</u> <ul style="list-style-type: none"> Idade mediana global
Grupos por região (%)	<u>Unidade de Distribuição</u> <ul style="list-style-type: none"> Região <u>Unidade de Conteúdo</u> <ul style="list-style-type: none"> % da quantidade de grupos segundo a região em relação ao total de grupos de cada censo <u>Unidade de Classificação</u> <ul style="list-style-type: none"> % da quantidade de grupos do último censo (decrecente) <u>Apresentação do resultado</u> <ul style="list-style-type: none"> Tabular <u>Totalizações</u> <ul style="list-style-type: none"> Fixa em 100% para cada censo

Tabela 11 - Indicadores do site "Séries Históricas" (séries básicas)

Unidade de Análise: Grupos de pesquisa

Unidades de Distribuição	Descrição do resultado
<ul style="list-style-type: none"> • Região • Unidade da Federação • Instituição • Ano de criação • Grande área do conhecimento • Área do conhecimento 	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de grupos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo <p><u>Unidades de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de grupos do último censo (decrecente) <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total das quantidades de grupos • Fixa em 100% para os % de grupos <p><u>Observação</u></p> <ul style="list-style-type: none"> • Esta definição de resultado é comum para todas as unidades de distribuição
Ano de criação (*)	<p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Apresentar média e mediana do número de anos de existência dos grupos de pesquisa
Grande área do conhecimento (*)	<p><u>Unidades de agrupamento</u></p> <ul style="list-style-type: none"> • Dividir as grandes áreas em três grandes grupos de grandes áreas: a) Ciências da Natureza (“Engenharias e Ciência da Computação” e “Ciências Exatas”); b) Ciências da Vida (“Ciências da Saúde”, “Ciências Biológicas” e “Ciências Agrárias”); c) Humanidades (“Ciências Humanas”, “Ciências Sociais Aplicadas” e “Linguística, Letras e Artes”)

(*) considerações específicas para uma determinada unidade de distribuição

Tabela 12 - Indicadores do site “Séries Históricas” disponíveis para “Grupos de Pesquisa”

Unidade de Análise: Linhas de pesquisa

Unidades de Distribuição	Descrição do resultado
Número por grupo de pesquisa (domínio de valores de “1” a “9” e “10 ou +” linhas por grupo)	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de grupos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • A própria unidade de distribuição (crescente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total de grupos por censo • Quantidade média de linhas de pesquisa por grupo, em cada censo • Quantidade mediana de linhas de pesquisa por grupo, em cada censo <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular

Setor de aplicação	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de linhas de pesquisa segundo a unidade de distribuição • % da quantidade de linhas de pesquisa segundo a unidade de distribuição em relação ao total de linhas de pesquisa de cada censo <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de linhas de pesquisa do último censo (decrecente) <p><u>Apresentação do resultado:</u></p> <ul style="list-style-type: none"> • Tabular
--------------------	--

Tabela 13 - Indicadores do site “Séries Históricas” disponíveis para “Linhas de pesquisa”

Unidade de Análise: Pesquisadores

Unidades de Distribuição	Descrição do resultado
Número por grupo de pesquisa (domínio de valores de “1” a “12” e “13 ou +” pesquisadores por grupo)	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de grupos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • A própria unidade de distribuição (crescente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total de grupos por censo • Quantidade média de pesquisadores por grupo, em cada censo • Quantidade mediana de pesquisadores por grupo, em cada censo <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular
Titulação máxima do pesquisador	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de pesquisadores segundo a unidade de distribuição • % da quantidade de pesquisadores segundo a unidade de distribuição em relação ao total de pesquisadores de cada censo <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de pesquisadores do último censo (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total de pesquisadores por censo • Total do % de pesquisadores por censo (deve ser 100%) <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular
Titulação máxima e liderança	<ul style="list-style-type: none"> • Idem à unidade de distribuição “”, com a apresentação do resultado, entretanto, dividida em duas tabelas, sendo uma para pesquisadores líderes de grupo e outra para pesquisadores não líderes.

Faixa etária

Unidades de Conteúdo

- Quantidade de pesquisadores segundo a unidade de distribuição
- Quantidade de pesquisadores líderes segundo a unidade de distribuição
- Quantidade de pesquisadores não líderes segundo a unidade de distribuição
- % da quantidade de pesquisadores líderes segundo a unidade de distribuição em relação à quantidade total de pesquisadores de acordo com a unidade de distribuição
- % da quantidade de pesquisadores não líderes segundo a unidade de distribuição em relação à quantidade de pesquisadores de acordo com a unidade de distribuição

Unidade de Classificação

- Faixa etária (crescente)

Totalizações

- Total de pesquisadores por censo
- Total de pesquisadores líderes por censo
- Total de pesquisadores não líderes por censo
- Idade mediana dos pesquisadores, dos pesquisadores líderes e dos pesquisadores não líderes por censo
- % dos totais de pesquisadores líderes e não líderes por censo

Apresentação do resultado

- Tabular para os indicadores de quantidade
- Tabular para os indicadores percentuais

Liderança e sexo

Unidade de Distribuição Vertical

- Liderança

Unidades de Conteúdo

- Quantidade de pesquisadores masculinos segundo a unidade de distribuição
- Quantidade de pesquisadores femininos segundo a unidade de distribuição
- % da quantidade de pesquisadores masculinos segundo a unidade de distribuição em relação à quantidade de pesquisadores de acordo com a unidade de distribuição, por censo
- % da quantidade de pesquisadores femininos segundo a unidade de distribuição em relação à quantidade de pesquisadores de acordo com a unidade de distribuição, por censo
- % da quantidade de pesquisadores masculinos segundo a unidade de distribuição em relação ao total de pesquisadores masculinos do censo
- % da quantidade de pesquisadores femininos segundo a unidade de distribuição em relação ao total de pesquisadores femininos do censo

Unidade de Classificação

- Liderança (ordem alfabética)

Apresentação do resultado

- Tabular para os indicadores de quantidade
- Tabular para os percentuais em relação ao total por condição de liderança
- Tabular para os percentuais em relação ao total de cada sexo

<ul style="list-style-type: none"> • Região • UF • Instituição • (*) a definição do resultado é válida para os três indicadores 	<u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> • Quantidade de pesquisadores, considerando dupla contagem, segundo a unidade de distribuição • Quantidade de pesquisadores, não considerando dupla contagem, segundo a unidade de distribuição • Quantidade de pesquisadores doutores distintos, não considerando dupla contagem, segundo a unidade de distribuição • % da quantidade de pesquisadores doutores, não considerando dupla contagem, em relação ao total de pesquisadores doutores do censo
	<u>Unidades de Classificação</u> <ul style="list-style-type: none"> • Quantidade de pesquisadores doutores do último censo (decrecente)
	<u>Totalizações</u> <ul style="list-style-type: none"> • Totais das quantidades, considerando dupla contagem mesmo para unidades de conteúdo que são distribuídas sem dupla contagem • % do total da quantidade de pesquisadores doutores (deve ser 100%)
	<u>Apresentação do resultado</u> <ul style="list-style-type: none"> • Tabular

Tabela 14 - Indicadores do site “Séries Históricas” disponíveis para “Pesquisadores”

Unidade de Análise: Produção C&T

Unidade de Distribuição	Descrição do resultado
Grande área do conhecimento	<u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> • Quantidade de produções, com período de produção entre os últimos quatro anos anteriores ao do censo (incluindo o ano do censo) • Quantidade média de produções por ano por pesquisador doutor
	<u>Unidade de Filtro</u> <ul style="list-style-type: none"> • Tipo de produção = “Bibliográfica”
	<u>Unidade de Classificação</u> <ul style="list-style-type: none"> • Tipo de produção e Grande área do conhecimento (ordem alfabética)
	<u>Totalizações</u> <ul style="list-style-type: none"> • Total da quantidade de produção para cada censo • Média geral de produções por ano por pesquisador doutor para cada censo
	<u>Apresentação do resultado</u> <ul style="list-style-type: none"> • Tabular para cada subtipo de produção (“Artigos completos de circulação nacional”, “Artigos completos de circulação internacional”, “Trabalhos completos publicados em anais”, “Livros”, “Capítulos de livro”, “Produção Técnica”, “Teses”, “Dissertações”)

Tabela 15 - Indicadores do site “Séries Históricas” disponíveis para “Produção C&T”

RQF-02: Visualização de indicadores no site “Súmula Estatística”

No site “Súmula Estatística” serão publicados indicadores gerados estaticamente a partir das informações disponíveis no Data Mart do Diretório de Grupos de Pesquisa,

contendo informações de cada um dos censos anteriores, de forma que as análises sejam rapidamente apresentadas para os usuários. Nas definições de indicadores para o site “Súmula Estatística” deve-se considerar que o resultado é apresentado sempre para um censo específico, selecionado previamente pelo usuário. Os indicadores disponíveis para cada uma das unidades de análise são apresentados a seguir.

Unidade de Análise: Grupos de pesquisa

Unidades de Distribuição	Descrição do resultado
Região	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de grupos segundo a unidade de distribuição • % da quantidade de grupos em relação ao total <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de grupos (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total da quantidade de grupos • Total do % (deve ser 100%) <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular • Gráfico de setores (pizza) com a unidade de conteúdo “% da quantidade de grupos”
Unidade da Federação	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de grupos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total do censo • % acumulado da quantidade de grupos em relação ao total <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de grupos (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total da quantidade de grupos • Total do % (deve ser 100%) • Não há totalização para o % acumulado <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular

Instituição

Unidades de Conteúdo

- Quantidade de grupos segundo a unidade de distribuição
- % da quantidade de grupos segundo a unidade de distribuição em relação ao total do censo
- Quantidade total de pesquisadores distintos associados a grupos da instituição
- Quantidade total de pesquisadores distintos doutores associados a grupos da instituição

Unidade de Classificação

- Quantidade de grupos (decrecente)

Totalizações

- Total da quantidade de grupos
- Total do % (deve ser 100%)
- Total da quantidade de pesquisadores
- Total da quantidade de pesquisadores doutores

Apresentação do resultado

- Tabular
 - Gráfico de Linha: eixo “x” com número acumulado de instituições; eixo “y” com “% acumulado da quantidade de grupos em relação ao total de grupos do censo”
-

Grande área do conhecimento

Unidades de Conteúdo

- Quantidade de grupos segundo a unidade de distribuição
- % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo

Unidade de Agrupamento

- Dividir as grandes áreas em três grandes grupos de grandes áreas: a) Ciências da Natureza (“Engenharias” e “Ciência da Computação”); b) Ciências da Vida (“Ciências da Saúde”, “Ciências Biológicas” e “Ciências Agrárias”); c) Humanidades (“Ciências Humanas”, “Ciências Sociais Aplicadas” e “Linguística, Letras e Artes”)

Unidade de Classificação

- Quantidade de grupos em cada grande grupo de grande área (decrecente)

Totalizações

- Total da quantidade de grupos
- Total do % (deve ser 100%)

Apresentação do resultado

- Tabular
 - Gráfico de setores (pizza) com a unidade de conteúdo “% da quantidade total de grupos do censo” para cada um dos três grandes grupos de grandes áreas
-

Área do conhecimento	<hr/> <u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> • Quantidade de grupos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total do censo • % acumulado da quantidade de grupos em relação ao total <u>Totalizações</u> <ul style="list-style-type: none"> • Total da quantidade de grupos • Total do % (deve ser 100%) • Não há totalização para o % acumulado <u>Unidade de Classificação</u> <ul style="list-style-type: none"> • Quantidade de grupos (decrecente) <u>Apresentação do resultado</u> <ul style="list-style-type: none"> • Tabular <hr/>
Setor de aplicação	<hr/> <u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> • Quantidade de grupos distintos com linhas associadas à unidade de distribuição • Quantidade de linhas de pesquisa segundo a unidade de distribuição • % da quantidade de grupos em relação ao total de grupos do censo <u>Unidade de Classificação</u> <ul style="list-style-type: none"> • Quantidade de grupos (decrecente) <u>Totalizações</u> <ul style="list-style-type: none"> • Não há totalizações <u>Apresentação do resultado</u> <ul style="list-style-type: none"> • Tabular <hr/>
Ano de formação	<hr/> <u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> • Quantidade de grupos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos formados até o censo • % acumulado da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos formados até o censo <u>Unidade de Classificação</u> <ul style="list-style-type: none"> • Ano de formação (do mais recente para o menos recente) <u>Totalizações</u> <ul style="list-style-type: none"> • Total da quantidade de grupos • Total do % (deve ser 100%) <u>Apresentação do resultado</u> <ul style="list-style-type: none"> • Tabular • Gráfico de barras com a unidade de conteúdo “Quantidade de grupos” <hr/>

Tabela 16 - Indicadores do site “Súmula Estatística” disponíveis para “Grupos de pesquisa”

Unidade de Análise: Linhas de pesquisa

Unidades de Distribuição	Descrição do resultado
Quantidade de linhas de pesquisa por grupo de pesquisa (domínio de valores de "1" a "9" e "10 ou +" linhas por grupo)	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de grupos de pesquisa distintos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo • % acumulado da quantidade de grupos em relação ao total de grupos do censo <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de linhas de pesquisa por grupo (crescente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total da quantidade de grupos • Total do % (deve ser 100%) • Não há totalização para o % acumulado • Média do número de linhas por grupo • Mediana do número de linhas por grupo <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular
Grande área do conhecimento	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de linhas de pesquisa segundo a unidade de distribuição • Quantidade de pesquisadores, considerando dupla contagem, associados às linhas de pesquisa • Quantidade de grupos de pesquisa distintos • Relação L/G • Relação P/G • Relação P/L <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de linhas de pesquisa (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total da quantidade de linhas de pesquisa • Total da quantidade de pesquisadores • Total da quantidade de grupos • Relações L/G, P/G e P/L aplicadas aos totais do censo <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular
Área do conhecimento	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de linhas de pesquisa segundo a unidade de distribuição • Quantidade de pesquisadores, considerando dupla contagem, associados às linhas de pesquisa • Quantidade de grupos de pesquisa distintos <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de grupos de pesquisa (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total da quantidade de linhas de pesquisa • Total da quantidade de pesquisadores • Total da quantidade de grupos <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular

Tabela 17 - Indicadores do site “Súmula Estatística” disponíveis para “Linhas de pesquisa”

Unidade de Análise: Pesquisadores

Unidade de Distribuição	Descrição do resultado
Quantidade de pesquisadores por grupo de pesquisa (domínio de valores de “1” a “14” e “15 ou +” pesquisadores por grupo)	<u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> • Quantidade de grupos de pesquisa distintos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo • % acumulado da quantidade de grupos em relação ao total de grupos do censo
	<u>Unidade de Classificação</u> <ul style="list-style-type: none"> • Quantidade de pesquisadores por grupo (crescente)
	<u>Totalizações</u> <ul style="list-style-type: none"> • Total da quantidade de grupos • Total do % (deve ser 100%) • Não há totalização para o % acumulado • Média do número de pesquisadores por grupo • Mediana do número de pesquisadores por grupo
	<u>Apresentação do resultado</u> <ul style="list-style-type: none"> • Tabular • Gráfico de Linha: eixo “x” com número de pesquisadores por grupo; eixo “y” com “quantidade de grupos”
Titulação e Liderança (divisão tabular)	<u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> • Quantidade de pesquisadores segundo a titulação • % da quantidade de pesquisadores segundo a titulação em relação ao total de pesquisadores
	<u>Unidade de Classificação</u> <ul style="list-style-type: none"> • Quantidade de pesquisadores (decrescente)
	<u>Totalizações</u> <ul style="list-style-type: none"> • Total da quantidade de pesquisadores • Total do % (deve ser 100%)
	<u>Apresentação do resultado</u> <ul style="list-style-type: none"> • Tabular para os pesquisadores (independente da condição de liderança) • Tabular para os pesquisadores líderes • Tabular para os pesquisadores não líderes • Gráfico de barras: eixo “x” com o status de liderança; eixo “y” com o total de pesquisadores; barras sobrepostas divididas pela titulação (“Doutorado”, “Mestrado” e “Graduação ou Especialização”)

Unidade da Federação

Unidades de Conteúdo

- Quantidade de pesquisadores distintos segundo a unidade de distribuição
- Quantidade de pesquisadores distintos, com distribuição horizontal por titulação (“Doutorado”, “Mestrado”, “Especialização” e “Graduação”), segundo a unidade de distribuição
- % da quantidade de pesquisadores doutores segundo a unidade de distribuição em relação à quantidade de pesquisadores de acordo com a unidade de distribuição

Unidade de Classificação

- Quantidade de pesquisadores doutores (decrecente)

Totalizações

- Total da quantidade de pesquisadores
- % da quantidade total de pesquisadores doutores em relação à quantidade total de pesquisadores do censo

Apresentação do resultado

- Tabular

Titulação (distribuição horizontal) e Região (distribuição vertical)

Unidades de Conteúdo

- Quantidade de pesquisadores distintos segundo a região
- Quantidade de pesquisadores distintos, com distribuição horizontal por titulação (“Doutorado”, “Mestrado”, “Especialização” e “Graduação”)
- % da quantidade de pesquisadores por titulação em relação à quantidade de pesquisadores da região

Unidade de Classificação

- Quantidade de pesquisadores doutores (decrecente)

Totalizações

- Total da quantidade de pesquisadores
- % da quantidade total de pesquisadores em relação à quantidade total de pesquisadores do censo

Apresentação do resultado

- Tabular para a quantidade de pesquisadores
- Tabular para os % da quantidade de pesquisadores por titulação

- Grande área do conhecimento
- Instituição

Unidade de Conteúdo

- Quantidade de pesquisadores distintos segundo a unidade de distribuição
- Quantidade de pesquisadores doutores distintos segundo a unidade de distribuição
- % da quantidade de pesquisadores doutores segundo a unidade de distribuição em relação à quantidade de pesquisadores de acordo com a unidade de distribuição

Unidade de Classificação

- % da quantidade de pesquisadores doutores em relação à quantidade de pesquisadores (decrecente)

Totalizações

- Total da quantidade de pesquisadores
- Total da quantidade de pesquisadores doutores
- % da quantidade total de pesquisadores doutores em relação à quantidade total de pesquisadores do censo

Apresentação do resultado

- Tabular

Observação

- Esta definição é válida para ambas as unidades de distribuição

Faixa etária e Sexo
(distribuição horizontal)

Unidades de Conteúdo

- Quantidade de pesquisadores segundo a faixa etária
- Quantidade de pesquisadores do sexo masculino segundo a faixa etária
- Quantidade de pesquisadores do sexo feminino segundo a faixa etária
- % da quantidade de pesquisadores do sexo masculino em relação à quantidade de pesquisadores da faixa etária
- % da quantidade de pesquisadores do sexo feminino em relação à quantidade de pesquisadores da faixa etária

Unidade de Classificação

- Faixa etária (crescente)

Totalizações

- Totais das quantidades de pesquisadores e quantidades por sexo
- % da quantidade total de pesquisadores por sexo em relação à quantidade total de pesquisadores do censo

Apresentação do resultado

- Tabular para todos os pesquisadores
- Gráfico de barras horizontais: eixo “y” com faixas etárias; eixo “x” com percentual de pesquisadores; barras sobrepostas com os % das quantidades de pesquisadores dos sexos masculino e feminino em relação à quantidade de pesquisadores
- Tabular para os pesquisadores líderes
- Tabular para os pesquisadores não líderes

- Estrangeiros por Unidade da Federação
- Estrangeiros por Grande área do conhecimento
- Estrangeiros por Área do conhecimento

Unidades de Conteúdo

- Quantidade total de pesquisadores segundo a unidade de distribuição
- Quantidade total de pesquisadores estrangeiros segundo a unidade de distribuição
- % da quantidade de pesquisadores estrangeiros segundo a unidade de distribuição em relação à quantidade total de pesquisadores estrangeiros do censo
- % da quantidade de pesquisadores estrangeiros segundo a unidade de distribuição em relação à quantidade de pesquisadores de acordo com a unidade de distribuição

Unidade de Filtro

- Nacionalidade = “Estrangeira”

Unidade de Classificação

- % da quantidade de pesquisadores estrangeiros em relação à quantidade total de pesquisadores estrangeiros (decrecente)

Totalizações

- Totais das quantidades de pesquisadores e pesquisadores estrangeiros
- % da quantidade total de pesquisadores estrangeiros em relação à quantidade total de pesquisadores do censo

Apresentação do resultado

- Tabular

Observação

- Esta definição é válida para ambas as unidades de distribuição
-

Estrangeiros por País de origem	<u>Unidades de Conteúdo</u>
	<ul style="list-style-type: none"> • Quantidade de pesquisadores estrangeiros segundo a unidade de distribuição • % da quantidade de pesquisadores estrangeiros segundo a unidade de distribuição em relação à quantidade total de pesquisadores estrangeiros do censo
	<u>Unidade de Filtro</u>
	<ul style="list-style-type: none"> • Nacionalidade = “Estrangeira”
	<u>Unidade de Classificação</u>
	<ul style="list-style-type: none"> • % da quantidade de pesquisadores estrangeiros em relação à quantidade total de pesquisadores estrangeiros (decrecente)
	<u>Totalizações</u>
	<ul style="list-style-type: none"> • Total da quantidade de pesquisadores estrangeiros
	<u>Apresentação do resultado</u>
	<ul style="list-style-type: none"> • Tabular

Tabela 18 - Indicadores do site “Súmula Estatística” disponíveis para “Pesquisadores”

Unidade de Análise: Estudantes

Unidades de Distribuição	Descrição do resultado
Quantidade de estudantes por grupo de pesquisa (domínio de valores de “0” a “14” e “15 ou +” estudantes por grupo)	<u>Unidades de Conteúdo</u>
	<ul style="list-style-type: none"> • Quantidade de grupos de pesquisa distintos segundo a unidade de distribuição • % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo • % acumulado da quantidade de grupos em relação ao total de grupos do censo
	<u>Unidade de Classificação</u>
	<ul style="list-style-type: none"> • Quantidade de estudantes por grupo (crescente)
	<u>Totalizações</u>
	<ul style="list-style-type: none"> • Total da quantidade de grupos • Total do % (deve ser 100%) • Não há totalização para o % acumulado • Média do número de estudantes por grupo • Mediana do número de estudantes por grupo
	<u>Apresentação do resultado</u>
	<ul style="list-style-type: none"> • Tabular

<ul style="list-style-type: none"> • Região • Unidade da Federação 	<hr/> <p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de estudantes distintos segundo a unidade de distribuição • % da quantidade de estudantes segundo a unidade de distribuição em relação ao total de estudantes do censo • Quantidade de estudantes doutorandos distintos segundo a unidade de distribuição • % da quantidade de estudantes doutorandos segundo a unidade de distribuição em relação ao total de estudantes doutorandos do censo • % da quantidade de estudantes doutorandos segundo a unidade de distribuição em relação à quantidade de estudantes de acordo com a unidade de distribuição <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de estudantes doutorandos (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total da quantidade de estudantes • Total do % de estudantes (deve ser 100%) • Total da quantidade de estudantes doutorandos • Total do % de estudantes doutorandos (deve ser 100%) • % da quantidade total de estudantes doutorandos em relação à quantidade total de estudantes do censo <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular <p><u>Observação</u></p> <ul style="list-style-type: none"> • Esta definição é válida para ambas as unidades de distribuição <hr/>
<p>Grande área do conhecimento e Nível de formação (distribuição horizontal)</p>	<hr/> <p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de estudantes distintos segundo a unidade de distribuição • Quantidade de estudantes distintos segundo o nível de formação (“Doutorando”, “Mestrando”, e “Graduando”) • % das quantidades de pesquisadores por nível de formação em relação à quantidade de pesquisadores <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Quantidade de estudantes (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Totais das quantidades de estudantes • % das quantidades totais de pesquisadores em relação à quantidade total de pesquisadores do censo <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular <hr/>

Tabela 19 - Indicadores do site “Súmula Estatística” disponíveis para “Estudantes”

Unidade de Análise: Técnicos

Unidades de Distribuição	Descrição do resultado
Quantidade de técnicos por grupo de pesquisa (domínio de valores de “0” a “10” e “11 ou +” técnicos por grupo)	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> Quantidade de grupos de pesquisa distintos segundo a unidade de distribuição % da quantidade de grupos segundo a unidade de distribuição em relação ao total de grupos do censo <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> Quantidade de técnicos por grupo (crescente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> Total da quantidade de grupos Total do % (deve ser 100%) <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> Tabular
Grande área do conhecimento predominante no grupo	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> Quantidade de técnicos segundo a unidade de distribuição Quantidade de grupos distintos com técnicos segundo a unidade de distribuição Relação entre a quantidade de técnicos segundo a unidade de distribuição e a quantidade de grupos com técnicos de acordo com a unidade de distribuição Quantidade total de grupos distintos segundo a unidade de distribuição com ou sem técnicos Relação entre a quantidade de técnicos segundo a unidade de distribuição e a quantidade total de grupos de acordo com a unidade de distribuição <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> Quantidade de técnicos (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> Total da quantidade de técnicos Total da quantidade de grupos com técnicos Relação entre o total de técnicos e o total de grupos com técnicos Total da quantidade total de grupos Relação entre o total de técnicos e o total de grupos <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> Tabular
<ul style="list-style-type: none"> Área do conhecimento e Grau de escolaridade (distribuição horizontal) Tipo de atividade e Grau de escolaridade (distribuição horizontal) 	<p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> Quantidade de técnicos segundo a unidade de distribuição Quantidade de técnicos por grau de escolaridade (“Mestrado e Doutorado”, “Graduação e Especialização” e “Primeiro e Segundo Graus”) <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> Unidade de distribuição (ordem alfabética) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> Totais de todas as colunas de quantidades <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> Tabular <p><u>Observação</u></p> <ul style="list-style-type: none"> Esta definição é válida para ambas as unidades de distribuição

Tabela 20 - Indicadores do site “Súmula Estatística” disponíveis para “Técnicos”

Unidade de Análise: Produção C&T

Unidade de Distribuição	Descrição do resultado
Grande área do conhecimento	<u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> Quantidade de produções segundo a unidade de distribuição Quantidade média de produções por ano segundo a unidade de distribuição Número médio de produções por pesquisador doutor por ano segundo a unidade de distribuição
	<u>Unidade de Filtro</u> <ul style="list-style-type: none"> Tipo de produção = "Bibliográfica"
	<u>Unidade de Classificação</u> <ul style="list-style-type: none"> Tipo de produção e Grande área do conhecimento (ordem alfabética)
	<u>Totalizações</u> <ul style="list-style-type: none"> Total da quantidade de produção Total da quantidade média de produções por ano Número médio de produções por pesquisador doutor do censo
	<u>Apresentação do resultado</u> <ul style="list-style-type: none"> Tabular para cada um dos subtipos de produção Gráfico de barras horizontais: eixo "y" com as grandes áreas do conhecimento; eixo "x" com o número médio de produções por pesquisador doutor por ano; uma barra para cada subtipo de produção

Tabela 21 - Indicadores do site "Súmula Estatística" disponíveis para "Produção C&T"

Unidade de Análise: Empresas

Unidades de Distribuição	Descrição do resultado
<ul style="list-style-type: none"> Grande área Área Região UF Instituição 	<u>Unidades de Conteúdo</u> <ul style="list-style-type: none"> Quantidade de grupos distintos que mencionaram relacionamento com empresas segundo a unidade de distribuição Quantidade total de grupos distintos segundo a unidade de distribuição % entre a quantidade total de grupos segundo a unidade de distribuição e a quantidade de grupos distintos que mencionaram relacionamento com empresas em relação à unidade de distribuição
	<u>Unidade de Classificação</u> <ul style="list-style-type: none"> A própria unidade de distribuição
	<u>Totalizações</u> <ul style="list-style-type: none"> Total de grupos distintos que mencionaram relacionamento com empresas Total de grupos distintos do censo % entre a quantidade total de grupos do censo e a quantidade de grupos distintos que mencionaram relacionamento com empresas
	<u>Apresentação do resultado</u> <ul style="list-style-type: none"> Tabular
	<u>Observação</u> <ul style="list-style-type: none"> Esta definição é válida para ambas as unidades de distribuição

<ul style="list-style-type: none"> • Tipo de relacionamento • Forma de remuneração 	<hr/> <p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Quantidade de grupos distintos que mencionaram relacionamento com empresas segundo a unidade de distribuição • Quantidade de empresas distintas mencionadas pelos grupos segundo a unidade de distribuição <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • A própria unidade de distribuição <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total de grupos • Total de empresas <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular <p><u>Observação</u></p> <ul style="list-style-type: none"> • Esta definição é válida para ambas as unidades de distribuição <hr/>
Empresas mais mencionadas	<hr/> <p><u>Unidades de Conteúdo</u></p> <ul style="list-style-type: none"> • Número de menções que cada empresa teve • % em relação ao total de empresas mencionadas <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Número de menções (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total de menções • % do total de menções em relação ao total de empresas mencionadas <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular <p><u>Observação</u></p> <ul style="list-style-type: none"> • Apresentar as 30 empresas com maior número de menções e o total de vezes e % relativo às demais empresas mencionadas <hr/>
Setor da Atividade Econômica	<hr/> <p><u>Unidade de Conteúdo</u></p> <ul style="list-style-type: none"> • Número de empresa segundo a unidade de distribuição • % em relação ao total de empresas mencionadas <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Número de empresas (decrecente) <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total de empresas • % total (deve ser 100%) <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular <hr/>
<ul style="list-style-type: none"> • Natureza Jurídica • Faixa de número de empregados 	<hr/> <p><u>Unidade de Conteúdo</u></p> <ul style="list-style-type: none"> • Número de empresas segundo a unidade de distribuição <p><u>Unidade de Classificação</u></p> <ul style="list-style-type: none"> • Unidade de distribuição <p><u>Totalizações</u></p> <ul style="list-style-type: none"> • Total de empresas <p><u>Apresentação do resultado</u></p> <ul style="list-style-type: none"> • Tabular <p><u>Observação</u></p> <ul style="list-style-type: none"> • Esta definição é válida para ambas as unidades de distribuição <hr/>

Tabela 22 - Indicadores do site “Súmula Estatística” disponíveis para “Empresas”

RQF-03: Visualização de indicadores no site “Plano Tabular”

No site “Plano Tabular” estarão disponíveis indicadores gerados dinamicamente a partir das informações armazenadas no Data Mart do Diretório de Grupos de Pesquisa. Tais indicadores contêm informações de cada um dos censos anteriores e do censo atual. As funcionalidades que devem ser disponibilizadas no site “Plano Tabular” são as que se seguem.

- Seleção e combinação das unidades: a combinação das unidades de análise, de conteúdo, de distribuição, de filtro e de classificação será permitida de forma dinâmica, respeitando as regras de relacionamento entre essas unidades definidas em termos de requisitos deste documento.
- Resultado dinâmico: a geração do resultado para as unidades selecionadas será feita de forma dinâmica acessando a base de dados do Data Mart do Diretório de Grupos de Pesquisa, com a finalidade de obter os valores mais atualizados de indicadores disponíveis nessa base. Estes serão apresentados na forma tabular e acompanhados de uma descrição do resultado bem como de notas de rodapé selecionadas automaticamente pelo site, de acordo com a combinação das unidades utilizadas.
- Impressão e download: estarão disponíveis opções para que o usuário possa gerar uma versão do resultado de indicadores para impressão ou para download.
- Operações OLAP: podem ser aplicadas ao resultado obtido para gerar uma nova visão de análise ou para detalhá-lo; a funcionalidade de cada operação OLAP disponível na visualização de indicadores está descrita na especificação complementar.

- Integração com a Busca Textual: sempre que visualizados, os indicadores terão hiperlinks ativos para a integração do site “Plano Tabular” com o site da “Busca Textual”. No momento em que é selecionado qualquer um dos indicadores devem-se enviar para o site da “Busca Textual” a unidade de análise e todas as unidades de filtro que foram utilizadas na sua composição, para a apresentação de uma lista de itens (grupos, linhas, pesquisadores, estudantes, técnicos ou produções) correspondentes.

RQF-04: Visualização de indicadores no site “Estratificação de Grupos”

No site “Estratificação de Grupos” serão publicados indicadores gerados estaticamente a partir das informações disponíveis no Data Mart do Diretório de Grupos de Pesquisa, contendo dados relativos ao resultado da estratificação dos grupos de pesquisa de um determinado censo. A aplicação de uma rotina de estratificação classifica os grupos de pesquisa em três categorias (estratos) possíveis:

- **eF**: Grupos em formação
- **eC**: Grupos em consolidação
- **Co**: Grupos consolidados

Apesar de a implementação da metodologia de estratificação de grupos de pesquisa (<http://lattes.cnpq.br/censo2002/estratificacao/metodologia.html>) não ser uma das responsabilidades diretas do Data Mart de Grupos de Pesquisa, este repositório, entretanto, deve prever o armazenamento da classificação final dos grupos, de forma a permitir a extração de indicadores relacionados a essa característica.

Para cada quantidade referente a cada uma das unidades de conteúdo pode ser considerada ou não a dupla contagem dependendo do tipo de indicador que será

apresentado. Esses indicadores disponíveis na estratificação de grupos combinam uma ou mais das seguintes unidades de distribuição para a visualização dos valores referentes às unidades de conteúdo:

- Grande área predominante nas atividades do grupo;
- Área predominante nas atividades do grupo;
- Estrato do grupo de pesquisa;
- Unidade da Federação;
- Instituição de Ensino Superior.

RQF-05: Visualização de indicadores (todas as unidades de análise)

Sempre que uma tabela de indicadores ou gráfico forem apresentados, estes deverão vir acompanhados de uma descrição detalhada e, quando necessário, de notas de rodapé. Tal descrição deve apresentar o conteúdo e os critérios utilizados para seleção dos dados empregados na contabilização dos indicadores, e as notas de rodapé devem explicar condições específicas que caracterizam os indicadores apresentados.

RQF-06: Visualização de indicadores da Unidade de Análise “Grupos de pesquisa”

Quando houver um total de grupos de pesquisa menor que 30 e maior que zero, serão colocados 'x' vermelhos nas células correspondentes na tabela e será mostrada a nota de rodapé correspondente.

RQF-07: Visualização do resultado da “Busca Textual”

No site “Busca Textual”, para os critérios específicos de seleção e palavras-chave informados pelo usuário, será apresentado o resultado da busca contendo as informações para cada uma das unidades de análise disponíveis, conforme demonstrado no quadro a seguir.

Unidades de Análise	Resultado da Busca Textual
Grupos de pesquisa	<ul style="list-style-type: none"> • Nome do grupo • Sigla da instituição do grupo • Área predominante • Nome do(s) líder(es)
Linhas de pesquisa	<ul style="list-style-type: none"> • Nome da linha de pesquisa • Nome do grupo • Sigla da instituição do grupo • Área de conhecimento oficial (1º. Campo relacionado) • Setor de Aplicação (1º. Campo relacionado)
Pesquisadores	<ul style="list-style-type: none"> • Nome do pesquisador • Sigla da instituição do grupo • Titulação máxima • Nome do grupo
Estudantes	<ul style="list-style-type: none"> • Nome do estudante • Sigla da instituição do grupo • Nível de treinamento • Sigla da instituição do curso • Nome do grupo
Produção C&T	<ul style="list-style-type: none"> • Título da produção • Tipo e subtipo da produção • Autor(es) (na forma de referência bibliográfica)

Tabela 23 - Conteúdo apresentado como resultado do site “Busca Textual”

Requisitos Não Funcionais

Os requisitos abaixo descrevem restrições e condições não diretamente visíveis ao usuário final das aplicações que utilizarão o Data Mart do Diretório de Grupos de Pesquisa como fonte de dados para a visualização de indicadores.

RQnF-01: Performance - Tempo de resposta e consumo de memória

O tempo máximo para a execução de consultas realizadas pelas aplicações referentes aos sites “Plano Tabular” e “Busca Textual”, combinado com a estimativa do número de acessos simultâneos esperado para essas aplicações, deve estar de acordo com as configurações estabelecidas para *timeout* de página e capacidade de memória. Também o processamento dos servidores Web e de banco de dados a

serem utilizados para disponibilizar essas aplicações deve ocorrer adequadamente, visto que elas geram indicadores ou realizam busca de forma dinâmica, com acesso à base de dados do Data Mart do Diretório de Grupos de Pesquisa.

RQnF-02: Segurança - Integridade

Devem ficar claras para os usuários das aplicações que acessam informações do Data Mart do Diretório de Grupos de Pesquisa as condições em que os indicadores são gerados com dupla contagem ou não. A implementação deste requisito está prevista por meio da inclusão de “Notas de rodapé” a serem apresentadas na visualização dos indicadores.

RQnF-03: Segurança - Disponibilidade

Os sites que disponibilizam informações do Data Mart do Diretório de Grupos de Pesquisa possuem, de uma forma geral, a característica de disponibilidade integral (24 x 7). Deve-se considerar, entretanto, que o processo de carga e atualização do “Censo Atual” será executado periodicamente e poderá afetar consideravelmente a performance dessas aplicações. O processo de carga deve automaticamente desabilitar as aplicações enquanto o acesso a elas estiver sendo executado. Ao término do processo, o acesso é liberado. Caso o acesso seja bloqueado, os usuários das aplicações poderão, através de uma página HTML, obter informações detalhadas sobre o processo de carga em si e uma estimativa de tempo para a sua conclusão.

RQnF-04: Segurança - Confidencialidade

Por não possuírem qualquer característica de sigilo, as informações disponibilizadas pelo Data Mart do Diretório de Grupos de Pesquisa estarão disponíveis sem

qualquer tipo de restrição ao público em geral, através das aplicações (sites) que disponibilizarão essas informações de forma dinâmica ou estática.

RQnF-05: Multidimensionalidade - Conformidade

O Data Mart do Diretório de Grupos de Pesquisa deve ser projetado para disponibilizar tanto indicadores com dupla contagem, como já utilizados nas atuais versões das aplicações para o “Censo 2002”, quanto indicadores sem dupla contagem. Estes, apesar de exigirem maior recurso de processamento para serem calculados, serão disponibilizados através das aplicações que oferecem informações estáticas aos usuários.

RQnF-06: Multidimensionalidade - Integrabilidade

Os dados operacionais do Diretório de Grupos de Pesquisa serão utilizados pelo processo de carga para atualizar o “Censo Atual”. Sempre que for necessária a geração de uma nova versão de censo histórico (em 2004, por exemplo), o “Censo Atual” será convertido para o censo histórico e a próxima execução da carga e atualização do Data Mart voltará novamente a gerar uma versão do “Censo Atual”. A periodicidade para a geração de censos históricos está inicialmente definida para ser executada a cada dois anos, e a atualização do “Censo Atual” não possui ainda uma periodicidade a ser considerada para a execução da carga de dados.

RQnF-07: Multidimensionalidade - Interpretabilidade

O usuário deve ser informado sobre:

- as possíveis diferenças entre informações que constam na base de dados operacional do Diretório de Grupos de Pesquisa e aquelas disponibilizadas como “Censo Atual” na base de dados do Data Mart, em virtude da configuração adotada para a periodicidade de atualização desse censo;

- as possíveis diferenças entre indicadores referentes à produção de pesquisadores e estudantes nos censos históricos, pelo fato de os detalhes da produção de cada indivíduo não serem armazenados na forma histórica tal como será feito com os indicadores da produção – pesquisadores e estudantes podem atualizar seus currículos Lattes com informações de produção de anos anteriores, incluindo ou excluindo essas informações de seus currículos, e a consulta da produção de censos históricos mostrará indicadores com valores divergentes daqueles obtidos pela relação atualizada das produções desses indivíduos considerando os mesmos parâmetros adotados no cálculo dos indicadores.

RQnF-08: Usabilidade - Operabilidade e facilidade de uso

Por serem disponibilizadas através de aplicações (sites) Web, as consultas de indicadores e a busca textual já possuem naturalmente uma interface intuitiva no que se refere à sua operação pelos usuários. Essas aplicações devem, adicionalmente, apresentar mensagens e notas explicativas resumidas em todas as páginas onde houver necessidade de que o usuário execute operações que estejam fora do escopo tradicional de navegação em um site. Devem ser também exibidas aos usuários notas técnicas sempre que eles solicitarem, na forma de uma “ajuda” on-line. Tais notas devem descrever características das informações disponibilizadas e detalhar funcionalidades das aplicações utilizadas para acessar a base de dados do Data Mart do Diretório de Grupos de Pesquisa.

RQnF-09: Usabilidade - Flexibilidade

Os usuários das aplicações poderão utilizar apenas critérios simples de filtros para definir um subconjunto de dados que serão considerados na construção dos

indicadores ou na realização da busca textual. Critérios simples implicam a escolha de apenas um único argumento de filtro para cada unidade de distribuição utilizada com essa finalidade. Condições de filtros dos tipos “maior que”, “menor que”, “diferente de”, “entre X e Y” não estão previstas para a implementação das aplicações de visualização de indicadores.

4.2.4 Especificação complementar

EC-01: Descrição da consulta de indicadores: regras de formação

Toda a consulta de indicadores deverá vir acompanhada de uma descrição detalhada de seu conteúdo (unidades de análise, conteúdo e distribuição) e dos critérios de formação dos indicadores (unidades de filtro). Essa descrição será formada pelos seguintes itens:

- descrição da(s) unidade(s) de conteúdo da unidade de análise que está sendo consultada;
- descrição da unidade de distribuição horizontal (formadora de colunas adicionais na tabela de resultados);
- descrição das unidades de filtro e de seus respectivos valores;
- descrição da(s) unidade(s) de distribuição vertical (formadora das diferentes linhas apresentadas na tabela de resultados).

Exemplo: “Distribuição das linhas de pesquisa, de pesquisadores e de grupos, segundo a área de conhecimento predominante nas atividades do grupo, para grande área de conhecimento predominante nas atividades do grupo igual a ‘Ciências Exatas e da Terra’ do Censo ‘2002’.”

Tipo de Unidade	Seleção
Unidade de Análise	<ul style="list-style-type: none"> Linhas de pesquisa
Unidades de Conteúdo	<ul style="list-style-type: none"> Quantidade de linhas de pesquisa Quantidade de pesquisadores Quantidade de grupos
Unidades de Filtro	<ul style="list-style-type: none"> Censo = "2002" Grande área = "Ciências Exatas e da Terra"
Unidade de Distribuição (vertical)	<ul style="list-style-type: none"> Área do conhecimento predominante do grupo

Tabela 24 - Unidades utilizadas para a geração da descrição de uma consulta do "Plano Tabular"

EC-02: Descrição de notas de rodapé: regras de formação

Notas de rodapé devem ser mostradas na consulta de indicadores de acordo com as unidades de análise, de conteúdo e de distribuição selecionadas para a consulta, conforme a tabela abaixo.

Unidades de Análise	Nota de rodapé	Condição de visualização
Grupos de pesquisa	"Pode haver dupla contagem no número de pesquisadores, estudantes e técnicos."	Quando forem visualizadas uma ou mais das unidades de conteúdo: <ul style="list-style-type: none"> "Pesquisadores" "Estudantes" "Técnicos"
	"Esta informação só pode ser fornecida quando a célula corresponder a mais de 30 grupos de pesquisa."	Quando houver um total de grupos de pesquisa menor que 30 e maior que zero. Ver requisito funcional "Visualização de indicadores da unidade de análise 'Grupos de Pesquisa'".
Linhas de pesquisa	"Pode haver dupla contagem de pesquisadores."	Quando a unidade de conteúdo "Pesquisadores" for visualizada.
	"Não existe dupla contagem no número de Pesquisadores na menor unidade de informação."	Em qualquer consulta, independente de unidade de conteúdo ou distribuição.
	"Cada Linha de pesquisa pode estar associada a até 3 Grandes áreas ou Áreas do conhecimento ou a até 3 Setores de aplicação."	Quando a unidade de distribuição selecionada for "Grande área", "Área" ou "Setor de aplicação".
Pesquisadores	"Não existe dupla contagem no número de Pesquisadores na menor unidade de informação."	Em qualquer consulta, independente de unidade de distribuição.
Estudantes	"Não existe dupla contagem no número de Estudantes na menor unidade de informação."	Em qualquer consulta, independente de unidade de distribuição.
Pessoal técnico	"Pode haver dupla contagem no número de Técnicos na menor unidade de informação."	Em qualquer consulta, independente de unidade de distribuição.
Produção C&T	"Pode haver dupla contagem no total de itens da produção C&T, tendo em vista poder existir mais de um autor por produção e um autor poder participar de mais de um grupo de pesquisa."	Em qualquer consulta, independente de unidade de distribuição.

Tabela 25 - Regras de formação de notas de rodapé

EC-03: Fontes de dados das Unidades de Análise, Conteúdo e Distribuição

As unidades de análise, conteúdo e distribuição serão geradas a partir de informações armazenadas no Sistema Diretório de Grupos de Pesquisa utilizando-se as informações já mantidas nas dimensões padronizadas do Data Warehouse ou através de informações originárias de fontes externas (CAPES).

Unidades de Análise	Fonte de Dados
<ul style="list-style-type: none"> • Grupos de pesquisa • Linhas de pesquisa • Pesquisadores • Estudantes • Pessoal Técnico • Empresas 	Informações do Sistema “Diretório de Grupos de Pesquisa” e de fontes externas para as unidades de conteúdo que identificam propriedades de pesquisadores associadas à CAPES.
<ul style="list-style-type: none"> • Produção C&T 	Informações da unidade de análise “Produção” do Data Mart de Currículos Lattes

Tabela 26 - Fontes de dados das Unidades de Análise e Conteúdo

Dimensões	Fonte de Dados
<ul style="list-style-type: none"> • Censo 	Gerada automaticamente quando da realização de um censo, a partir das informações armazenadas como censo “Atual”
<ul style="list-style-type: none"> • Empresa 	Relação de empresas associadas aos grupos de pesquisa no Sistema “Diretório de Grupos de Pesquisa” (tabela EM_EMPRESA_GR)
<ul style="list-style-type: none"> • Distribuição geográfica 	Dimensões “Geografia” (tabela DI_GEOGRAFIA) e “Instituição” (tabela DI_INSTITUICAO) padronizadas no DW
<ul style="list-style-type: none"> • Árvore do conhecimento 	Dimensão “Área do conhecimento” (tabela DI_AREA_CONHEC) padronizada no DW e dimensão “Setor de aplicação” a ser padronizada com informações da base de dados operacional
<ul style="list-style-type: none"> • Características pessoais 	Gerada automaticamente a partir da combinação de valores das unidades de distribuição que constituem essa dimensão
<ul style="list-style-type: none"> • Atividade técnica 	Relação de atividades técnicas do Sistema “Diretório de Grupos de Pesquisa” (tabela EM_ATIVIDADE_TECNICA)
<ul style="list-style-type: none"> • Tipo de produção 	Dimensão “Tipo de Produção” (tabela DI_TIPO_PRODUCAO) padronizada no DW
<ul style="list-style-type: none"> • Tipo de relacionamento 	Relação de tipos de relacionamento do Sistema “Diretório de Grupos de Pesquisa” (tabela EN_RELACAO_PREDOMINANTE)
<ul style="list-style-type: none"> • Forma de remuneração 	Relação de formas de remuneração do Sistema “Diretório de Grupos de Pesquisa” (tabela EN_TIPO_REMUNERACAO)
<ul style="list-style-type: none"> • Setor da atividade econômica 	Relação de setores econômicos do Sistema “Diretório de Grupos de Pesquisa” (tabela EN_SETOR_ATIVIDADE_ECONOMICA)
<ul style="list-style-type: none"> • Pesquisadores • Estudantes 	Dimensão “Pessoa” (tabela DI_PESSOA) padronizada no DW

Tabela 27 - Fontes de dados das Unidades de Distribuição (Dimensões)

EC-04: Tipos de Relacionamento entre Empresas e Grupos

São 14 os tipos de relacionamento previstos até o momento entre empresas e grupos de pesquisa, descritos no quadro a seguir.

#	Tipos de Relacionamento
1.	Atividades de consultoria técnica não englobadas em qualquer das categorias anteriores
2.	Atividades de engenharia não rotineiras, inclusive o desenvolvimento de protótipo cabeça de série ou planta piloto para o parceiro
3.	Atividades de engenharia não rotineiras, inclusive o desenvolvimento ou a fabricação de equipamentos para o grupo
4.	Desenvolvimento de software não-rotineiro para o grupo pelo parceiro
5.	Desenvolvimento de software para o parceiro pelo grupo
6.	Fornecimento, pelo grupo, de insumos materiais para as atividades do parceiro, sem vinculação a um projeto específico de interesse mútuo
7.	Fornecimento, pelo parceiro, de insumos materiais para as atividades de pesquisa do grupo, sem vinculação a um projeto específico de interesse mútuo
8.	Pesquisa científica com considerações de uso imediato dos resultados
9.	Pesquisa científica sem considerações de uso imediato dos resultados
10.	Transferência de tecnologia desenvolvida pelo grupo para o parceiro
11.	Transferência de tecnologia desenvolvida pelo parceiro para o grupo
12.	Treinamento de pessoal do grupo pelo parceiro, incluindo cursos e treinamento "em serviço"
13.	Treinamento de pessoal do parceiro pelo grupo, incluindo cursos e treinamento "em serviço"
14.	Outros tipos predominantes de relacionamento que não se enquadrem em nenhum dos anteriores

Tabela 28 - Tipos de relacionamento possíveis entre empresas e grupos de pesquisa

EC-05: Formas de Remuneração entre Empresas e Grupos

São 10 as formas de remuneração previstas até o momento entre empresas e grupos de pesquisa, descritas no quadro a seguir.

#	Formas de Remuneração
1.	Fornecimento de bolsas para o grupo pelo parceiro
2.	Parceria com transferência de recursos de qualquer espécie nos dois sentidos
3.	Parceria sem a transferência de recursos de qualquer espécie, envolvendo exclusivamente relacionamento de risco
4.	Transferência de insumos materiais para as atividades de pesquisa do grupo
5.	Transferência de insumos materiais para as atividades do parceiro
6.	Transferência de recursos financeiros do grupo para o parceiro
7.	Transferência de recursos financeiros do parceiro para o grupo
8.	Transferência física temporária de recursos humanos do grupo para as atividades do parceiro
9.	Transferência física temporária de recursos humanos do parceiro para as atividades de pesquisa do grupo
10.	Outras formas de remuneração que não se enquadrem em nenhuma das anteriores

Tabela 29 - Formas de remuneração possíveis entre empresas e grupos de pesquisa

EC-06: Operações OLAP disponíveis no site “Plano Tabular”

As operações apresentadas no quadro a seguir estarão disponíveis para um usuário após ele visualizar os indicadores de uma consulta no site “Plano Tabular”.

Operação	Descrição
<i>Drill down</i>	<ul style="list-style-type: none"> • O usuário poderá utilizar o mouse para selecionar uma das linhas da consulta bem como uma nova unidade de agrupamento horizontal • Todas as unidades de agrupamento visíveis da linha selecionada devem ser transformadas em unidades de filtro, de acordo os valores destas unidades para essa linha • Um novo resultado será apresentado com o resultado dos indicadores, já considerando as novas definições (unidades de filtro e de agrupamento)
<i>Sort</i>	<ul style="list-style-type: none"> • O usuário poderá utilizar o mouse para selecionar uma das colunas da consulta para ser definida como uma unidade de classificação • Qualquer uma das colunas referentes a unidades de conteúdo pode ser selecionada, entretanto somente a coluna de agrupamento de nível mais interno (menor granularidade) poderá ser selecionada para unidade de classificação • Quando a coluna que já está definida como unidade de classificação for novamente selecionada, o resultado será uma inversão na ordem de classificação entre crescente e decrescente (somente para colunas referentes a unidades de conteúdo) • Uma coluna referente a uma unidade de agrupamento será sempre classificada de forma crescente ou alfabética • Uma coluna referente a uma unidade de conteúdo será sempre classificada de forma crescente ao ser selecionada pela primeira vez

Tabela 30 - Operações OLAP disponíveis no “Plano Tabular”

4.2.5 Glossário

Glossário conceitual

Grupo de pesquisa

É um conjunto de indivíduos organizados hierarquicamente a) cujo fundamento organizador da hierarquia são a experiência, o destaque e a liderança no terreno científico ou tecnológico; b) que estão envolvidos de forma profissional e permanentemente com atividades de pesquisa; c) cujo trabalho se organiza em torno de linhas comuns de pesquisa; e d) e que, em algum grau, compartilham instalações e equipamentos.

Líder de grupo

O pesquisador líder de grupo é o personagem que detém a liderança acadêmica e intelectual naquele ambiente de pesquisa. Normalmente, tem a responsabilidade de coordenação e planejamento dos trabalhos de pesquisa do grupo. Sua função aglutina os esforços dos demais pesquisadores e aponta horizontes e novas áreas de atuação dos trabalhos.

Linha de pesquisa

Linhas de pesquisa representam temas aglutinadores de estudos científicos que se fundamentam em tradição investigativa, estudos dos quais se originam projetos cujos resultados guardam afinidades entre si.

Pesquisador

Pesquisadores são os membros graduados ou pós-graduados da equipe de pesquisa direta e criativamente envolvidos com a realização de projetos e com a produção científica, tecnológica e artística do grupo. Estagiários pós-doutores são considerados pesquisadores.

Estudante

São estudantes (bolsistas ou não) em iniciação científica e pós-graduação (exceto os estagiários em nível de pós-doutoramento, que devem ser considerados pesquisadores) aqueles que participam ativamente das linhas de pesquisa desenvolvidas pelo grupo como parte de suas atividades discentes, sob a orientação de pesquisadores do grupo.

Censo

Um censo corresponde a uma visão estática das informações do Diretório de Grupos de Pesquisa, fixadas em um determinado momento. Usualmente essa visão é gerada para cada período de dois anos. O censo atual corresponde à visão dinâmica da base de informações do Diretório de Grupos de Pesquisa.

Glossário técnico

Unidade de Análise

Conjunto de indicadores referentes a um mesmo processo ou entidade de negócio, dentro do escopo do projeto que está sendo tratado. Uma unidade de análise representa uma ou mais características mensuráveis, medidas ou métricas a respeito do processo ou da entidade a ser analisada. As unidades de análise podem ser agrupadas em assuntos. Neste projeto, por exemplo, o “Diretório de Grupos de Pesquisa” corresponde a um assunto; “Grupos de Pesquisa” e “Linhas de Pesquisa” são unidades de análise dentro desse assunto.

Unidade de Conteúdo

Corresponde a cada um dos indicadores que podem ser analisados dentro de uma unidade de análise. Unidades de conteúdo possuem valores mensuráveis e que permitem operações algébricas e estatísticas, tais como totalização, cálculo de médias ou distribuição percentual. Neste projeto, por exemplo, a unidade de análise “Grupos de Pesquisa” pode ser analisada sob diferentes unidades de conteúdo, entre elas, a quantidade de grupos de pesquisa e a quantidade média de linhas de pesquisa por grupo.

Unidade de Distribuição

São utilizadas para classificar o valor dos indicadores representados pelas unidades de conteúdo, sendo retratadas por um conjunto de hierarquias e dimensões, que tipicamente qualificam os indicadores em distribuições temporais, geográficas, institucionais ou outras específicas para cada unidade de análise ou de conteúdo. Cada uma destas pode ter seu subconjunto de unidades de distribuição válidas de serem aplicadas.

Unidade de Filtro

Pode ser: a) uma unidade de distribuição utilizada para selecionar um subconjunto de valores de uma ou mais unidades de conteúdo de uma unidade de análise – por exemplo, em “vendas realizadas pelas filiais da Região Sul” a unidade de distribuição “região” está sendo utilizada como unidade de filtro para selecionar informações referentes apenas a uma região específica; b) ou uma faixa de valores de uma unidade de conteúdo utilizada para restringir o domínio de valores dos indicadores que devem ser analisados – por exemplo, em “filiais com vendas realizadas superiores a R\$ 5.000,00 no mês atual” a unidade de conteúdo “valor das vendas” está sendo utilizada como unidade de corte para filtrar o conjunto de filiais que satisfaçam o valor mínimo de venda a ser analisado.

Unidade de Agrupamento

É uma unidade de distribuição utilizada para subdividir os valores dos indicadores representados pelas unidades de conteúdo de uma unidade de análise em linhas (agrupamento horizontal) ou colunas (agrupamento vertical); uma consulta com unidade de agrupamento é usualmente descrita como “total de <unidade de conteúdo> por <unidade de distribuição>”; quanto maior a quantidade de unidades

de agrupamento horizontal ou vertical utilizadas, maior será o número de linhas ou colunas resultantes em uma consulta.

Unidade de Classificação

É uma unidade de distribuição ou de conteúdo utilizada para classificar o resultado de uma consulta de forma ascendente ou descendente, alfabeticamente para valores de unidades de distribuição com conteúdo textual ou numericamente para unidades de conteúdo.

Operações OLAP

São operações de agregação aplicadas sobre um conjunto de dados, com o objetivo de gerar resultados (somatórios e contagens sobre as unidades de análise) mais genéricos ou mais detalhados através da combinação de unidades de distribuição disponíveis (filtro, agrupamento ou classificação).

4.3 Conclusões

Neste capítulo foi apresentado um estudo de caso no qual foi aplicada a proposta de documentação de requisitos de projetos de DW definida no capítulo anterior. A utilização da proposta neste estudo de caso permitiu a sua validação e o seu refinamento com a finalidade de elaborar a estratégia mais adequada para a documentação dos requisitos de um DW.

5 CONCLUSÕES E TRABALHOS FUTUROS

5.1 Contexto

A análise de requisitos está inserida no contexto de um processo de desenvolvimento de software como uma disciplina intermediária entre as disciplinas de modelagem de negócios e de análise e projeto do sistema. Conceitualmente, é importante destacar as diferentes atividades que são de responsabilidade de cada uma dessas disciplinas, mesmo que tais atividades sejam sobrepostas em um cronograma de atividades de desenvolvimento.

5.1.1 Análise de requisitos x modelagem de processos

De uma forma geral, a modelagem de processos pode ser realizada não somente com o objetivo de conhecer um processo de negócio para que se possa automatizá-lo, mas também com objetivos que não os de desenvolvimento de sistemas de informação, por exemplo, em uma reengenharia de processos ou em um processo de otimização de procedimentos e regras de negócio. Processos de negócio que regem as atividades operacionais e transacionais possuem características bastante diferentes dos processos voltados à análise de informações e ao apoio decisório.

O mapeamento de processos decisórios pode levar ao mapeamento de novos processos que possuem características transacionais e são usualmente voltados para a manutenção e o controle de atividades geradas a partir dos resultados das análises.

A análise de requisitos, quando aplicada ao desenvolvimento de um Data Warehouse, deve focar-se nos processos de análise e não nos processos de controle conseqüentes das análises realizadas. Ao mapear as necessidades de

análise devem-se mostrar aos usuários exatamente as operações que eles executam ou desejam executar, descritas em um formato direto e pouco generalizado. Um modelo de dados dimensional permite quase que uma infinidade de combinações de consultas, mas essa visão técnica e generalista nem sempre é bem compreendida pelos usuários.

5.1.2 Análise de requisitos x análise e projeto do Sistema

Com a análise de requisitos busca-se documentar as necessidades dos usuários, de forma que eles possam compreender facilmente o resultado dessa documentação. Aspectos técnicos que abrangem a definição de elementos de implementação, como a modelagem de tabelas de fatos e de dimensões, fazem parte do escopo da disciplina de análise e projeto da base de dados dimensional, que dará suporte ao desenvolvimento de aplicações de consulta e análise de dados. Nessa disciplina é que o analista ou o projetista da base de dados fará o mapeamento entre os elementos conceituais e os elementos físicos de implementação, sendo possível que uma unidade de análise seja mapeada em mais de uma tabela de fato ou que uma tabela de fato seja projetada para armazenar mais de uma unidade de conteúdo.

Neste trabalho foi apresentada uma proposta de nomenclatura não técnica para os elementos que compõem um sistema de informação voltado a análises de dados transacionais. Os conceitos propostos (unidade de análise, de conteúdo, de distribuição, etc.) assemelham-se aos conceitos técnicos de projeto de um Data Warehouse, como fatos, dimensões e hierarquias; essa relação, entretanto, não deve ser seguida com rigor (o mapeamento de “um para um” nem sempre é o mais recomendável).

Um outro exemplo para ilustrar esse mapeamento, além do Diretório de Grupos de Pesquisa no Brasil, é o projeto do Data Mart de Fomento (<http://lattes.cnpq.br/dmfomento>), em que as unidades de análise apuradas como requisitos de implementação foram mapeadas não para tabelas de fatos, mas sim para uma tabela de dimensão. Nesse projeto, apenas uma tabela de fato foi suficiente para a implementação de um Data Mart e a tabela de dimensão “Linha de Pesquisa” foi criada com hierarquias que, como funcionalidades para os usuários, foram disponibilizadas como unidades de análise. A Tabela 31 ilustra algumas das hierarquias (grande linha de atuação e linha de atuação) utilizadas na modelagem das unidades de análise apuradas no levantamento de requisitos.

Unidades de Análise	Descrição
Bolsas no País	Grande linha de atuação que reúne as linhas de bolsas de formação e qualificação de pesquisadores e as bolsas de estímulo à pesquisa e estímulo à inovação, de média e longa duração, concedidas pelo CNPq no País
• Formação e qualificação de pesquisadores no País	Linha de atuação que reúne as concessões do CNPq nas modalidades de bolsas de Iniciação Científica, Aperfeiçoamento, Estágio e Especialização, Mestrado, Doutorado, Pesquisa de Campo e Pós-Doutorado no País
• Estímulo à pesquisa	Linha de atuação que reúne as concessões do CNPq nas modalidades de bolsas de Produtividade em Pesquisa, Pesquisador Visitante, Recém-Doutor e Desenvolvimento Científico Regional, Apoio Técnico e Fixação de Doutores
• Estímulo à inovação para a competitividade	Linha de atuação que reúne as concessões do CNPq nas modalidades de bolsas de Iniciação Tecnológica e Industrial, Desenvolvimento Tecnológico e Industrial e Especialista Visitante
• Outras bolsas no País	Inclui as concessões do CNPq nas modalidades de bolsas de Programas de Capacitação Institucional (PCI), a saber: Desenvolvimento de Gestão Institucional, Especialista Visitante Institucional, Desenvolvimento Técnico e Industrial/PCI, Iniciação Técnica e Industrial/PCI. Inclui também as bolsas concedidas através do Convênio CNPq/Ministério da Saúde (Programa Saúde da Família), nas modalidades de Apoio Técnico em Saúde e Interiorização do Trabalho em Saúde
Bolsas no exterior	Grande linha de atuação do CNPq que reúne as bolsas de formação e qualificação de pesquisadores, de média e longa duração, concedidas pelo CNPq no exterior
• Formação e qualificação de pesquisadores no exterior	Linha de atuação que reúne as concessões do CNPq nas modalidades de bolsas de Aperfeiçoamento, Estágio e Especialização, Mestrado, Doutorado, Doutorado Sanduíche e Pós-Doutorado e Estágio Sênior no exterior
Fomento à pesquisa	Grande linha de atuação que reúne as linhas de apoio a projetos de pesquisa, apoio à editoração (publicações técnico-científicas) e a eventos datados, financiados pelo CNPq

• Apoio a projeto de pesquisa	Linha de atuação que reúne as concessões do CNPq nas modalidades de Auxílio Pesquisa, Auxílio Integrado, Auxílio Projeto Conjunto de Pesquisa e Apoio a Núcleos de Excelência-PRONEX
• Apoio a editoração	Linha de atuação que reúne as concessões do CNPq na modalidade de Auxílio Editoração (apoio a publicações técnico-científicas)
• Apoio a eventos	Linha de atuação que reúne as concessões do CNPq nas modalidades de Auxílio Especialista Visitante, Auxílio Pesquisador Visitante, Auxílio Participação em Eventos Científicos, Auxílio Promoção de Eventos Científicos, Auxílio Estágio/Especialização no País e Auxílio Estágio/Especialização no exterior, incluindo também as bolsas de curta duração

Tabela 31 - Unidades de Análise mapeadas como tabela de dimensão

5.2 Linguagem e comunicação

A análise de requisitos deve não somente exprimir a real necessidade de usuários para a implementação de um sistema de informação mas também ser expressa em uma linguagem compreensível para eles.

Devido à importância da disciplina de análise de requisitos no processo de desenvolvimento do DW, a possibilidade de serem cometidos erros de especificação nessa fase coloca em risco todo o projeto. Novamente justifica-se destacar a importância de valer-se de uma linguagem acessível e não técnica para a documentação dos requisitos.

A utilização de casos de uso permite a modelagem das principais análises (ações) que o usuário realiza ou deseja realizar com o apoio de um sistema de informação.

Os requisitos funcionais, quando detalhados na forma de uma descrição das possíveis análises a serem realizadas, permitem a validação dos requisitos de informação, para garantir que estes atendam às reais e atuais necessidades dos usuários.

5.3 Trabalhos futuros

Neste trabalho, o foco foi direcionado para apenas uma das etapas do processo de desenvolvimento de software (a análise de requisitos), aplicada à construção de um Data Warehouse, onde um estudo de robustez da taxonomia proposta seria útil. Outras etapas compõem o processo de desenvolvimento de sistemas de informação em geral, e, também como trabalhos futuros, sugere-se que essas etapas do processo sejam avaliadas e estendidas, com o objetivo de se criarem propostas para a aplicação de cada uma dessas etapas no desenvolvimento de projetos de DW.

5.4 Considerações finais

Da mesma forma que em um sistema de informação em geral, mostra-se a necessidade de se realizar a análise de requisitos em um projeto de DW, uma vez que este é o último documento não técnico que será lido (e possivelmente compreendido) pelos usuários finais da aplicação.

Ao formalizar um modelo de documentação, no qual o resultado a ser obtido da disciplina de análise de requisitos está mapeado, busca-se criar um facilitador na preparação e aplicação de técnicas como de entrevistas ou questionários. Sabendo-se o tipo de resposta que se deseja obter, torna-se menos árdua a tarefa de escolher as perguntas mais adequadas para que essas respostas sejam alcançadas.

REFERÊNCIAS

- ALHIR, Sinan Si. **UML in a nutshell**: A desktop quick reference. O'Reilly, 1998.
- BIO, S. R. **Sistemas de informação**: um enfoque gerencial. São Paulo: Atlas, 1985.
- CAMPOS, F. M. P. de. Os sistemas de informação e as modernas tendências da Tecnologia dos Negócios ERA. **Revista de Administração de Empresas**, v. 34, n. 6, p. 33-45, nov./dez. 1994.
- BOOCH, Grady; RUMBAUGH, James; JACOBSON, Ivar. **UML Guia do Usuário**. Rio de Janeiro: Campus, 2000.
- BRUCKNER, Robert; LIST, Beate; SCHIEFER, Josef. Developing Requirements for Data Warehouse Systems with Use Cases. In: SEVENTH AMERICAS CONFERENCE ON INFORMATION SYSTEMS, 2001. Disponível em: <<http://www.ifs.tuwien.ac.at/~bruckner/pubs/>>. Acesso em: 17 fev. 2003.
- FURLAN, José Davi. **Modelagem de objetos através da UML**: análise e desenho orientados a objetos. São Paulo: Makron Books, 1998.
- FOWLER, Martin; SCOTT, Kendall. **UML Essencial**: um breve guia para a linguagem padrão de modelagem de objetos. 2. ed. Porto Alegre: Bookman, 2000.
- GARCIA, Ana et al. FBCDataWare: um Data Warehouse para cardiologia. In: 4º SIMPÓSIO ARGENTINO DE INFORMÁTICA E SAÚDE – SADIO, 10., 2001. Disponível em: <http://www.sis.org.ar/tlibres/B/b_10.PDF>. Acesso em: 14 jan. 2003.
- GUIMARÃES, R. **Avaliação e fomento de C&T no Brasil**: propostas para os anos 90. Brasília: MCT/CNPq, 1994.
- IEEE. IEEE Guide to Software Requirements Specifications. IEEE Inc., 1984.
- INMON, Ana et al. **Como construir o Data Warehouse**. 2 ed., Rio de Janeiro: Campus, 1997.

- INMON, Bill. **The Infinite Data Warehouse**: Cost-Effectively Expanding Your Near-Line Storage. Mountain View, CA: Legato Systems, Inc. 2000. Disponível em: [<http://www.legato.com/resources/brochures/F064.pdf>]. Acesso em: 14 jan. 2003.
- JACOBSON, Ivar; BOOCH, Grady; RUMBAUGH, James. **The Unified Software Development Process**. Addison Wesley, 1999.
- KIMBALL et al. **The Data Warehouse Lifecycle Toolkit**: Expert Methods For Designing, Developing and Deploying Data Warehouses. New York: John Wiley & Sons Inc., 1998.
- KIMBALL, Ralph; ROSS, Margy. **The Data Warehouse Toolkit**: The complete guide to dimensional modeling. 2. ed. USA: Wiley, 2002.
- KIMBALL, Ralph. **Drilling Down, Up, and Across**: Understanding the vocabulary of navigating dimensions. DBMS OnLine. 1996. Disponível em: <<http://www.dbmsmag.com/9603d05.html>>. Acesso em: 14 abr.2003.
- KOTONYA, Gerald; SOMMERVILLE, Ian. **Requirements Engineering**: Processes and Techniques. John Wiley & Sons, 1998.
- LAGO, Sandra M. S. **Uma metodologia para avaliação da necessidade e da viabilidade de implantação de ferramentas de business intelligence**. 2002. Dissertação (Mestrado em Engenharia de Produção) – Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, 2002.
- LAUDON, Kenneth; LAUDON, Jane. **Management Information Systems-Organization and Technology**. 4. ed. EUA: Macmillan Publishing Company, 1996.
- LIST, Beate; SCHIEFER, Josef; TJOA, A Min. **Process-Oriented Requirement Analysis Supporting the Data Warehouse Design Process**: A Use Case Driven Approach. Institute of Software Technology (E188), Vienna University of Technology, 2001. Favoritenstr. 9-11 / 188, A-1040 Vienna, Austria. LNCS 1873, p. 593 ff. Disponível em: <<http://link.springer.de/link/service/series/0558/bibs/1873/18730593.htm>>. Acesso em: 17 fev. 2003.
- MACHADO, F. M. R. **Projeto de Data Warehouse**: uma visão multidimensional. São Paulo: Érica, 2000.

MARTINS, G. M.; GALVÃO, G. Diretório de Grupos de Pesquisa no Brasil: Perspectivas de Fomento e Avaliação. **Educação Brasileira**, v. 16, n. 33, p. 11-29, 1994.

MELLO, João A. B. de. **Uma proposta de modelo de dados para suporte ao processamento transacional e de apoio informacional simultaneamente**. 2002. Dissertação (Mestrado em Engenharia de Produção) – Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, 2002.

O'BRIAN, J. A. **Sistemas de informação e as decisões gerenciais na era da internet**. 9. ed. São Paulo: Saraiva, 2001.

PAIM, Fábio R. S.; CASTRO, Jaelson F. B. Enhancing Data Warehouse Design with the NFR Framework. In: WER02 - V WORKSHOP ON REQUIREMENTS ENGINEERING, 2002, Valencia, Espanha. Disponível em: <<http://www.cin.ufpe.br/~ler/frsp/fabio-sub.pdf>>. Acesso em: 21 jul. 2003.

POLLONI, E. G. F. **Administrando sistemas de informação**. São Paulo: Futura, 2000.

RATIONAL Software Corporation. **Rational Unified Process**. Rational, 2002. 1 CD-ROM.

REZENDE, D. A.; ABREU, A. F. **Tecnologia da informação aplicada a sistemas de informações empresariais**: o papel estratégico da informação e dos sistemas de informação nas empresas. São Paulo: Atlas, 2000.

SELL, Denilson. **Uma arquitetura para distribuição de componentes tecnológicos de sistemas de informações baseados em Data Warehouse**. 2001. Dissertação (Mestrado em Engenharia de Produção) – Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, 2001.

THOMSEN, Erik. **OLAP**: Construindo sistemas de informações multidimensionais. Rio de Janeiro: Campus, 2002.

WG Systems. **Arquitetura de um sistema de BI**. Disponível em <http://www.wgsystems.com.br/bi/bi_arquitetura.htm>. Acesso em: 28 mar. 2003.