

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em
Engenharia de Produção

Tatiana Sousa Gonzaga

**UMA METODOLOGIA PARA O DESENVOLVIMENTO DE INSTRUMENTOS DE
ANÁLISE MULTIDIMENSIONAL DA INFORMAÇÃO EM PROJETOS DE
GOVERNO ELETRÔNICO VOLTADO AO CIDADÃO**

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

Orientador: Prof. Aran Bey Tcholakian Morales, Dr.

Florianópolis
2005

Tatiana Sousa Gonzaga

**UMA METODOLOGIA PARA O DESENVOLVIMENTO DE INSTRUMENTOS DE
ANÁLISE MULTIDIMENSIONAL DA INFORMAÇÃO EM PROJETOS DE
GOVERNO ELETRÔNICO VOLTADO AO CIDADÃO**

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, fevereiro de 2005.

Prof. Edson Pacheco Paladini, Dr.
Coordenador do Programa

BANCA EXAMINADORA

Prof. Aran Bey Tcholakian Morales, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof. José Leomar Todesco, Dr.
Universidade Federal de Santa Catarina

Prof. Roberto C. S. Pacheco, Dr.
Universidade Federal de Santa Catarina

Prof. Vinícius Medina Kern, Dr.
Universidade Federal de Santa Catarina

Aos meus pais

AGRADECIMENTOS

A Deus, que procura sempre me mostrar o caminho e fazendo-me crer que tudo vai dar certo.

Ao Grupo Stela, por acreditar em mim e ser o principal formador da minha vida profissional.

Aos meus pais, Ivonete e Wolney Gonzaga, por serem os melhores pais do mundo e por terem sido os principais responsáveis pela pessoa que me tornei.

Ao meu irmão Flávio, por ser tão especial e, claro, pelas constantes consultorias.

Ao meu irmão Fabian, que embora me dê trabalho, me ajuda muito a crescer e a compreender a vida.

Aos meus avós que sempre me mencionam em suas orações.

Ao Daniel Tomazini Silva, companheiro em todos os momentos, pelo apoio, carinho e compreensão.

Ao Aran Bey Tcholakian Morales, que de orientador e coordenador passou a ser um excelente companheiro de trabalho e um amigo. É realmente difícil descrever em palavras a admiração que tenho por ele.

Aos meus companheiros do Grupo Stela, especialmente a Denílson Sell, Fabiano Beppler, Ricardo Rieke, Sandra Martins, Luciano Liduário, Rodolfo Almeida e Kristiany Zamai, que nunca negaram ajuda e contribuem constantemente para o meu aprendizado.

Aos meus amigos, que com sua presença, gestos e palavras colaboraram com minha alegria diária.

Enfim, a todos aqueles que contribuíram de alguma forma para que este trabalho tenha sido concluído.

RESUMO

GONZAGA, Tatiana Sousa. **Uma metodologia para o desenvolvimento de instrumentos de análise multidimensional da informação em projetos de governo eletrônico voltados ao cidadão**. 2005. 130 f. Dissertação (Mestrado em Engenharia de Produção). Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2005.

O projeto de sistemas multidimensionais carece de uma linguagem comum para expressar requisitos e especificações técnicas. A partir do estudo de *Data Warehouse* (DW) e OLAP, esta dissertação propõe uma metodologia para que tais sistemas traduzam os requisitos de análise dos usuários, apresentando informações mais próximas do domínio do usuário. A nomenclatura técnica usual de projetos de DW é comparada à nomenclatura conceitual apresentada por uma pesquisa precursora do presente trabalho. A partir da observação das unidades de informação percebidas por peritos em domínios de problema, é concebido um modelo composto por assunto, unidade de análise, unidade de filtro e unidade de conteúdo, que guardam correspondência com os conceitos técnicos de *Data Mart*, fato, dimensão e medida. A metodologia é completada por uma seqüência de etapas a realizar para a modelagem de DW, usando os quatro conceitos introduzidos, além de ser adaptada aos conceitos de usabilidade na *Web*. A principal contribuição desta dissertação é a criação de uma metodologia aplicável para a implementação de instrumentos de análise multidimensional da informação. A metodologia inclui uma nova nomenclatura conceitual que permite que analistas de sistemas e analistas de negócio comuniquem-se sem recorrer ao jargão técnico ou específico do negócio, facilitando a tradução dos requisitos em processos de recuperação e apresentação da informação.

Palavras-chave: *Data Warehouse*, OLAP, apoio à decisão, governo eletrônico.

ABSTRACT

GONZAGA, Tatiana Sousa. **Uma metodologia para o desenvolvimento de instrumentos de análise multidimensional da informação em projetos de governo eletrônico voltados ao cidadão.** 2005. 130 f. Dissertação (Mestrado em Engenharia de Produção). Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2005.

Multidimensional systems lack of a common language for expressing requirements and technical specifications. By studying Data Warehouse (DW) and OLAP, this thesis proposes a methodology for these systems to translate the users' analysis requirements, presenting information that is closer to what the user can manage. The technical nomenclature used in DW projects is compared to a conceptual nomenclature presented by a precursor research from the present work. By observing the information units noticed by experts, a model that contains subject, analysis unit, filter unit and content unit was implemented, which corresponds to the following technical concepts: Data Mart, fact, dimension and measure. The methodology is completed by a set of steps for implementing a DW data model, using the four presented concepts, and is adapted to Web usability concepts. The main contribution of this work is the creation of an applicable methodology for the implementation of multidimensional analysis instruments. The methodology includes a new conceptual nomenclature that allows system analysts and business analysts to communicate in a common language, facilitating the translation of the requirements in information extraction and presentation.

Keywords: Data Warehouse, OLAP, decision support, electronic government.

SUMÁRIO

RESUMO	5
ABSTRACT	6
LISTA DE ILUSTRAÇÕES	10
LISTA DE QUADROS	12
LISTA DE ABREVIATURAS E SIGLAS	13
1 INTRODUÇÃO	15
1.1 Motivação	15
1.2 Objetivo geral.....	19
1.3 Objetivos específicos.....	19
1.4 Justificativa	20
1.5 Método	22
1.6 Estrutura do trabalho	23
1.7 Delimitação da dissertação	24
2 SISTEMAS DE INFORMAÇÃO	26
2.1 Introdução	26
2.2 Arquitetura de sistemas de informação	26
2.3 Dado, informação e conhecimento	28
2.4 Sistemas de informação	29
2.4.1 Evolução	30
2.4.2 Classificação.....	31
2.5 Business Intelligence e as Tecnologias de Informação.....	33
2.5.1 Definição.....	33
2.5.2 Arquitetura de um sistema de <i>Business Intelligence</i>	33
2.6 Data Warehouse	35
2.6.1 Definições, objetivos e características	35
2.6.2 Processo de desenvolvimento de um DW	36
2.6.2.1 Modelo dimensional.....	38
2.6.2.2 Projeto físico	40
2.6.2.3 Projeto e desenvolvimento da área de <i>data staging</i>	41
2.7 Considerações finais	41
3 Área de apresentação	43
3.1 Introdução	43

3.2	Conceitos OLAP	43
3.3	Funcionalidades de uma ferramenta OLAP	45
3.4	Características do processamento OLAP	47
3.5	Arquiteturas	49
3.6	Usabilidade na Web	50
3.6.1	CrITÉrios de avaliaÇo da usabilidade	52
3.6.1.1	<i>Checklists</i>	52
3.6.1.2	CrITÉrios de Jakob Nielsen	53
3.7	WOLAP	55
3.7.1	WOLAP e usabilidade	56
3.8	ConsideraÇes finais	56
4	METODOLOGIA PARA O DESENVOLVIMENTO DE INSTRUMENTOS DE ANLISE MULTIDIMENSIONAL	58
4.1	IntroduÇo	58
4.2	Metodologia proposta	58
4.2.1	Nomenclatura conceitual	61
4.2.1.1	Assunto	62
4.2.1.2	Unidade de anlise	62
4.2.1.3	Unidade de contedo	63
4.2.1.4	Unidade de filtro	65
4.2.1.5	ComparaÇo entre nomenclaturas	67
4.2.2	Etapas	68
4.3	ConsideraÇes finais	71
5	APLICAÇO – DIRETRIO DE GRUPOS DE PESQUISA NO BRASIL	73
5.1	IntroduÇo	73
5.2	Plataforma Lattes: uma plataforma de E-gov para gesto de CT&I	73
5.2.1	Unidades de informaÇo, ontologias e padronizaÇo XML	75
5.2.2	Fontes e sistemas de informaÇo	77
5.2.3	Portais e serviÇos <i>Web</i>	78
5.2.4	Sistemas de conhecimento e agentes	78
5.3	O Diretrio de Grupos de Pesquisa no Brasil	81
5.3.1	As fases do DGP	83
5.3.2	O <i>Data Mart</i> do Diretrio de Grupos de Pesquisa	85
5.3.3	Instrumentos de informaÇo do DGP	89
5.3.4	Plano Tabular	91

5.3.4.1	Aplicação da metodologia proposta	91
5.3.4.2	Funcionamento	96
5.3.4.3	Operações OLAP	99
5.3.4.4	Usabilidade	101
5.3.4.5	Integração do Plano Tabular com outros sistemas	106
5.3.5	Vantagens da utilização da metodologia proposta no DGP	108
5.4	Considerações finais	110
6	CONCLUSÕES E TRABALHOS FUTUROS	112
6.1	Conclusões	112
6.2	Trabalhos futuros	114
	REFERÊNCIAS BIBLIOGRÁFICAS	116
	APÊNDICE A – <i>Data Mart</i> do Diretório de Grupos de Pesquisa no Brasil.....	121
	APÊNDICE B – Evolução do Plano Tabular	125
	APÊNDICE C – Operações OLAP do Plano Tabular	128

LISTA DE ILUSTRAÇÕES

Figura 1 - Integração entre os analistas e os sistemas DW/OLAP	18
Figura 2 - Representação da metodologia utilizada no trabalho	23
Figura 3 - Visões de um ASI	28
Figura 4 - Arquitetura de um sistema de <i>Business Intelligence</i>	34
Figura 5 - Ciclo de vida de um projeto de DW	36
Figura 6 - Um modelo estrela	39
Figura 7 - Contexto da metodologia proposta em sistemas DW/OLAP	60
Figura 8 - Possíveis tabelas de fato para o assunto “Vendas”	63
Figura 9 - Possíveis medidas para a tabela Fato_Vendas	64
Figura 10 - Possíveis tabelas de dimensão para o assunto “Vendas”	66
Figura 11 - <i>Data Mart</i> “Vendas”	67
Figura 12 - Etapas da metodologia proposta	70
Figura 13 - Arquitetura conceitual proposta para projetos de <i>E-gov</i>	74
Figura 14 - Fragmento do DMGRUPO	88
Figura 15 - Fragmento do DMGRUPO	89
Figura 16 - Plano Tabular e suas unidades de informação	95
Figura 17 - Consulta da unidade de análise “grupos de pesquisa” pela unidade de filtro “região”	97
Figura 18 - Consulta da unidade de análise “pesquisadores” pela unidade de filtro “região” e unidade de conteúdo “formação”	98
Figura 19 - Apresentação das informações textuais para a unidade de análise “grupos”	99
Figura 20 - Elementos de suporte de navegação do Plano Tabular	102
Figura 21 - Integração do Plano Tabular com a Busca Textual	107
Figura 22 - Integração do Plano Tabular com o Lattes RedesGP	108
Figura 23 - Plano Tabular em sua versão 2.0	125
Figura 24 - Exemplo de consulta no Plano Tabular versão 2.0	125
Figura 25 - Plano Tabular em sua versão 3.0	126
Figura 26 - Exemplo de consulta do Plano Tabular versão 3.0	126
Figura 27 - Plano Tabular em sua versão 4.0	127
Figura 28 - Plano Tabular em sua versão 5.0	127
Figura 29 - <i>Drill Down</i>	128

Figura 30 - Drill Out	128
Figura 31 - Slice	129
Figura 32 - Filtering	129
Figura 33 - Paging	130
Figura 34 - Sort	130

LISTA DE QUADROS

Quadro 1 - A evolução dos sistemas de informação.....	31
Quadro 2 - Tipos de sistemas de informação.....	32
Quadro 3 - Funcionalidades OLAP	47
Quadro 4 - Benefícios das arquiteturas MOLAP e ROLAP.....	50
Quadro 5 - Recomendações de usabilidade na <i>Web</i>	53
Quadro 6 - Recomendações mais atuais de usabilidade na <i>Web</i>	54
Quadro 7 - Nomenclatura proposta	61
Quadro 8 - Relação entre a nomenclatura conceitual e técnica	68
Quadro 9 - Passos de navegação do sistema	70
Quadro 10 - Comparação dos elementos do <i>Data Mart</i> com a nomenclatura conceitual e a técnica	86
Quadro 11 - Unidades de análise, conteúdo e filtro do DMGRUPO	88
Quadro 12 - Unidades de análise, de filtro e de conteúdo do Plano Tabular	94
Quadro 13 - Passos de navegação no Plano Tabular	96
Quadro 14 - Operações OLAP disponíveis no Plano Tabular	100
Quadro 15 - Suportes de navegação do Plano Tabular	103
Quadro 16 - Avaliação do sistema com relação aos critérios de usabilidade na <i>Web</i>	105
Quadro 17 - Comparação de atividades executadas para apresentação de dados no DGP	110

LISTA DE ABREVIATURAS E SIGLAS

ASI – Arquitetura de Sistemas de Informação
BD – Banco de Dados
BI – *Business Intelligence*
CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico
C&T – Ciência e Tecnologia
CT&A – Científica, Tecnológica e Artística
CT&I – Ciência, Tecnologia e Inovação
CV-Lattes – Currículo Lattes
CWM – *Common Warehouse Metamodel*
DGP – Diretório de Grupos de Pesquisa
DM – *Data Mining*
DMGRUPO – *Data Mart* de Grupos de Pesquisa
DOLAP – *Desktop On-line Analytical Processing*
DSS – *Decision Support System*
DW – *Data Warehouse*
EIS – Sistema de Informação para Executivos
ERP – *Enterprise Resource Planning*
ETL – Extração, Transformação e Carga de Dados
GUI – *Graphical User Interface*
HOLAP – *Hybrid On-line Analytical Processing*
IC – Iniciação Científica
IES – Instituições de Ensino Superior
ISO – *International Organization for Standardization*
LMPL – Linguagem de Marcação da Plataforma Lattes
MOLAP – *Multidimensional On-line Analytical Processing*
OLAP – *On-line Analytical Processing*
OLTP – *On-line Transaction Processing*
PIBIC – Programa Institucional de Bolsas de Iniciação Científica
ROLAP – *Relational On-line Analytical Processing*
SAD – Sistema de Apoio a Decisão
SI – Sistema de informação
SPT – Sistemas de Processamento Transacional

SQL – *Structure Query Language*

TI – Tecnologia de Informação

UF – Unidade da Federação

URL – *Uniform Resource Locator*

WAN – *Wide Area Network*

WOLAP – *Web On-line Analytical Processing*

XML – *eXtended Markup Language*

1 INTRODUÇÃO

1.1 *Motivação*

Muitas organizações possuem projetos de informática, representados pelos seus sistemas transacionais, ou seja, os sistemas de apoio às operações do dia-a-dia. Porém, a existência da informação operacional não garante sua oferta de informação necessária à tomada de decisão gerencial. Com a utilização de sistemas transacionais, pode-se afirmar que a informação está disponível, mas a um custo de processamento muito elevado, o que pode comprometer o funcionamento desses sistemas por motivos de desempenho. Além disso, a extração e a interpretação de informações a serem utilizadas para o apoio gerencial dependem de recursos humanos especializados, e essa dependência dificulta o acesso às informações pelos tomadores de decisão e, principalmente, de usuários com interesse vertical (e.g. cidadão em aplicação de governo eletrônico).

Quando as organizações perceberam esse fato, começaram a desenvolver projetos de informação como complemento aos de informática e, para isso, passaram a utilizar uma tecnologia denominada *Data Warehouse* (DW). A crescente utilização dos DW está relacionada à necessidade do domínio de informações estratégicas para garantir respostas rápidas (MACHADO, 2000), evitando problemas de confiabilidade nas informações fornecidas e de desempenho dos sistemas transacionais.

Para realizar a apresentação das informações gerenciais, os projetos de informação incorporam, também, ferramentas de processamento analítico da informação (OLAP – *On-line Analytical Processing*), direcionadas aos usuários de

negócio que não possuem conhecimentos técnicos ou detalhes do projeto de informação.

Entre os potenciais usuários temos o cidadão, que não possui o conhecimento sobre o negócio em questão no mesmo nível do analista de negócio. Portanto, as ferramentas OLAP disponíveis, associadas aos projetos de DW, acabam sendo complexas e de difícil manipulação e navegabilidade pelos usuários do sistema.

A complexidade existe desde o processo de levantamento de requisitos para o projeto de DW, pois existe uma dificuldade de comunicação entre os analistas de sistemas e os analistas de negócio. Na fase de definição dos requisitos de negócio do processo de construção de um DW faz-se uma verificação das necessidades e exigências do usuário de negócio em conjunto com a realidade dos dados (KIMBALL; ROSS, 2002). O objetivo de se apurarem os requisitos de um DW é prover uma especificação compreensível dos principais aspectos técnicos e permitir que a implementação do sistema se complete com sucesso.

Pensando-se na visão do usuário de negócio e na construção de sistemas de análises multidimensionais que reflitam requisitos do negócio, é possível destacar uma primeira pergunta de pesquisa:

- ✍ Como sistematizar a apresentação de análises multidimensionais da informação mais próxima da visão dos usuários, analistas de negócio e cidadãos, em termos de manipulação e navegabilidade, permitindo ao mesmo tempo usufruir todas as potencialidades que as análises desse tipo possibilitam?

A Figura 1 ilustra os processos do ambiente do problema. No processo 1, o analista de sistemas faz o levantamento das necessidades e exigências da organização junto ao analista de negócios. De acordo com a realidade dos dados e com esses requisitos, o analista de sistemas implementa os repositórios de DW e as ferramentas de processamento analítico OLAP (processos 2 e 3). No processo 4, as informações extraídas dos repositórios de dados são apresentadas aos analistas de negócios e ao cidadão comum. Assim, embora uma nomenclatura que auxilie na identificação dos requisitos de um DW e que defina a interface de sistemas multidimensionais possa responder à primeira questão de pesquisa (itens 1 e 4 da Figura 1), ainda é necessário que os sistemas DW/OLAP compreendam essa nomenclatura (itens 2 e 3 da Figura 1), o que nos leva a formular uma segunda pergunta de pesquisa:

- ✍ O que deve ser feito para que os sistemas DW/OLAP traduzam os requisitos de análise do usuário em processos de recuperação e apresentação de informações?

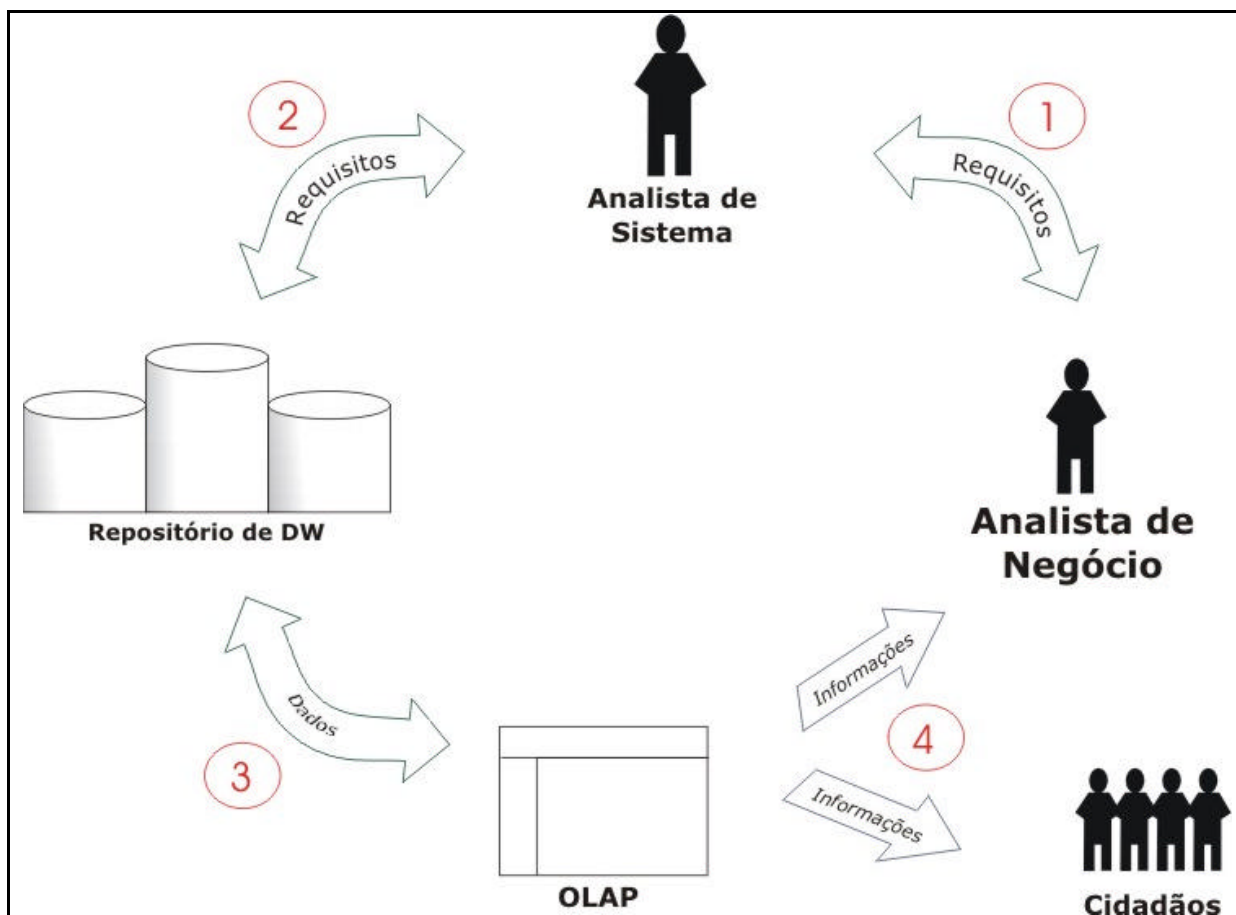


Figura 1 - Integração entre os analistas e os sistemas DW/OLAP

Fonte: Da autora

Com o intuito de responder a essas perguntas, este trabalho estabelece uma metodologia de definição do domínio de conhecimento que se deseja tratar, por meio do desenvolvimento de: (a) uma nomenclatura conceitual simples e de fácil entendimento pelos integrantes do sistema, a qual especifique unidades de informação que podem ser cruzadas e analisadas, e auxilie na análise resultante do trabalho do usuário; e (b) um conjunto de etapas que permitam a construção de sistemas de análises multidimensionais, fazendo uso da nomenclatura definida. Um dos princípios da metodologia é permitir que a apresentação das informações não se desvie de seu principal objetivo: o apoio aos processos de tomada de decisão.

1.2 *Objetivo geral*

O objetivo geral deste trabalho é propor uma metodologia para a construção de sistemas de análises multidimensionais para a *Web* e voltados aos cidadãos, através da tradução dos requisitos de análise do usuário incluídos na estrutura de um *Data Warehouse*. Isso deve facilitar a recuperação e apresentação das informações do domínio do problema que o usuário possui e que seja aplicável em diferentes projetos de informação nos quais se utiliza a estrutura DW/OLAP.

1.3 *Objetivos específicos*

- ✍ Desenvolvimento de uma nomenclatura conceitual orientada a usuários de negócio que traduza os requisitos de análise do usuário em processos de recuperação da informação.
- ✍ Desenvolvimento de um conjunto de etapas para a construção de sistemas de *Data Warehouse* e sistemas de *front-end*.
- ✍ Definição de unidades de informação identificáveis em sistemas de DW e OLAP que permitam a construção de indicadores e possibilitem ao usuário moldar sua própria consulta, utilizando-a em suas análises e em seus processos decisórios.
- ✍ Aplicação da metodologia em caso prático, visando analisar aspectos de implementação e adequação a um domínio específico.

1.4 Justificativa

A Engenharia de Produção é uma área voltada a projeto, desenvolvimento e gerência de sistemas que envolvam pessoas, materiais, tecnologias e o próprio ambiente em que esses itens se encontram. Associada ao avanço da produção de serviços e à evolução da Tecnologia da Informação (TI), analisa e avalia a aplicação de seus sistemas e processos com o intuito de suprir as necessidades do mercado e da sociedade. Assim, considerando-se esse contexto e a necessidade da informação no processo decisório para manter as organizações competitivas no mercado, é válido ressaltar a importância em se investir na área de engenharia e gestão do conhecimento e no desenvolvimento de instrumentos de análise competentes.

Segundo Laudon e Laudon (1996), a TI reúne os meios necessários para gerenciar e utilizar informação, incluindo computação, telecomunicações e sistemas, contemplando, portanto, componentes de hardware, infra-estrutura e periféricos, software e a gestão de dados e informação. Pan e Leidner (2002) afirmam que a TI passou a incluir a gestão do conhecimento proporcionado pela interação de seus componentes, tendo se tornado, segundo Tait e Pacheco (2000), um elemento integrante do centro estratégico das empresas e um fator gerador de conhecimento.

No entanto, um dos principais desafios para a área de TI tem sido encontrar a melhor metodologia para projetos de desenvolvimento de sistemas de informação para governos (TAIT; PACHECO, 2000). Embora já sejam significativos os avanços na organização, publicação e produção de informações em páginas de governo, muitos desafios no tocante a projetos de informação permanecem recorrentes em todo novo desenvolvimento (PACHECO, 2003).

Portanto, os desafios referentes aos projetos de plataformas eletrônicas de governos perduram mesmo diante da diversidade de tecnologias e metodologias de construção de sistemas de informação. Dado que investimentos de governos têm se intensificado nos últimos anos (especialmente de países desenvolvidos), são necessários modelos específicos para projetos de *E-Gov*, que sejam capazes de contemplar as melhores metodologias de Engenharia de Software com os recursos mais adequados da Tecnologia da Informação (PACHECO, 2003).

Uma das tecnologias utilizadas neste trabalho é o *Data Warehouse* (DW), o qual cumpre um importante papel na gestão de conhecimento organizacional. Por ser um repositório de dados histórico, integrado e orientado a assuntos de interesse da organização, é capaz de apresentar as informações necessárias para as tomadas de decisão gerencial. Este trabalho está contextualizado na camada de sistemas de conhecimento da Plataforma Lattes, uma plataforma de *E-Gov* cuja arquitetura foi proposta por Pacheco (2003). Segundo o autor, sistemas de conhecimento são projetados para gerar novos conhecimentos a partir das fontes de informação da plataforma, sendo essas fontes os diversos tipos de repositório de dados, inclusive DW. No entanto, a Engenharia do Conhecimento chama a atenção para o fato de o DW guardar informação e que para que esta gere conhecimento é secundário lhe agregar contexto e semântica. Assim, uma metodologia que visa aprofundar o DW/OLAP das necessidades e da compreensão que o usuário tem do domínio é um avanço significativo no preenchimento do hiato entre DW e Engenharia do Conhecimento no sentido de dar à informação do DW valor de conhecimento.

Portanto, através do estudo dessa e de outras tecnologias de armazenamento, extração e apresentação de informações, propõe-se uma metodologia para a construção de sistemas de análise multidimensional que possam

ser utilizados por qualquer tipo de usuário, não somente o usuário que compreenda DW ou OLAP. Por conseguinte, a comunicação entre analistas de sistemas e analistas de negócios fica facilitada, e os cidadãos poderão compreender e utilizar o sistema.

Com isso, espera-se aproximar os resultados da recuperação e apresentação de informações essenciais aos diferentes processos de tomada de decisão, de forma que propiciem benefícios às organizações da sociedade do conhecimento e atendam às suas necessidades e de seus analistas de negócio.

1.5 Método

Visando à concepção e ao desenvolvimento deste trabalho e com o intuito de alcançar os objetivos identificados, foram definidas as seguintes etapas:

1. estudo sobre sistemas de informação, sobre *Business Intelligence* e sobre os principais conceitos relacionados à tecnologia *Data Warehouse*;
2. estudo sobre a tecnologia OLAP, seus conceitos e principalmente suas funcionalidades.
3. estudo sobre a usabilidade na *Web* e seus critérios de avaliação;
4. desenvolvimento da metodologia: nesta etapa são estabelecidos uma nomenclatura conceitual e um conjunto de etapas para o desenvolvimento de sistemas de análise multidimensionais; e
5. aplicação da metodologia proposta na construção de instrumentos de informação a uma plataforma governamental.

A Figura 2 a seguir apresenta a metodologia descrita.

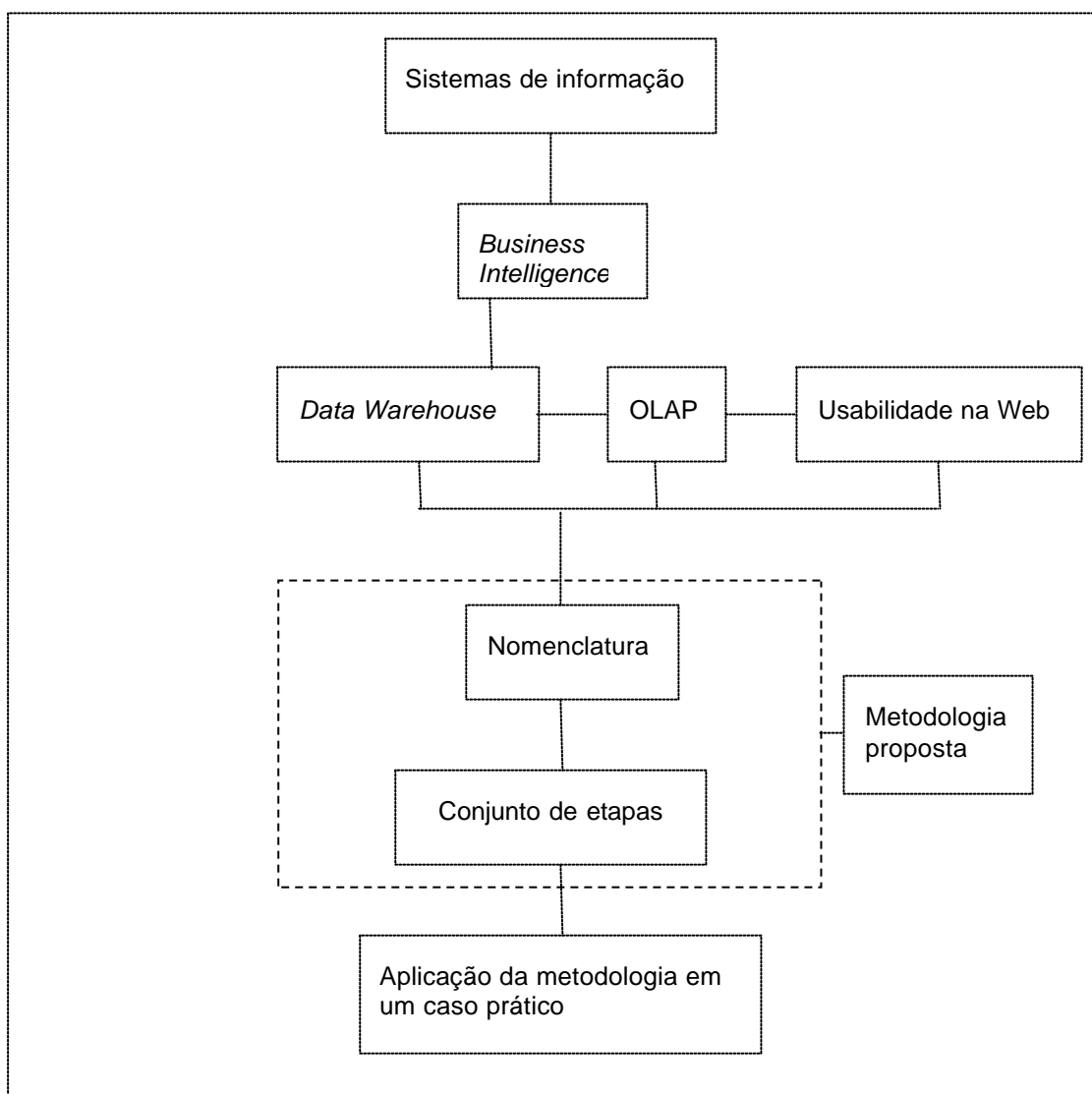


Figura 2 - Representação da metodologia utilizada no trabalho

Fonte: Da autora

1.6 Estrutura do trabalho

No segundo capítulo, são apresentados os principais conceitos referentes aos sistemas de informação e às tecnologias relacionadas a eles, enfatizando-se o papel de ambos nos processos de tomada de decisão das organizações. Em seguida, são mostrados alguns conceitos relevantes sobre a tecnologia *Data Warehouse*.

No terceiro capítulo, são explicitados aspectos teóricos sobre OLAP, suas características, algumas de suas funcionalidades e seus tipos de arquitetura. São

também apresentados conceitos básicos de usabilidade na *Web*, devido à sua importância na apresentação das informações ao usuário final.

No quarto capítulo, descreve-se a metodologia proposta neste trabalho, relacionando-a com os conceitos da tecnologia OLAP e de usabilidade na *Web* mencionados anteriormente. Também nesse capítulo serão respondidas as problemáticas levantadas neste primeiro capítulo.

Para complementar e exemplificar a proposta, o quinto capítulo apresenta a metodologia aplicada em um caso real e mostra sua relação com os conceitos abordados no trabalho.

No quinto capítulo, apresenta-se a aplicação da metodologia proposta a um caso real. Discutem-se questões práticas e implementativas da aplicação da metodologia proposta.

No capítulo seis, são relatadas as conclusões referentes aos resultados da pesquisa bem como são feitas algumas recomendações e sugestões importantes para o prosseguimento de estudos e pesquisas na área abordada.

Por fim, listam-se as referências bibliográficas utilizadas e os apêndices elaborados para o desenvolvimento deste trabalho.

1.7 Delimitação da dissertação

Os conceitos das tecnologias OLAP e *Web*, individualmente, apresentam uma literatura extensa. No entanto, a literatura não distingue os conceitos de aplicações OLAP e aplicações OLAP para a *Web* (WOLAP). Essa diferença se faz necessária basicamente quando abordamos os tópicos de usabilidade (porque os ambientes *desktop* e *Web* são diferentes) e para ser útil ao usuário cidadão. Por esse motivo,

consideramos a falta de literatura sobre aplicações WOLAP uma limitação de pesquisa.

A metodologia proposta neste trabalho é específica para sistemas que utilizam as tecnologias *Data Warehouse*, OLAP e *Web*, visto que é baseada nas necessidades de análises do usuário cidadão. Essa limitação não significa dizer que não é possível desenvolver uma nova metodologia para o caso da utilização de outras tecnologias ou mesmo fazer uma adaptação com a atual.

Por fim, a metodologia e, principalmente, o caso real proposto neste trabalho são desenvolvidos para oferecer apenas algumas operações OLAP. Assim como as tecnologias utilizadas, essa escolha também foi feita com base nas necessidades do usuário final.

2 SISTEMAS DE INFORMAÇÃO

2.1 Introdução

A informação está presente na grande maioria das atividades de uma organização, assumindo a base para suas análises gerenciais e seus processos de tomada de decisão. Os sistemas de informação (SI) cumprem um importante papel no fornecimento dessas informações.

Os SI podem ser contextualizados no que chamamos de Arquitetura de Sistemas de Informação (ASI), que engloba tanto os aspectos organizacionais e de negócio quanto as Tecnologias de Informação, os próprios sistemas de informação e os usuários envolvidos (TAIT, 2000). Dessa forma, a visão sobre os SI foi modificada, os quais passaram de simples sistemas de apoio a tarefas rotineiras para sistemas de apoio estratégico e gerencial da organização.

Neste trabalho, o conceito de SI adotado aborda a visão de apoio ao processo decisório da organização utilizando os conceitos de repositório de dados *Data Warehouse*, um importante componente da arquitetura do conceito de *Business Intelligence* (BI). Detalhes relacionados a outros conceitos tratados pela ASI não são mencionados por estarem fora do escopo deste trabalho.

2.2 Arquitetura de sistemas de informação

Arquitetura de Sistemas de Informação (ASI) é um conjunto de elementos cuja finalidade é proporcionar um mapeamento da organização no tocante aos itens envolvidos no processo de desenvolvimento e implantação de SI (ZACHMAN, 1987).

Nesse conceito, também, destaca-se a necessária integração entre a visão organizacional, os sistemas de informação, a TI, os negócios e os usuários, que, ao fazerem parte de uma ASI, contribuem para o desenvolvimento e uso adequado dos sistemas de informação (TAIT, 1994).

Segundo Earl (1993), a estrutura tecnológica, a qual o autor chama de arquitetura, compreende quatro elementos: (1) a computação (hardware e software), (2) as comunicações (redes de telecomunicações), (3) os dados (relacionamentos entre eles) e (4) as aplicações. Já de acordo com Laudon e Laudon (1996), a arquitetura de informação é a forma particular que a TI tem, em uma organização, de alcançar metas selecionadas ou funções, estando estritamente ligada ao aspecto tecnológico pelo arranjo de hardware, software, dados e redes de computadores, na linha seguida por Earl (1993).

Tait (1994) propõe uma estrutura simplificada de ASI organizacional, conforme mostra a Figura 3. Para chegar nessa proposta, a autora partiu da premissa de que uma ASI deve combinar três componentes básicos: (1) organização, (2) sistemas e (3) tecnologia, de forma que se possa incluir ainda negócios e usuários. Portanto, a estrutura se apresenta em cinco níveis: (1) visão da organização, (2) visão dos negócios, (3) visão dos sistemas, (4) visão da tecnologia e (5) visão dos usuários.

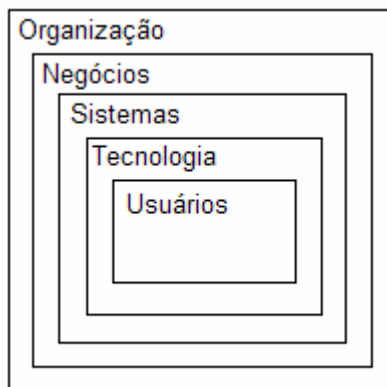


Figura 3 - Visões de um ASI
Fonte: TAIT, 1994

2.3 Dado, informação e conhecimento

Os termos “dado”, “informação” e “conhecimento” são freqüentemente confundidos e utilizados incorretamente devido à proximidade de seus significados. O’Brien (2003, p. 23) define “dado” como fato ou observação crua, normalmente, sobre fenômenos físicos ou transações de negócios, ou seja, recursos de matéria-prima que são processados para construir o produto chamado de informação.

Já informação é o resultado de dados processados e convertidos em um contexto significativo e útil para usuários finais específicos (O’BIEN, 2003, p. 23). Para Laudon e Laudon (1999, p. 10), informação é um conjunto de dados aos quais seres humanos deram forma para torná-los significativos e úteis.

No tocante ao conhecimento, este caracteriza um conjunto de ferramentas conceituais e de categorias usadas pelos seres humanos para criar, coleccionar, armazenar e compartilhar a informação (LAUDON; LAUDON, 1999, p. 10).

Dado, informação e conhecimento possuem uma importante relação. Davenport e Prusak (1999) dão maior ênfase à informação e afirmam que ela serve

como conexão entre os dados brutos e o conhecimento que se pode obter eventualmente.

Segundo Malhorta (1998), a confusão entre os conceitos sobre os termos dado, informação e conhecimento resultou em enormes gastos em iniciativas tecnológicas, raramente produzindo aquilo que as organizações pretendiam. O êxito e o fracasso organizacional podem depender freqüentemente do que se sabe sobre esses conceitos (dado, informação e conhecimento): quais deles se têm, de quais se necessita e o que é possível ou não fazer com cada um deles.

Obter dados de diferentes fontes, processá-los para produzir informações relevantes e usar essas informações para o processo de tomada de decisão são procedimentos comuns na área de negócios bem como na vida diária. Os cuidados com a informação, seu valor, sua qualidade e sua segurança são fundamentais para todo tipo de organização.

Um fato bastante comum na atualidade é uma organização possuir sistemas transacionais, os quais têm muitos dados que são úteis apenas para registro de transações, como contabilidade, controle de estoque e produção. Os sistemas de informação apareceram, então, como soluções para ajudar a organização a lidar com seus dados e informações, para que estes pudessem auxiliá-la nos processos de tomada de decisão.

2.4 Sistemas de informação

Laudon e Laudon (1999, p. 4) definem sistema de informação como um conjunto de componentes inter-relacionados que trabalham juntos para coletar, recuperar, processar, armazenar e distribuir informação com a finalidade de facilitar

o planejamento, o controle, a coordenação, a análise e o processo decisório nas organizações.

Os sistemas de informação desempenham três papéis fundamentais em uma organização: apoio às operações do dia-a-dia da organização, apoio ao processo decisório e, através deste, apoio à vantagem competitiva por meio das decisões gerenciais (O'BRIEN, 2003, p. 9).

2.4.1 Evolução

Os sistemas de informação, juntamente com seu conceito e sua importância, evoluíram com o decorrer dos anos. Eles passaram pelas etapas de operacionalização das tarefas rotineiras e de integração entre os diversos sistemas de informação da organização até o suporte ao gerenciamento (LAUDON; LAUDON, 1999).

Segundo O'Brien (2003), os papéis dos sistemas de informação computadorizados têm se expandido no curso do tempo, causando impacto em usuários finais e em gerentes das organizações. O Quadro 1 apresenta um resumo dos papéis dos sistemas de informação ao longo dos anos.

PERÍODO	PAPÉIS DOS SISTEMAS DE INFORMAÇÃO
Processamento de dados: 1950 – 1960	Sistemas de processamento eletrônico de dados: processamento de transações, manutenção de registros e aplicações contábeis tradicionais.
Relatórios administrativos: 1960 – 1970	Sistemas de informação gerencial: relatórios administrativos de informações pré-estipuladas para apoiar a tomada de decisão.
Apoio à decisão: 1970 – 1980	Sistemas de Apoio à Decisão (SAD): apoio interativo e <i>ad hoc</i> ao processo de tomada de decisão gerencial.

<p>Apoio estratégico e ao usuário final: 1980 – 1990</p>	<p>Sistemas de computação do usuário final: apoio direto à computação para produtividade do usuário final e colaboração de grupos de trabalho.</p> <p>Sistemas de suporte a executivos: informações críticas para a alta administração.</p> <p>Sistemas especialistas: conselho especializado baseado no conhecimento para os usuários finais.</p> <p>Sistemas de informação estratégica.</p> <p>Produtos e serviços estratégicos para vantagem competitiva.</p>
<p>Empresa e conexão em rede global: 1990 – 2000</p>	<p>Sistemas de informação interconectados: sistemas direcionados ao usuário final, à empresa e à computação, às comunicações e à colaboração interorganizacionais, incluindo operações e administração globais na internet, intranets, extranets e outras redes empresariais e mundiais.</p>

Quadro 1 - A evolução dos sistemas de informação

Fonte: Adaptado de O'BRIEN, 2003

Segundo Laudon e Laudon (1999), a TI reúne os meios necessários para gerenciar e utilizar a informação, incluindo computação, telecomunicações e sistemas. Dessa forma contempla, portanto, componentes de hardware, software, infra-estrutura, periféricos e gestão de dados de informação.

2.4.2 Classificação

Os sistemas de informação podem ser classificados de maneiras diferentes.

O Quadro 2 classifica-os quanto ao nível organizacional (SELL, 2001):

NÍVEL ORGANIZACIONAL	TIPO DE SISTEMA DE INFORMAÇÃO
Nível operacional	<i>Transaction Processing Systems</i> (TPS) ou Sistemas de Processamentos Transacionais (STP)
Nível de gerência de conhecimento	<p><i>Knowledge Work Systems</i> (KWS) ou Sistemas de Controle de Conhecimento (SCC)</p> <p><i>Office Automation Systems</i> (OAS) ou Sistemas de Automação de Escritório (SAE)</p>

Nível gerencial	<i>Decision Support Systems</i> (DSS) ou Sistemas de Apoio à Decisão (SAD) <i>Management Information Systems</i> (MIS) ou Sistemas de Informação Gerenciais (SIG)
Nível executivo	<i>Executive Support Systems</i> (ESS) ou Sistemas de Informação para Executivos (EIS)

Quadro 2 - Tipos de sistemas de informação

Fonte: Adaptado de SELL, 2001

Os sistemas de nível operacional são responsáveis por monitorar as atividades e as transações elementares da organização. Essas envolvem o registro e o monitoramento de atividades rotineiras necessárias para conduzir o negócio (MOTTA, 2001).

Os sistemas de nível de gerência de conhecimento são utilizados para auxiliar na criação, integração e disseminação de novos conhecimentos e informações (MOTTA, 2001).

Os sistemas de nível gerencial são utilizados para monitorar, controlar e avaliar a utilização de recursos de modo a atingir os objetivos e as metas da organização bem como ajudar no processo de tomada de decisão (MOTTA, 2001). Como exemplo de sistemas de nível gerencial é possível citar os Sistemas de Apoio à Decisão (SAD), os quais surgiram devido à necessidade de se fornecerem informações rapidamente, à competição cada vez maior entre as organizações e à disponibilidade de tecnologias para armazenar e buscar eficientemente as informações.

Os sistemas de nível executivo auxiliam a alta gerência na determinação de estratégias, na definição de objetivos da organização e no planejamento em longo prazo (MOTTA, 2001). Os Sistemas de Informação para Executivos (EIS) são exemplos de sistemas de nível executivo. Segundo Inmon (2002), os EIS foram

feitos sob medida para ajudar executivos a tomarem decisões, permitindo que eles localizem problemas com precisão e detectem tendências de vital importância para a gerência da organização.

2.5 *Business Intelligence e as Tecnologias de Informação*

2.5.1 Definição

Com o intuito de dar suporte à tomada de decisão na busca de vantagem competitiva organizacional, surgiu o conceito de Business Intelligence (BI), a evolução dos Sistemas de Apoio à Decisão e dos Sistemas de Suporte a Executivos.

Business Intelligence é um ambiente em que usuários de negócio recebem dados confiáveis, consistentes, entendíveis e manipuláveis por tempo e de forma simples. Esses dados pretendem auxiliar os usuários em suas análises de modo que eles compreendam o estado do negócio no passado, no presente e também no futuro (DMREVIEW, 2005).

2.5.2 Arquitetura de um sistema de *Business Intelligence*

A Figura 4 apresenta uma arquitetura de um sistema típico de BI, a qual se divide nas seguintes etapas (WG SYSTEMS, 2003a):

- ✍ módulo ETL (Extração, Transformação e Carga de Dados): é a parte responsável pela coleta das informações nas mais diversas fontes, desde sistemas ERP (*Enterprise Resource Planning*) até arquivos de texto ou planilhas de cálculo;

- ☞ *Data Warehouse*: repositório onde ficam concentrados os dados extraídos dos sistemas transacionais. A vantagem de se utilizar um repositório de dados é a possibilidade de se armazenarem informações históricas, dando melhor suporte para as análises posteriores;
- ☞ *front-end*: é a parte visível ao usuário, podendo ser em forma de relatórios padronizados, análises OLAP (*On-line Analytical Processing*) e outras funções, como *Data Mining* (mineração de dados) e projeções de cenários futuros.



Figura 4 - Arquitetura de um sistema de *Business Intelligence*
Fonte: WG SYSTEMS, 2003a

2.6 Data Warehouse

2.6.1 Definições, objetivos e características

Inmon (2002) define *Data Warehouse* (DW) como um conjunto de dados baseado em assuntos, integrado, não-volátil e variável em relação ao tempo, de apoio às decisões gerenciais. Machado (2000) discute essas características:

- ✍ orientado por assunto: o DW agrupa as informações por áreas de interesse da organização, em contraste com os sistemas transacionais, que são orientados por processos;
- ✍ variável em relação ao tempo: os dados de um DW representam os resultados operacionais em um determinado momento, assim sendo, eles não podem ser modificados;
- ✍ não-volátil: um DW possui duas operações básicas, a carga e o acesso aos dados. Esses dados são transformados, purificados, carregados no DW e lá ficam até que se decida que não são mais necessários; e
- ✍ integrado: os dados de um DW possuem um alto nível de integração, o que significa que inconsistências e ambigüidades podem ser eliminadas.

Resumidamente, o DW proporciona uma sólida e concisa integração de dados da organização para a realização de análises gerenciais estratégicas. Ele se preocupa em integrar e consolidar as informações, sumariando, filtrando, limpando e armazenando esses dados num repositório único para que estejam preparados para análise e suporte aos processos decisórios (MACHADO, 2000, p. 13).

Em geral, o DW oferece uma melhoria no desempenho do ambiente operacional e a disponibilidade de informações de melhor qualidade. As decisões a

serem tomadas serão embasadas em fatos, e não em intuições. Poderão ser descobertos novos mercados, novas oportunidades e novos produtos, e isso é um diferencial competitivo para as organizações.

Uma abordagem de utilização de DW é o *Data Mart*, que caracteriza um subconjunto de dados de um DW (MACHADO, 2000), ou seja, uma pequena porção de dados do DW organizada por assunto. Ele serve de fonte para os dados que comporão os bancos de dados individuais e, geralmente, é direcionado para um departamento ou para uma área específica da organização;

2.6.2 Processo de desenvolvimento de um DW

Uma forma de apresentar o processo de desenvolvimento de um DW pode ser resumida de acordo com a Figura 5.

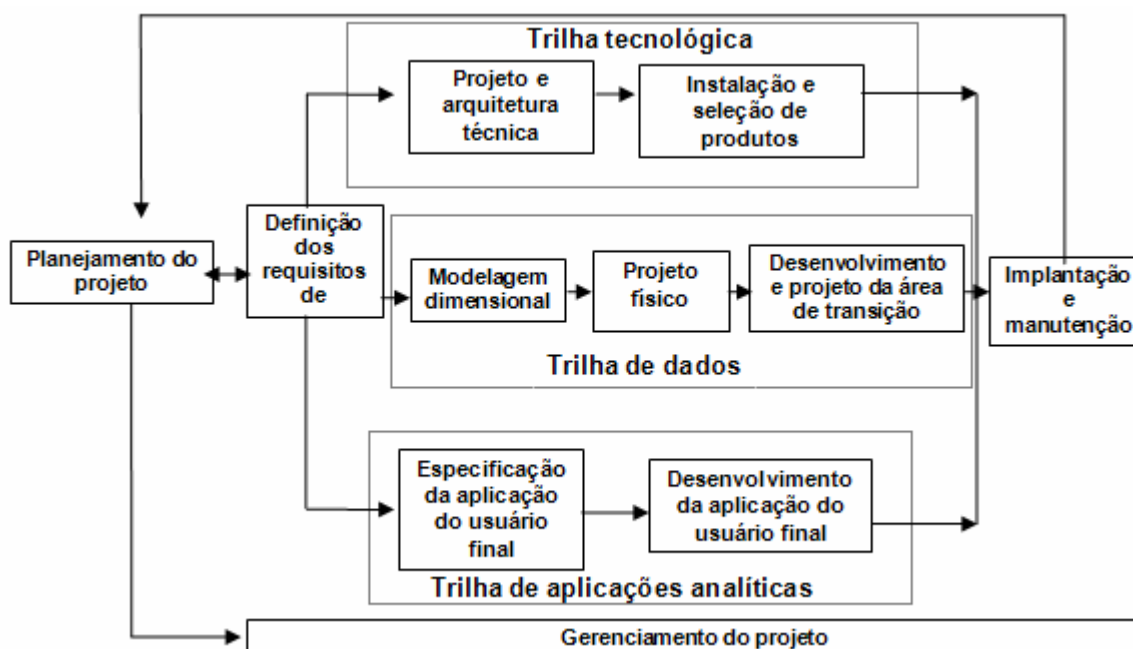


Figura 5 - Ciclo de vida de um projeto de DW

Fonte: KIMBALL; ROSS, 2002

O planejamento (primeira caixa da Figura 5) constitui as seguintes tarefas: avaliação das disponibilidades da organização (investimento, motivação, praticabilidade), definição do escopo e justificativa do projeto, identificação da equipe e suas responsabilidades, e elaboração do cronograma do projeto (KIMBALL; ROSS, 2002).

Na fase de definição dos requisitos de negócios (segunda caixa da Figura 5) faz-se o levantamento das necessidades e das exigências do usuário de negócio em conjunto com a realidade dos dados, com o intuito de se detectarem problemas e oportunidades (KIMBALL; ROSS, 2002). A definição dos requisitos é indispensável para se entender o que acontece na organização e para se definir como deverá ser o sistema. Segundo Thomsen (2002), a documentação dos requisitos e dos processos envolvidos com a implementação pode ser útil para todos os momentos do projeto, inclusive para a comunicação e o consenso entre os implementadores e os usuários. A seta bidirecional entre as fases de planejamento e de definição dos requisitos mostra que há muita interação entre elas.

A partir daí, o diagrama de ciclo de vida de um DW divide-se em três trilhas paralelas: (1) tecnológica, (2) de dados e (3) de aplicações analíticas. A trilha tecnológica prevê o projeto técnico da arquitetura que estabelece a estrutura geral e a infra-estrutura do DW, isto é, a definição de hardware, software, servidores, rede de comunicações, estações de trabalho e, posteriormente, a seleção e a instalação dos produtos específicos. A trilha central, chamada de trilha de dados, inicia-se pela tradução dos requisitos em um modelo dimensional, sendo este transformado em uma estrutura física para depois proceder aos processos de extração, transformação e carga dos dados (ETL). A trilha de aplicações analíticas consiste na especificação e no desenvolvimento das aplicações analíticas dirigidas ao usuário final. Uma vez

desenvolvidas as três trilhas, prevê-se uma fase de treinamento e suporte para distribuição, seguida da manutenção contínua do projeto (KIMBALL; ROSS, 2002).

As trilhas de aplicações analíticas serão abordadas no próximo capítulo, pois são necessárias para a apresentação da proposta deste trabalho.

2.6.2.1 Modelo dimensional

A modelagem dimensional é uma técnica de concepção e visualização de um modelo de dados que descreve aspectos comuns de negócios, sendo empregada especialmente para reestruturar dados e apresentá-los em visões que suportem análises gerenciais (MACHADO, 2000).

O principal tipo de modelo dimensional é o chamado modelo estrela ou *star schema*, composto de uma tabela dominante no centro chamada de fato e de um conjunto de tabelas menores chamadas de dimensões (KIMBALL et al., 1998). As tabelas de fato contêm várias chaves, enquanto as de dimensão possuem uma única chave primária que faz a ligação com a tabela de fatos. Um exemplo típico de modelo estrela pode ser visualizado na Figura 6.

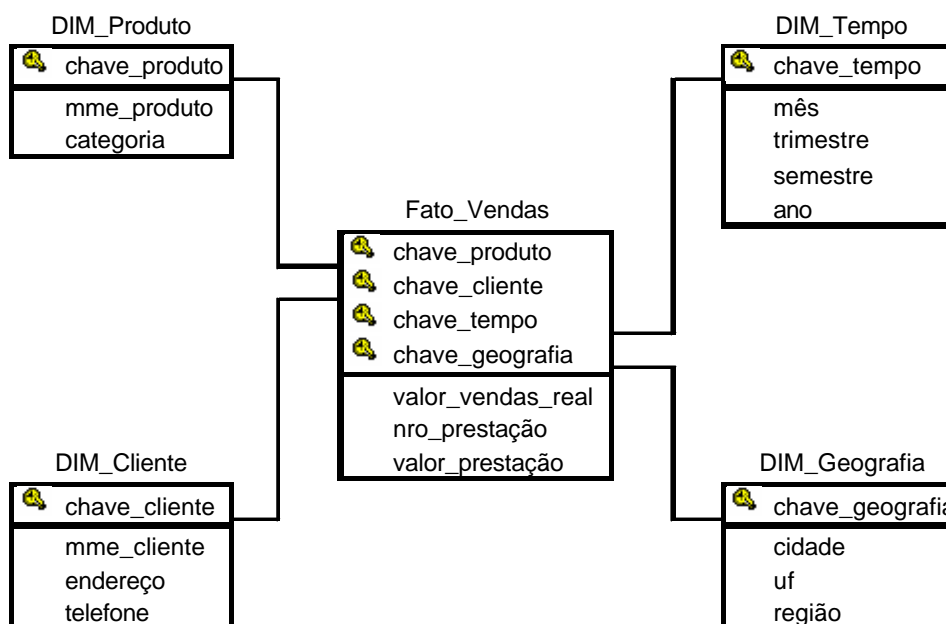


Figura 6 - Um modelo estrela

Fonte: Adaptado de KIMBALL; ROSS, 2002

Segundo Machado (2000), fato é uma coleção de dados numéricos do negócio a serem analisados, enquanto uma dimensão se refere ao contexto em que um determinado fato ocorreu, como períodos de tempo, produtos, mercados, clientes e fornecedores, ou seja, elementos que possam descrever o contexto de um determinado fato. A medida, também um componente do modelo estrela, é um atributo que quantifica um determinado fato, representando o desempenho de um indicador em relação às dimensões que participam do fato (MACHADO, 2000). O contexto de uma medida é estabelecido em função dessas dimensões.

Segundo Kimball et al. (1998), a forma de organização dos dados do modelo dimensional propicia flexibilidade para eventuais ajustes que se façam necessários no modelo, como, por exemplo, adicionar novos fatos à tabela de fatos, desde que os dados correspondam ao mesmo nível de detalhe, adicionar novas dimensões, adicionar novos atributos às dimensões existentes e redefinir o nível de detalhe dos dados.

Outro tipo de estrutura bastante comum é o modelo de dados “flocos de neve” ou *snow flake*. Machado (2000) define esse modelo como uma extensão do modelo estrela, sendo o resultado da decomposição de uma ou mais dimensões que possuem hierarquias entre seus membros.

2.6.2.2 Projeto físico

O modelo dimensional deve ser traduzido para um modelo (projeto) físico. O modelo físico consiste na especificação para o banco de dados físico, por exemplo, nomes das colunas, tipos de dados e chaves primárias. Nesta fase, criam-se também os índices e os agregados, alternativas efetivas para o desempenho do sistema (KIMBALL; ROSS, 2002).

Segundo Kimball e Ross (2002), agregados são linhas físicas em um banco de dados, quase sempre criados a partir da soma de registros organizados com o intuito de melhorar o desempenho de consultas.

São duas as abordagens de utilização de agregados: (1) definir novas tabelas de fatos ou (2) definir campos nível, sendo que a primeira torna a manutenção, a carga e a utilização dos dados mais simples (KIMBALL et al., 1998).

Para melhorar o desempenho do modelo, podem ser construídos diversos agregados que representem os agrupamentos mais comuns dentro das dimensões do DW. Porém, esse aumento do desempenho significa aumento do consumo de espaço de armazenamento (KIMBALL et al., 1998).

2.6.2.3 Projeto e desenvolvimento da área de *data staging*

O objetivo desta fase é fazer a integração de informações de fontes múltiplas e complexas. Esta etapa é uma das mais críticas de um projeto de DW, pois uma informação carregada erroneamente trará conseqüências imprevisíveis nas fases posteriores. Antes de os dados serem armazenados no DW de destino, eles precisam ser manipulados e freqüentemente armazenados em áreas intermediárias, chamadas de áreas de *data staging* (KIMBALL; ROSS, 2002).

Na área de *data staging* é realizada a etapa de ETL, que está dividida em três passos: (1) extração, (2) transformação e (3) carga dos dados. Extração é o primeiro passo para dispor os dados no DW. Primeiramente são definidas as origens das fontes de dados e, em seguida, são realizados a leitura, o entendimento e a extração dos dados (KIMBALL; ROSS, 2002). Essas origens podem ser muitas e de diferentes formatos, como, por exemplo, de sistemas OLTP, ERP, planilhas, arquivos-texto, etc. Definidas as fontes, deve-se transformar e limpar esses dados. Tal etapa é necessária porque os dados normalmente vêm inconsistentes, com lixo, e advêm de fontes diversas, desconhecidas e antigas. Portanto, é necessário padronizar os diferentes formatos para que o usuário não veja informações iguais apresentadas de forma diferente ao consultar o DW. Uma vez realizadas a limpeza e a transformação dos dados, estes devem ser carregados na base de dados.

2.7 Considerações finais

Os sistemas de informação têm revelado uma importância crescente nos dias atuais para qualquer tipo de organização. A necessidade de realizar as atividades de

forma eficaz, com relação ao tempo e aos objetivos, obriga a organização a desenvolver novos meios de executar tarefas antigas ou mesmo de inovar e reinventar novas tarefas para se manter em atividade no mercado.

As TIs têm sido alguns dos principais componentes dos sistemas de informação das organizações. Na atualidade, muitos investimentos estão sendo direcionados para soluções de *Business Intelligence* e *Data Warehouse*, pois elas contemplam diretamente as relações com o cliente e a manutenção de informações precisas e seguras para dar suporte à tomada de decisão, provocando alteração e modernização nos processos de trabalho bem como melhorias nos resultados gerenciais.

O DW é abordado pela sua importância como repositório de dados no processo decisório, transformando dados em informações e informações em conhecimento necessário e útil para as organizações. Essa importância deve-se à habilidade do DW em armazenar dados históricos, integrados e orientados por assunto de interesse das organizações, os quais podem revelar informações importantes sobre o negócio. O DW não deixa de ser um sistema de informação. Ambos são complementares na tarefa de subsidiar a tomada de decisão nas organizações.

Para que haja mais agilidade no fluxo informacional, as soluções ligadas aos sistemas de informação que fazem parte da organização devem ser integradas entre si. O mesmo se aplica à solução OLAP, a qual, como já visto neste capítulo, também pode ser um componente da arquitetura de BI. Os conceitos referentes a essa solução serão abordados no próximo capítulo.

3 Área de apresentação

3.1 Introdução

Este capítulo aborda conceitos relacionados a sistemas OLAP (*On-line Analytical Processing*), soluções utilizadas no apoio ao usuário final para a tomada de decisão estratégica. Entre os principais conceitos OLAP apresentados, é possível citar funcionalidades (operações) essenciais, características de processamento e arquiteturas de armazenamento.

Como o modelo proposto neste trabalho é aplicado a um OLAP para a *Web* (WOLAP) – um tipo de arquitetura de armazenamento OLAP definido adiante – também são abordados conceitos de usabilidade na *Web*, pois, como sistemas OLAP constituem tecnologias voltadas diretamente para o usuário, a usabilidade pode se tornar importante no auxílio ao desenvolvimento de interfaces.

3.2 Conceitos OLAP

Segundo Poe, Klauer e Brobst (1998), sistema transacional, atualmente chamado de *On-line Transaction Processing* (OLTP), é o processamento realizado por sistemas computacionais que têm a finalidade de capturar as transações dos negócios do empreendimento e dar suporte às atividades diárias de uma organização, sendo seus bancos de dados ditos de leitura e escrita, pois oferecem as atividades básicas de inserção, atualização, consulta e deleção de dados.

Para acessar e manipular os dados em bancos de dados OLTP, utiliza-se freqüentemente uma linguagem denominada *Structure Query Language* (SQL). No

entanto, essa linguagem nem sempre é conhecida pelos executivos e analistas de negócios. Outro problema referente às transações OLTP é que os dados são basicamente voláteis, ou seja, são manipulados em tempo real, não havendo, assim, o armazenamento dos dados históricos (VIEIRA, 2001).

Os sistemas OLTP levaram as organizações a acumular um grande volume de dados ao longo dos anos, sem que se soubesse a sua utilidade. Sendo assim, surgiram várias ferramentas de TIs, e, com isso, foi possível demonstrar que esses dados escondem informações valiosas para a organização.

Entre as TIs que surgiram, é possível citar o DW, visto no capítulo anterior, que permitiu armazenar e organizar grandes volumes de dados. Para a recuperação e a apresentação desses dados aos usuários de negócio, de forma que auxiliassem os processos decisórios, surgiu então a abordagem OLAP.

Segundo Kimball et al. (1998), OLAP constitui-se das atividades gerais e específicas de consulta e apresentação de dados numéricos e textuais provenientes do DW. Os sistemas OLAP utilizam a visão multidimensional dos dados, sendo ela natural, fácil e intuitiva, permitindo a visualização desses dados em diferentes perspectivas de negócios de uma organização.

Basicamente, as informações são representadas de forma que se permita seu livre manuseio. Essa representação deve oferecer informações estratégicas para o analista de negócio sem que ele tenha que se preocupar com detalhes tecnológicos do sistema. Os sistemas OLAP ajudam analistas a sintetizarem informações sobre a organização, através de comparações, visões personalizadas, análises históricas e projeções de dados, oferecendo respostas rápidas e consistentes às consultas iterativas, independentemente do tamanho e da complexidade do banco de dados.

Na atualidade, são várias as ferramentas OLAP disponíveis no mercado, entre elas: Maestro (Hyper Consultoria em Informática), Hyperion Wired for OLAP (Hyperion Solutions), DSS MicroStrategy (MicroStrategy) e BusinessObjects (BusinessObjects).

3.3 Funcionalidades de uma ferramenta OLAP

Numa seção típica OLAP, um usuário deseja obter informações (medidas de um fato) de várias perspectivas (dimensões) e em diferentes níveis de detalhe. Por exemplo, o valor de vendas (medidas de um fato) pode ser analisado por filial, por produto e por mês (dimensões). Um OLAP pode oferecer diversas funcionalidades para visualizar as informações de diferentes formas, sendo cada uma delas padronizada com um nome. Entre as principais funcionalidades disponibilizadas pelas ferramentas OLAP, destacam-se as apresentadas no Quadro 3 a seguir.

FUNCIONALIDADE	DESCRIÇÃO	EXEMPLO
<i>Drill Down</i>	Aumentar o nível de detalhe da informação e, conseqüentemente, diminuir o grau de granularidade ¹ (PARENTE, 2000).	Uma análise de vendas por Estado é alterada para uma análise de vendas das cidades de um Estado específico.
<i>Drill Up</i>	Diminuir o nível de detalhe e aumentar o grau de granularidade (PARENTE, 2000).	Uma análise de vendas é alterada de uma cidade para seu Estado correspondente.

¹ Granularidade diz respeito ao nível de detalhe ou de resumo com o qual serão armazenados os dados no DW. Quanto mais detalhe, mais baixo o nível de granularidade, quanto menos detalhe, mais alto esse nível será (INMON, 2002)

<i>Slice</i>	Cortar o cubo, mas manter a mesma perspectiva de visualização dos dados. Funciona como um filtro restringindo uma dimensão a somente algum ou alguns de seus valores (PARRINI, 2002).	Em uma dimensão tempo de um modelo, é selecionado somente o ano de 2000.
<i>Dice</i>	Mudar a perspectiva da visão multidimensional como se o cubo fosse girado. Permite descobrir comportamentos e tendências entre os valores das medidas analisadas, em diversas perspectivas (PARRINI, 2002).	A análise é alterada de região (linha) por ano (coluna) para ano (linha) por região (coluna).
<i>Drill Across</i>	É a ligação de duas ou mais tabelas de fato em uma mesma granularidade, ou seja, tabelas com o mesmo grupo de medidas e dimensões (KIMBALL, 1996).	Passar a visualizar a medida “total de vendas” de uma tabela de fatos chamada “Fato_Vendas” para outra “Fato_Pedido_Vendas” (através de uma tabela de dimensão Dim_Tempo).
<i>Drill Through</i>	Ocorre quando o usuário passa de uma informação contida em uma dimensão para uma outra (PARENTE, 2000).	O usuário está realizando uma análise na dimensão de tempo e no próximo passo analisa a informação por região (PARENTE, 2000).
<i>Drill Out</i>	É o detalhamento para informações externas, como fotos, som, arquivos-texto, tabelas.	O usuário seleciona um cliente ou fornecedor para visualizar um resumo sobre seus dados.
<i>Sort</i>	Tem a função de ordenar a informação (PARENTE, 2000), podendo ser aplicada a qualquer tipo de informação, não somente a valores numéricos.	Ordenar as instituições em ordem alfabética.
<i>Ranking</i>	Permite agrupar resultados numéricos por ordem de tamanho (PARENTE, 2000), refletindo somente na apresentação do resultado e não no resultado em si.	Ordenar a relação de filiais de acordo com os maiores volumes de vendas.
<i>Paging</i>	Apresentação dos resultados de uma consulta em várias páginas, permitindo a navegação do usuário.	Permitir um máximo de 10 resultados por página.
<i>Filtering</i>	Apresentação de consultas com restrições sobre atributos ou fatos, ou seja, restringe um fato a somente algum ou alguns de seus valores.	Em um fato Vendas, apenas os valores das vendas a prazo são apresentados.

<i>Alerts</i>	Utilizados para indicar situações de destaque em elementos dos relatórios, baseados em condições envolvendo objetos e variáveis (PARENTE, 2000).	Definir que os valores das vendas mensais inferiores a R\$ 50.000,00 devem aparecer com destaque em vermelho.
<i>Break</i>	Permite separar o resultado de uma análise em grupos de informações (PARENTE, 2000), possibilitando também a subtotalização de valores para cada grupo.	O usuário tem a necessidade de visualizar a informação por cidades, então ele solicita um <i>break</i> . O relatório será automaticamente agrupado por cidades, somando os valores mensuráveis por cidades (PARENTE, 2000).

Quadro 3 - Funcionalidades OLAP

Fonte: Adaptado de PARENTE, 2000 e PARRINI, 2002

Essas funcionalidades podem ser combinadas entre si, ou seja, pode-se realizar um *slice* e um *dice* ao mesmo tempo que um *drill down* ou um *drill up*.

Nem todas as ferramentas OLAP que estão no mercado suportam todas as funcionalidades mencionadas. A necessidade de cada funcionalidade OLAP depende dos objetivos da organização e de seus usuários.

3.4 Características do processamento OLAP

Codd et al. (1993) usaram o termo OLAP pela primeira vez através de um artigo intitulado em “*Providing OLAP to User-Analysts: An IT Mandate*” ou “Fornecendo OLAP para usuários-analistas: um mandato da TI”.

Tal artigo apresenta uma lista de 12 regras a serem utilizadas para a construção de sistemas OLAP. Kline, Gould e Zanevsky (2003) descreve essas regras, como se segue.

1. As informações devem ser apresentadas em tabelas multidimensionais.

2. Os dados devem ser acessados através das tabelas, chaves primárias e dos atributos.
3. Valores nulos devem ser apresentados como informação não disponível, e não como informações vazias.
4. Os metadados são armazenados e representados da mesma forma que os próprios dados.
5. Uma linguagem deve ser capaz de definir dados, *views*, acessos (autorizações), restrições de integridade, transações e manipulação de dados.
6. As *views* devem mostrar as atualizações de suas tabelas.
7. Um simples usuário operando deve ser capaz de realizar operações de consulta, atualização, inserção e deleção de dados.
8. Os *batches* e as operações dos usuários finais são separados da organização física dos arquivos da base de dados e dos métodos de acesso a esses arquivos.
9. Os *batches* e as operações dos usuários finais podem modificar o esquema de banco de dados sem ter que recriá-lo ou as aplicações construídas nele.
10. As restrições de integridade devem estar armazenadas e disponíveis no metadado, não em uma aplicação.
11. O fato de uma base de dados estar centralizada em uma máquina ou distribuída em várias não deve influir na manipulação dos dados do sistema.
12. Qualquer linha de processamento do sistema deve obedecer às mesmas restrições de integridade das operações de processamento.

As regras de Codd aqui descritas não devem ser vistas como obrigatórias no desenvolvimento de um sistema OLAP eficiente, principalmente porque foram escritas há muitos anos. Em princípio, a quanto mais itens o sistema atender, melhor será sua qualidade. Porém, o item essencial em um sistema OLAP é que suas características atendam às necessidades da organização e de seus usuários de negócio.

3.5 Arquiteturas

Muitos termos e nomenclaturas sobre arquiteturas OLAP estão surgindo e se difundindo. A nomenclatura referente à arquitetura de um OLAP está diretamente ligada às opções de local de armazenamento e processamento do sistema. As arquiteturas OLAP podem ser divididas basicamente em: ROLAP (*Relational On-Line Analytical Processing*), MOLAP (*Multidimensional On-Line Analytical Processing*), DOLAP (*Desktop On-Line Analytical Processing*) e HOLAP (*Hybrid On-Line Analytical Processing*).

Cada uma das arquiteturas citadas tem uma função específica, apresentando vantagens e desvantagens, enquanto as abordagens MOLAP e ROLAP são as mais utilizadas. O Quadro 4 resume algumas diferenças entre as duas:

BENEFÍCIOS	MOLAP	ROLAP
Visão multidimensional	X	X
Excelente desempenho	X	
Flexibilidade analítica	X	
Acesso a dados em tempo real		X

Grande capacidade de dados		X
Alavanca o DW		X
Fácil desenvolvimento	X	
Baixa manutenção de estrutura		X
Baixa manutenção de agregados	X	

Quadro 4 - Benefícios das arquiteturas MOLAP e ROLAP

Fonte: Traduzido e adaptado de ALALOUF, 1997

No momento da escolha entre as arquiteturas, o mais importante é entender os negócios da organização e selecionar qual melhor atende ao volume de dados e às necessidades de análise que ela possui. A arquitetura selecionada deve oferecer aos usuários uma boa flexibilidade, uma redução de complexidade e um bom desempenho nas análises.

3.6 Usabilidade na Web

No início do desenvolvimento dos sistemas de informação, os computadores eram utilizados profissionalmente pelos próprios desenvolvedores e por alguns usuários externos qualificados e treinados. Atividades de usabilidade nas TIs passaram a ser essenciais a partir do momento em que o público-alvo ampliou e passou a exigir programas amigáveis, em que treinamentos e manuais não fossem tão necessários.

Usabilidade é definida pela ISO/DIS 9241-11 (1995) como a capacidade que um sistema interativo oferece a usuários específicos com o intuito de obter objetivos específicos com efetividade, eficiência e satisfação em um contexto específico. Efetividade é a precisão e integridade com que os usuários obtêm seus objetivos;

eficiência são os recursos expendidos com relação à efetividade; e satisfação é o conforto e a acessibilidade de utilização do produto.

Segundo Cybis (2003), a dificuldade no desenvolvimento de interfaces com usabilidade se deve ao fato de elas constituírem fundamentalmente sistemas abertos, probabilísticos, não determinísticos, sujeitos às influências do ambiente e às interpretações dos usuários. Assim, a experiência da interação entre o usuário e o computador é individual, visto que cada pessoa é única em sua bagagem de conhecimento e experiência.

A teoria de usabilidade visa identificar os problemas de usabilidade em interfaces entre usuários e computadores, através de diagnóstico ou observação, e contribuir para a sua eliminação. Entre os objetivos a serem atingidos, têm-se (CYBIS, 2003):

- ✍ constatar, observar e registrar problemas efetivos de usabilidade durante a interação;
- ✍ calcular métricas objetivas para eficácia, eficiência e produtividade do usuário na interação com o sistema;
- ✍ diagnosticar as características do projeto que provavelmente atrapalhem a interação por estarem em desconformidade com padrões implícitos e explícitos de usabilidade;
- ✍ prever dificuldades de aprendizado na operação do sistema;
- ✍ prever os tempos de execução de tarefas informatizadas;
- ✍ conhecer a opinião do usuário em relação ao sistema;
- ✍ sugerir as ações de reprojeto mais evidentes diante dos problemas de interação efetivos ou diagnosticados.

3.6.1 Critérios de avaliação da usabilidade

3.6.1.1 Checklists

Checklist fundamenta-se basicamente na criação de um conjunto de listas, envolvendo princípios ergonômicos, sendo elas indicadores de qualidade para o critério de usabilidade. Muitos *checklists* são baseados em recomendações de manuais de desenvolvimento de interfaces gráficas, nas normas ISO 9241 e em guias para desenvolvimento de softwares.

Segundo Cybis (2003), a avaliação realizada através de *checklists* apresenta as seguintes potencialidades:

- ✍ possibilidade de ser realizada por projetistas, não exigindo especialistas em interfaces entre o usuário e o computador;
- ✍ sistematização da avaliação que garante resultados mais estáveis, pois as questões constantes do *checklist* sempre serão efetivamente verificadas;
- ✍ facilidade na identificação de problemas de usabilidade, devido à especificidade das questões do *checklist*;
- ✍ aumento da eficácia de uma avaliação, devido à redução da subjetividade normalmente associada a processos de avaliação; e
- ✍ redução de custo da avaliação, pois é um método de rápida aplicação.

Este método de avaliação permite a detecção de dois tipos de problema na utilização do software: (1) falhas no projeto da interface, envolvendo o acesso às funcionalidades, e (2) problemas na definição das funcionalidades, envolvendo a compatibilidade das funcionalidades com as tarefas-alvo do software.

3.6.1.2 Critérios de Jakob Nielsen

Jakob Nielsen, um dos pais da usabilidade na *Web*, descreve, em diferentes artigos, uma seleção de problemas e recomendações mais comuns que ainda são extremamente relevantes na usabilidade dos dias de hoje. O Quadro 5 resume seu primeiro artigo (NIELSEN, 1996):

CRITÉRIO	DESCRIÇÃO
<i>Frames</i>	<i>Frames</i> apresentam pequenos problemas na impressão da página e impedem usuários de recomendar uma URL para outros usuários.
Uso desnecessário de tecnologia	O público principal está interessado no conteúdo do site e na sua habilidade de oferecer bons serviços. O uso da tecnologia mais atual deve ser evitado, caso atrapalhe a consulta do usuário.
Animações	Imagens em movimento sobrecarregam a visão periférica do olho humano e tiram a concentração do conteúdo de texto que está sendo apresentado.
URLs complexas	Sempre que possível URLs devem conter conteúdo legível por usuários e nomes de arquivos que representem o seu conteúdo de informação. URLs longas causam problemas quando usuários tentam recomendá-las a outros usuários.
Páginas órfãs	Cada página deve ter um link para a página principal assim como alguma indicação de onde cada página se encontra dentro da hierarquia de informações do site.
Páginas longas	As opções visíveis, ou seja, apresentadas na tela, dominam. Os usuários nem sempre utilizam a barra de rolagem para visualizar o restante da página.
Suporte à navegação	Um link do logo do site no topo esquerdo de todas as páginas (que chamem a página principal) é bastante eficaz no suporte à navegação do usuário.
Cores dos <i>links</i>	Recomenda-se não modificar as cores padrão dos <i>links</i> para um melhor controle dos usuários: não visitados azuis e visitados vermelhos ou roxos.
Informação desatualizada	Manter a informação do site desatualizada pode causar a perda de credibilidade por parte dos usuários, principalmente quando eles possuem tantas outras opções de sites a recorrer.
Tempo de <i>download</i>	Deve-se, sempre, reduzir o tempo de <i>download</i> das páginas. Os usuários <i>Web</i> podem suportar um tempo máximo de 15 segundos antes que percam interesse.

Quadro 5 - Recomendações de usabilidade na *Web*

Fonte: Adaptado de NIELSEN, 1996

Essa seleção de recomendações de usabilidade definidas por Nielsen em 1996 levanta problemas comuns na usabilidade da *Web* ainda encontrados em muitos sites atuais. Infelizmente novas tecnologias e aplicações para a *Web* introduziram uma nova classe de problemas freqüentes, os quais podem ser resumidos no Quadro 6 (NIELSEN, 1999a, 1999b):

CRITÉRIO	DESCRIÇÃO
Botão “voltar”	O botão “voltar” é um guia para o usuário <i>Web</i> , sua utilização não deve ser desabilitada ou comprometida.
Nova janela de <i>browser</i>	Abrir novas janelas de <i>browser</i> polui a tela do usuário, além de desabilitar o botão “voltar”.
Uso de GUI <i>widgets</i> não padronizados	Consistência é um dos princípios mais importantes da usabilidade. As piores violações de consistência são na utilização de GUI <i>widgets</i> . Por exemplo, ao se escolher uma opção em um <i>radio button</i> , nada deve acontecer até que o usuário confirme sua escolha.
Biografias	Personalidade e ponto de vista são muito valorizados pelos usuários. É muito importante incluir os nomes dos autores caso os usuários precisem procurar trabalhos e informações desses autores.
Informação antiga	Informação antiga pode ser útil para leitores, mesmo quando a nova é mais valiosa. Manter informações antigas aumenta cerca de 10% no custo de um site, mas pode diminuir sua utilidade em 50% caso não existam.
Novas URLs	As URLs estão circulando entre os usuários da <i>Web</i> , sua modificação atrapalha a divulgação.
Contexto de títulos	Títulos, manchetes e cabeçalhos devem indicar o que o usuário irá encontrar, ajudando-os a navegar na <i>Web</i> .
Tempo de resposta do servidor	Recomenda-se investir em servidores rápidos e contratar um especialista para revisar a arquitetura do sistema e a qualidade do código para otimizar tempos de resposta.
Design com aparência de propaganda	É sempre bom evitar design que pareça ser anúncio ou propaganda, pois os usuários estão acostumados a fechá-los e ignorá-los.

Quadro 6 - Recomendações mais atuais de usabilidade na *Web*

Fonte: Adaptado de NIELSEN, 1999a, 1999b

Os critérios descritos neste quadro são válidos, mas devem ser bem analisados e adaptados de acordo com cada caso. Dependendo do site da *Web*, talvez seja necessário fugir à regra em algum critério. Por exemplo, se o objetivo do site for fazer um treinamento via *Web* que no final ofereça uma prova para testar os conhecimentos do usuário, desabilitar o botão “voltar” será uma necessidade, não um empecilho.

3.7 WOLAP

Além da diversidade de arquiteturas OLAP existe também uma abordagem de solução OLAP para acesso via *Web*, denominada *Web On-line Analytical Processing* (WOLAP).

Ahmad, Ahmad e Yahaya (2002) apresentam algumas razões da conveniência da utilização da abordagem WOLAP, são elas:

- ✍ o OLAP pode ser acessado em qualquer *browser* e plataforma;
- ✍ existe facilidade de manutenção, a qual é realizada apenas no servidor;
- ✍ as técnicas e habilidades utilizadas para navegar e selecionar a informação são semelhantes a outras aplicações da *Web*;
- ✍ a interface do *browser*, com *links* e botões, é imediatamente reconhecida pelos usuários da *Web*; e
- ✍ com técnicas adequadas de segurança, a internet pode ser utilizada como uma *Wide Area Network* (WAN) de baixo custo para o suporte à decisão.

Antes de utilizar essa arquitetura, é importante analisar os critérios mais atualizados sobre usabilidade na *Web*. Como essa é a abordagem utilizada no presente trabalho, alguns desses critérios serão tratados a seguir.

3.7.1 WOLAP e usabilidade

Segundo descreve Thomsen (2002), as etapas de implementação de um OLAP são:

- ✍ requisitos do usuário: análise dos problemas e restrições lógicas e físicas, e elaboração de um documento que poderá ser utilizado em todo o processo de desenvolvimento do OLAP;
- ✍ projeto da solução: definição de um modelo multidimensional que atenda aos requisitos do usuário dentro das restrições indicadas e com o objetivo de solucionar os problemas conhecidos; e
- ✍ agregações e análises mais complexas.

Essas etapas são similares ao *ciclo de vida de um projeto de DW* proposto por Kimball e Ross (2002) e apresentado neste trabalho (Figura 5 do capítulo 2).

Não foram encontradas na literatura as etapas para a construção de um WOLAP ou os conceitos de usabilidade para WOLAP. No entanto, no próximo capítulo este trabalho apresenta, como parte da sua metodologia, uma proposta de navegabilidade em um WOLAP. Esta proposta tem como preocupação o cidadão comum, além de levar em consideração os princípios de usabilidade para a construção de uma aplicação WOLAP.

3.8 Considerações finais

Os repositórios de dados (DW) são, geralmente, acompanhados de ferramentas analíticas utilizadas para a apresentação das informações neles armazenadas. O objetivo da ferramenta OLAP é cumprir com a função de recuperar

os dados armazenados no DW e apresentá-los aos usuários de negócio com o intuito de auxiliá-los nos processos decisórios.

A tecnologia OLAP é uma resposta do mercado para atender à demanda de análise multidimensional de dados corporativos, cálculos complexos e análise de tendência. Com isso, possibilita maior independência dos usuários na confecção de relatórios, envolvendo análises em múltiplas dimensões e com tempo de resposta aceitável.

O trabalho proposto faz uma combinação das tecnologias DW, OLAP e *Web*. O DW tem o objetivo de armazenar e prover a integração dos dados; OLAP, de extrair e apresentar as informações aos usuários de forma a oferecer os mecanismos que realmente obtenham os benefícios do DW; já o objetivo da *Web* é distribuir e compartilhar informações. A utilização adequada e em conjunto dessas tecnologias oferece meios para medir e gerenciar os negócios eficientemente.

Os conceitos de usabilidade aqui apresentados servirão para direcionar o desenvolvimento de um sistema mais próximo do domínio dos possíveis usuários, oferecendo-lhes melhor produtividade e eficiência em suas navegações. Este trabalho pretende propor uma solução alternativa para as análises multidimensionais que apresenta como vantagem os benefícios que a *Web* tem a oferecer, como, por exemplo, a facilidade na distribuição da informação.

Até este ponto foram definidos os conceitos utilizados como base para a solução proposta. Os próximos capítulos apresentam a metodologia elaborada e sua aplicação em um caso real.

4 METODOLOGIA PARA O DESENVOLVIMENTO DE INSTRUMENTOS DE ANÁLISE MULTIDIMENSIONAL

4.1 Introdução

Neste capítulo, serão respondidas as problemáticas levantadas no presente trabalho. Para que sistemas OLAP baseados em DW traduzam os requisitos de análise dos usuários em processos de recuperação e apresentação de informações na *Web*, e para sistematizar a apresentação de análises multidimensionais mais próximas da visão do domínio do problema que o usuário possui, este trabalho propõe uma metodologia para o desenvolvimento e a análise de sistemas multidimensionais.

4.2 Metodologia proposta

Nos capítulos anteriores foi evidenciada a importância da utilização dos repositórios de dados chamados de DW devido à sua habilidade em armazenar dados históricos, integrados e orientados por assunto de interesse das organizações, pois esses dados podem revelar informações importantes sobre o negócio. Por outro lado, os repositórios de dados devem estar acompanhados de ferramentas analíticas adequadas para a apresentação das informações ali armazenadas. O objetivo da ferramenta OLAP é cumprir essa função, recuperando e apresentando os dados armazenados no DW para os usuários de negócio com o intuito de auxiliá-los nos processos decisórios.

No entanto, um dos princípios sustentados neste trabalho é que as ferramentas OLAP disponíveis no mercado são, na sua maioria, complexas e de

difícil manipulação e navegabilidade por parte dos analistas de negócio e usuários em geral, como o cidadão. A dificuldade encontrada pelos analistas de negócio na utilização das ferramentas analíticas tem origem no processo de levantamento de requisitos para o projeto de DW. Como foi descrito no capítulo anterior, é na fase de definição dos requisitos de negócio do processo de construção de um DW que se faz o levantamento das necessidades e das exigências do usuário de negócio em conjunto com a realidade dos dados (KIMBALL; ROSS, 2002). O objetivo de apurar os requisitos de um DW é prover uma especificação compreensível dos principais aspectos técnicos e permitir que a implementação do sistema se complete com sucesso. O conjunto de requisitos levantados com os usuários expressa (TISSOT, 2004):

- ✍ os objetivos e benefícios da organização com relação ao projeto de DW (requisitos de negócios);
- ✍ as tarefas que os usuários estarão aptos a realizar com o auxílio do DW (requisitos de usuários);
- ✍ as funcionalidades que o sistema de *front-end* deve ter para que os usuários cumpram suas tarefas (requisitos funcionais); e
- ✍ as informações que serão necessárias para serem apresentadas e disponibilizadas pelo DW (requisitos de informação).

Assim, uma nomenclatura conceitual que faça a correspondência dos conceitos técnicos com a linguagem dos usuários é de grande valia para auxiliar na identificação dos requisitos citados (processos 1 e 2 da Figura 7) e para definir, junto aos futuros usuários, as funcionalidades, a interface e as formas de utilização e navegabilidade de um *front-end* para sistemas multidimensionais (processos 3 e 4 da Figura 7).

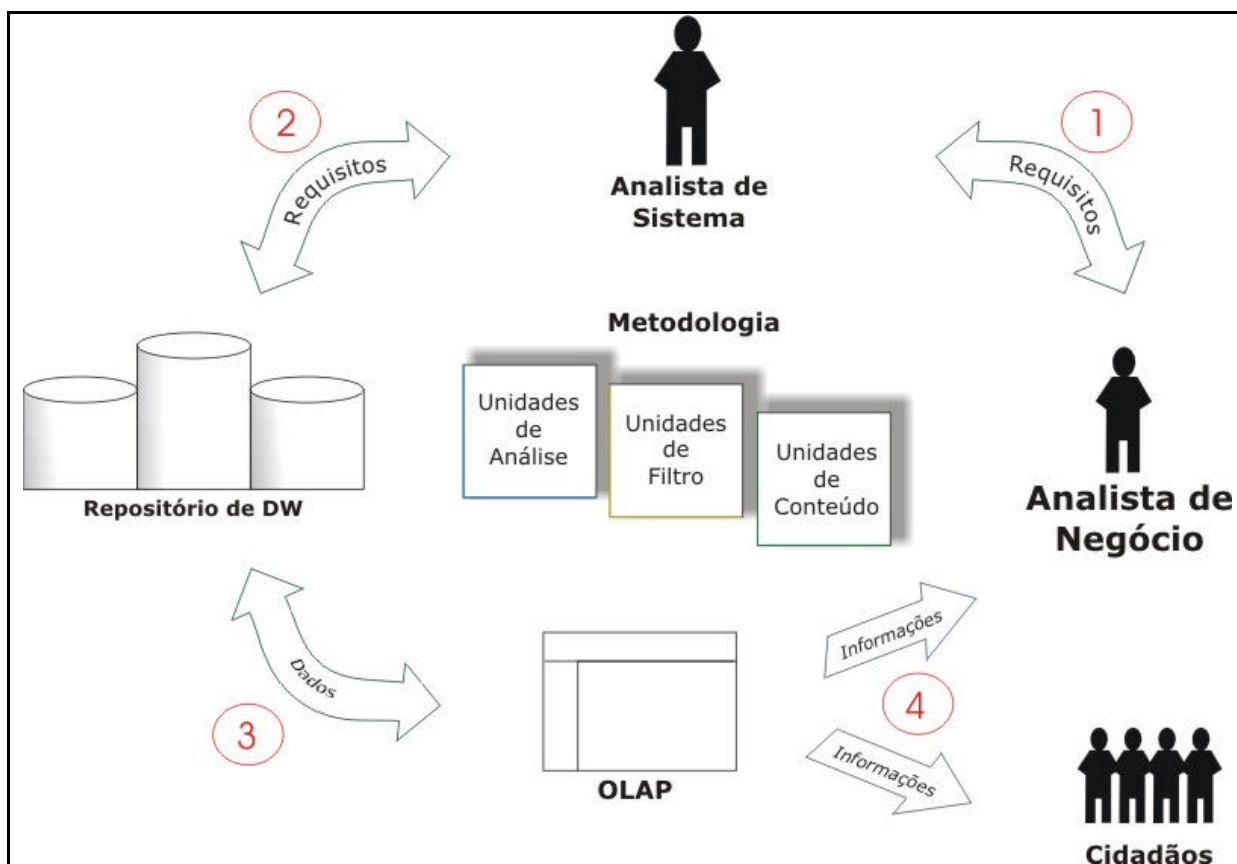


Figura 7 - Contexto da metodologia proposta em sistemas DW/OLAP
Fonte: Da autora

Para que a nomenclatura atinja os objetivos desejados, deve ser expressa em uma linguagem não técnica, própria, adequada e de fácil assimilação por parte de analistas de negócio e usuários do DW.

Com a nomenclatura conceitual definida, podem ser elaboradas as etapas da metodologia proposta, as quais fazem uso da nomenclatura e levam em consideração as principais funcionalidades dos sistemas WOLAP e os conceitos de usabilidade definidos no capítulo anterior.

4.2.1 Nomenclatura conceitual

Na construção de repositórios de dados do tipo DW, as análises de requisitos levantadas junto aos usuários devem refletir tanto a visão dos dados quanto a visão das aplicações analíticas que apresentam os dados. Dessa forma, a identificação dos requisitos modelados no repositório de dados e traduzidos em cubos de informação pelas técnicas OLAP será guiada por uma nomenclatura. Essa nomenclatura permite que os técnicos responsáveis pelo levantamento dos requisitos e pela modelagem dos dados correspondam às expectativas dos analistas de negócio. Enquanto isso, estes poderão utilizar as potencialidades do repositório de dados através de uma ferramenta intuitiva e de fácil entendimento que traduz os requisitos em informações relevantes para suas análises.

A nomenclatura é constituída dos seguintes itens: assunto, unidade de análise, unidade de conteúdo e unidade de filtro. Cada uma se refere a um elemento de um *Data Warehouse*. O Quadro 7 faz uma prévia de cada conceito da nomenclatura e de sua relação com um DW.

NOMENCLATURA CONCEITUAL	NOMENCLATURA TÉCNICA
Assunto	<i>Data Mart</i>
Unidade de análise	Tabela de fatos ou conjunto de tabelas de fatos.
Unidade de conteúdo	Medidas das tabelas de fatos ou composições com operadores ou funções sobre essas medidas.
Unidade de filtro	Tabela de dimensões.

Quadro 7 - Nomenclatura proposta

Fonte: Da autora

4.2.1.1 Assunto

O conceito técnico denominado de *Data Mart* refere-se, por definição, aos assuntos relacionados à mesma área funcional, ao mesmo departamento ou processo de negócio. Assim, por exemplo, “Vendas”, “Recursos humanos” e “Clientes” são conjuntos de informações que podem constituir um *Data Mart*, ou parte dele, sendo ele chamado de “assunto” pela nomenclatura proposta. Portanto, é possível representar um assunto através de um conjunto de tabelas de fatos e dimensões associadas a um mesmo processo de negócio. No entanto, uma tabela de fato e suas dimensões não são necessariamente associadas a um único assunto.

4.2.1.2 Unidade de análise

Um assunto pode possuir vários tópicos de interesse relacionados e que possam ser importantes para as análises. Esses tópicos indicam unidades de informação que representam características de um assunto específico, com visões e níveis de detalhamento diversos. Os tópicos são denominados, na nomenclatura proposta, de unidades de análise e representam características do assunto sobre o qual se necessita possuir informações e indicadores. As unidades de análise possuem três principais objetivos:

- ✍ permitir que o usuário analise e compreenda o assunto;
- ✍ permitir que o usuário levante ou confirme hipóteses sobre o assunto; e
- ✍ auxiliar no processo de tomada de decisões no que diz respeito ao assunto em questão.

A unidade de análise é implementada fisicamente em uma ou mais tabelas de fato de um *Data Mart*, sendo que uma tabela de fato pode também representar uma ou mais unidades de análise. Por exemplo, o assunto “Vendas” poderia ser analisado por unidades que representam as “Vendas à Vista”, “Vendas a Prazo” ou “Vendas por Cartão”, podendo ser representadas (Figura 8) em uma tabela *Fato_Vendas* ou em três tabelas específicas “*Fato_Vendas_Vista*”, “*Fato_Vendas_Prazo*” e “*Fato_Vendas_Cartao*”.

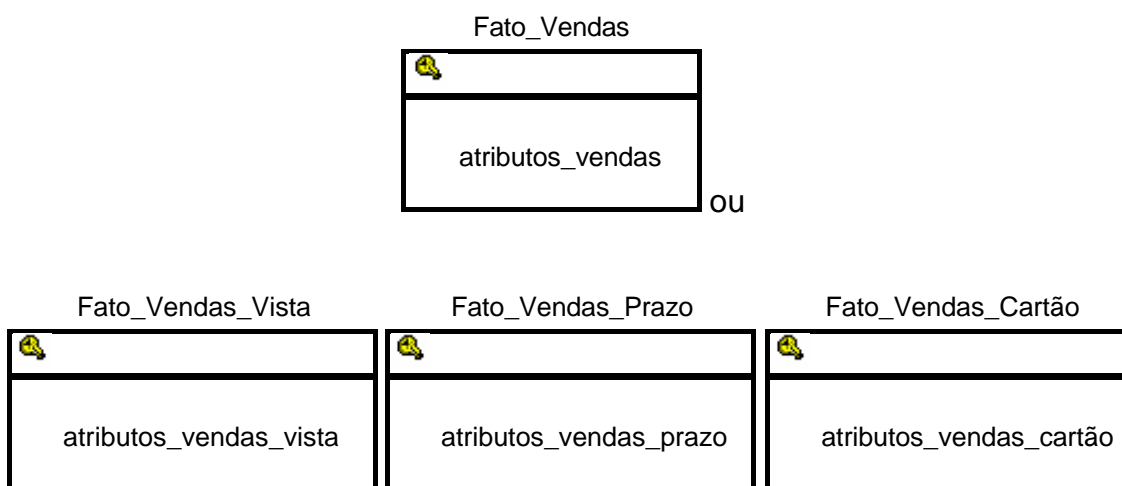


Figura 8 - Possíveis tabelas de fato para o assunto “Vendas”

Fonte: Da autora

4.2.1.3 Unidade de conteúdo

Uma vez determinados os tópicos de análises referentes a um determinado assunto, definem-se as informações que o usuário deseja visualizar. Essas informações podem ser indicadores ou possíveis atributos para a construção de indicadores e são denominadas de unidades de conteúdo. As unidades de conteúdo são fundamentais para a apresentação das informações relevantes referentes às unidades de análise e para permitir as diferentes análises relacionadas a essas unidades.

As unidades de conteúdo são representadas fisicamente pelos atributos que quantificam um determinado fato, também chamados de medidas de uma tabela de fato. Assim, uma unidade de conteúdo pode ser uma coluna da tabela de fato, relações entre medidas das tabelas de fatos ou fórmulas entre colunas distintas, isto é, valores resultantes da utilização de operadores algébricos, como soma, média, contagem e porcentagem. Por exemplo, no tópico (unidade de análise) “Vendas por Cartão” do assunto “Vendas”, as unidades de conteúdo utilizadas como indicadores podem ser “Valor das vendas em R\$”, “Número de prestações” ou “Valor das prestações”, entre outras possíveis medidas. Assim, caso seja utilizada apenas a tabela de fatos “Fato_Vendas” para o tópico “Vendas”, essa tabela conteria as medidas “valor_vendas_real”, “nro_prestacoes” e “valor_prestacao”, como mostra a Figura 9.

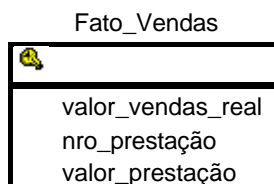


Figura 9 - Possíveis medidas para a tabela Fato_Vendas
Fonte: Da autora

Através das unidades de conteúdo podem ser incluídas algumas das principais funcionalidades de uma ferramenta OLAP, são elas:

- ✍ *operação sort e ranking*: quando os indicadores de um determinado tópico são agrupados e ordenados de acordo com algum padrão, por exemplo, ordem crescente, decrescente, alfabética ou ordem de tamanho baseado em valores numéricos;
- ✍ *operação drill out*: quando o usuário detalha o indicador com informações externas complementares ao selecioná-lo;

- ✍ operação *alert*: quando um indicador ou grupo específico de indicadores aparece em destaque em relação à grande maioria;
- ✍ operação *break*: quando um grupo de informações referente a um indicador é cortado, sendo seus valores mensuráveis totalizados;
- ✍ operação *paging*: quando um grupo de informações referente a um indicador é cortado e distribuído em páginas para que seja visualizado em partes; e
- ✍ operação *filtering*: quando os indicadores a serem apresentados podem ser filtrados de forma a se analisarem alguns deles (é realizado um corte nas unidades de conteúdo).

4.2.1.4 Unidade de filtro

As informações textuais, os indicadores e os índices que o usuário analisa sobre um tópico específico podem ter diferentes visões e níveis de detalhamento, os quais podem refletir tendências importantes sobre um determinado assunto. As visões possibilitam a navegação nas informações de diferentes pontos de vista e em diferentes níveis, o que permite analisar as informações em diversos contextos. As visões das unidades de conteúdo são fisicamente representadas pelas tabelas de dimensão do *Data Mart* e estão diretamente relacionadas às medidas da tabela de fato, visto que as chaves primárias das tabelas de dimensão são chaves estrangeiras das tabelas de fato. Essas visões são aqui denominadas de unidades de filtro.

Para exemplificar, as unidades de conteúdo do tópico “Vendas por Cartão” podem ser visualizadas de forma geográfica (região, UF, cidade) ou de forma

temporal (ano, semestre, trimestre, mês). Região, UF e cidade pertencem a uma mesma hierarquia, e cada uma representa uma unidade de filtro. Portanto, podem ser definidas a dimensão “Dim_Tempo”, com os atributos “mês”, “trimestre”, “semestre” e “ano”, e a dimensão “Dim_Geografia”, com os atributos “cidade”, “UF” e “região”, como mostra a Figura 10.

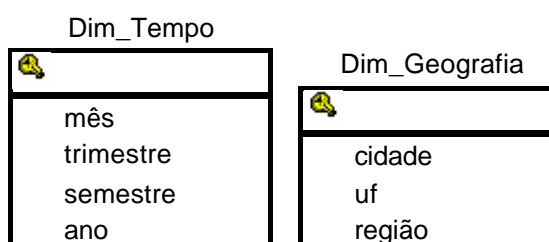


Figura 10 - Possíveis tabelas de dimensão para o assunto “Vendas”

Fonte: Da autora

Através das unidades de filtro podem ser incluídas algumas funcionalidades OLAP, que são:

- ✍ operação *slice*: quando é realizado um corte nas unidades de filtro, como, por exemplo, quando o usuário optar por visualizar uma única região geográfica ou um único Estado;
- ✍ operações *drill down* e *drill up*: quando o usuário detalha o indicador através dos atributos das dimensões hierárquicas (p. ex., dimensão tempo, dimensão geografia). Na primeira ocorre a diminuição da granularidade e, na segunda, o aumento;
- ✍ operação *drill through*: quando o usuário altera a forma de detalhamento do indicador de uma dimensão hierárquica para outra;
- ✍ operação *drill across*: quando o usuário modifica a visualização do indicador em uma dimensão hierárquica de uma tabela de fatos para outra;

☞ operação *dice*: a apresentação das medidas (unidades de conteúdo) cruzadas com as unidades de filtro pode ser realizada de várias formas: as unidades de conteúdo sendo as colunas e as de filtro, as linhas, ou vice-versa; o que diferencia é a forma de implementação da visualização do sistema de *front-end*. A operação *dice* é a alteração na forma de visualização das unidades de conteúdo e de filtro em linhas e colunas.

4.2.1.5 Comparação entre nomenclaturas

Na definição de cada conceito da nomenclatura nas seções anteriores, foram dados exemplos de unidades de análise, unidades de conteúdo e unidades de filtro para o assunto ou *Data Mart* “Vendas”. Utilizando essas unidades e adicionando mais algumas, pode-se resultar em um modelo como o da Figura 11.

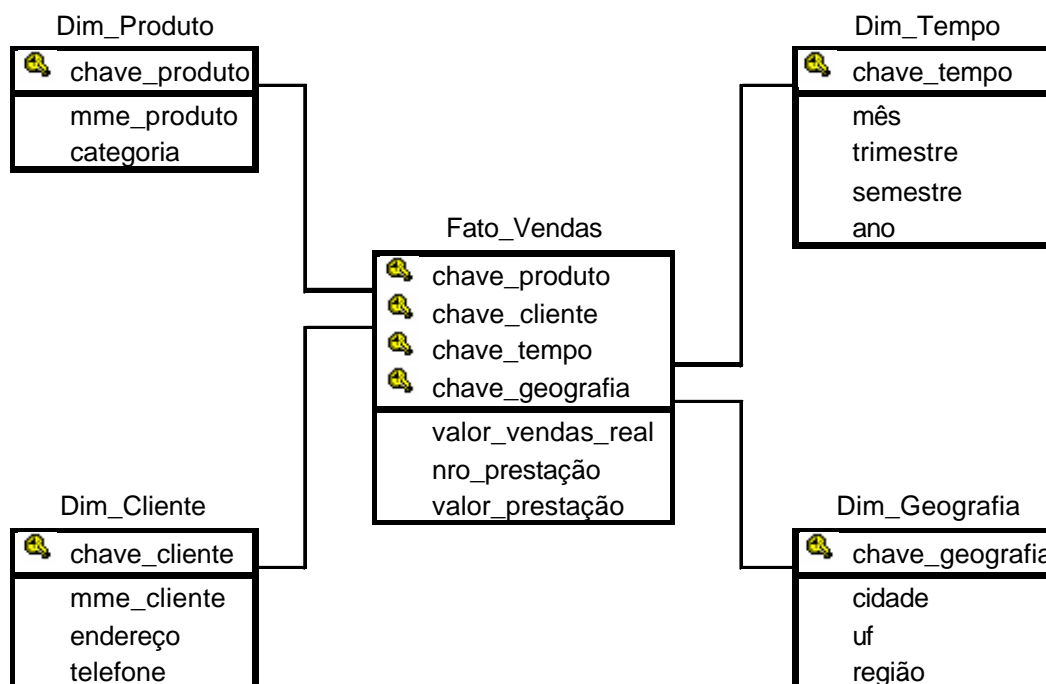


Figura 11 - Data Mart “Vendas”

Fonte: Adaptado de KIMBALL; ROSS, 2002

Os conceitos que compõem a nomenclatura proposta podem ser relacionados a conceitos técnicos conhecidos por analistas e projetistas de *Data Warehouse*. O Quadro 8 mostra a comparação entre esses conceitos.

NOMENCLATURA CONCEITUAL	NOMENCLATURA TÉCNICA	EXEMPLO
Assunto	<i>Data Mart</i> ou conjunto de tabelas de fatos e dimensões associadas ao mesmo processo de negócio.	Vendas
Unidade de análise	Tabela de fatos ou conjunto de tabelas de fatos.	Tabela Fato_Vendas.
Unidade de conteúdo	Medidas das tabelas de fatos ou composições com operadores ou funções sobre essas medidas. As funcionalidades OLAP associadas com as unidades de conteúdo são: operações de “Sort”, “Ranking”, “Drill Out”, “Alert”, “Break” e “Paging”.	Atributos da tabela Fato_Vendas: valor_vendas_real, nro_prestacao, valor_prestacao.
Unidade de filtro	Tabela de dimensões associadas às medidas representadas nas unidades de conteúdo. As funcionalidades OLAP associadas com as unidades de filtro são: operações de “Drill Within” (“Drill Down”, “Drill Up”), “Slice”, “Dice”, “Drill Through” e “Drill Across”.	Tabelas Dim_Tempo, Dim_Geografia, Dim_Produto, Dim_Cliente.

Quadro 8 - Relação entre a nomenclatura conceitual e técnica

Fonte: Da autora

4.2.2 Etapas

O fator principal de um projeto que venha a aplicar a metodologia proposta é identificar os elementos referentes a cada item da nomenclatura descrita. A abordagem utilizada na descrição da metodologia parte do pressuposto de que existe a necessidade de construção de um DW utilizando a forma de implementação *bottom-up*, ou seja, a utilização de *Data Marts* incrementais (KIMBALL; ROSS, 2002). Assim, as etapas podem ser descritas da seguinte forma:

1. a primeira etapa é definir qual será o *Data Mart*, ou seja, identificar o assunto relacionado à mesma área funcional ou ao mesmo processo de negócio. Exemplo: vendas;
2. a segunda etapa é selecionar os tópicos de interesse relacionados ao assunto definido. Esses tópicos devem ser relacionados à mesma área funcional ou ao mesmo processo de negócio. Esta etapa se refere à identificação das “unidades de análise”. Exemplo: vendas a prazo;
3. a terceira etapa é verificar os grupos de informações relevantes relacionadas às unidades de análise e que serão apresentadas aos usuários. Esta etapa se refere à identificação das “unidades de conteúdo”. Exemplo: número de prestações de uma venda a prazo; e
4. por último, na quarta etapa, deve-se compreender de que forma as unidades de análise podem ser exploradas, ou seja, quais as diferentes visões que podem ser mostradas aos usuários, qual a melhor forma de distribuir as informações e quais hierarquias utilizar. Esta etapa se refere à identificação das “unidades de filtro”. Exemplo: agrupamento temporal (mês, semestre e ano).

A Figura 12 apresentada a seguir resume as etapas da metodologia, utilizando a nomenclatura definida.

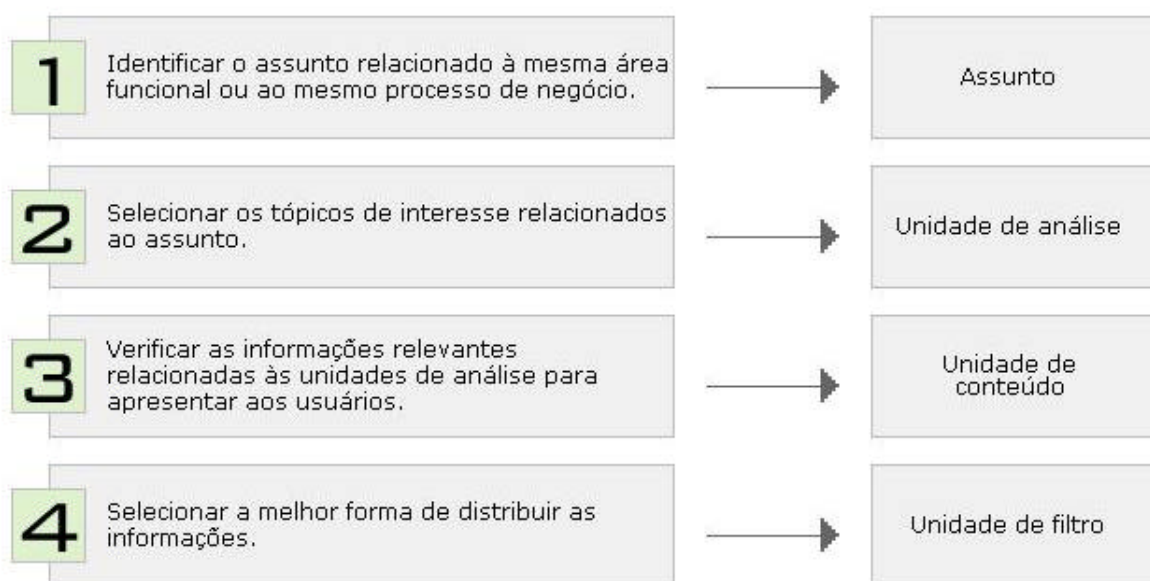


Figura 12 - Etapas da metodologia proposta

Fonte: Da autora

Para o desenvolvimento do sistema de *front-end*, existem diferentes possibilidades na distribuição das unidades de análise, conteúdo e filtro. Com o intuito de oferecer uma forma padronizada de navegação, propõe-se um ciclo de passos a serem seguidos na utilização do sistema, descrito no Quadro 9.

PASSO	PERGUNTA-BASE	MÉTODO
1	O que consultar?	Escolha da unidade de análise
2	Como é possível visualizar as informações?	Escolha da(s) unidade(s) de filtro
3	O que é possível visualizar no resultado?	Escolha da(s) unidade(s) de conteúdo
4	Como detalhar o resultado?	Operações OLAP diversas (<i>drill down, ranking, sort, etc.</i>)
5	O que é possível fazer com o resultado?	Salvar consultas, abrir em planilhas, etc.
6	O que mais?	Cabeçalhos, rodapés, notas, títulos dinâmicos, legendas, <i>helps</i> , FAQs, botão “voltar”, <i>links</i> diversos, etc.

Quadro 9 - Passos de navegação do sistema

Fonte: Da autora

O sistema de navegação utilizando a nomenclatura conceitual proposta oferece uma linguagem que o analista de negócio e o técnico entendem, proporcionando, assim, eficiência, eficácia e produtividade nas consultas.

Para conferir se o *front-end* atende às necessidades da organização e dos usuários bem como aos requisitos levantados no início do projeto, é interessante levar em consideração os objetivos apontados por autores como Cybis (2003). Outras formas de garantir que o sistema seja eficiente podem ser:

- ✍ fazer uma análise para conferir se o sistema atende a critérios de usabilidade de Jakob Nielsen e de outros autores de trabalhos atuais;
- ✍ conferir o sistema fazendo uso de *checklists* que contenham recomendações para o desenvolvimento de interfaces gráficas para que se detectem possíveis falhas com relação à funcionalidade e à praticabilidade do sistema.

4.3 Considerações finais

Este capítulo apresentou uma nomenclatura que faz a ligação entre os conceitos técnicos conhecidos por analistas de sistemas e os conceitos conhecidos pelos analistas de negócio. Identificar os itens dessa nomenclatura e aplicá-los a sistemas DW/OLAP constitui a metodologia proposta neste trabalho.

Com a utilização de tal metodologia aplicada a sistemas DW/OLAP obtém-se um sistema que traduz os requisitos de análise do usuário em processos de recuperação e apresentação de informações, permitindo que ele realize consultas por diferentes níveis de informação e acesse níveis distintos de detalhes. Tudo isso

fornece ao usuário informações relevantes, possibilitando-lhe a realização de análises multidimensionais de modo que tire decisivas conclusões sobre o negócio. Esta metodologia é aplicável em projetos diversificados de informação em que se utiliza a combinação entre as tecnologias *Data Warehouse* e OLAP.

5 APLICAÇÃO – DIRETÓRIO DE GRUPOS DE PESQUISA NO BRASIL

5.1 Introdução

O objetivo deste capítulo é apresentar a metodologia discutida no capítulo anterior empregada em um caso real. Essa metodologia foi aplicada na construção do Plano Tabular, um dos instrumentos de informação do Censo do Diretório de Grupos de Pesquisa no Brasil do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Para contextualizar o Plano Tabular, serão abordados, primeiramente, a Plataforma Lattes e sua arquitetura, especificando-se o contexto desse instrumento, e, em seguida, o Diretório de Grupos de Pesquisa (DGP), com seu histórico, finalidades, logística de funcionamento e estrutura de dados. Mais adiante, será apresentada uma descrição do *Data Mart* do DGP como suporte aos instrumentos de informação. Por último serão apresentados detalhes da implementação da metodologia e sua relação com os conceitos OLAP e de usabilidade na *Web* abordados.

5.2 Plataforma Lattes: uma plataforma de E-gov para gestão de CT&I

A Plataforma Lattes é um conjunto de sistemas de informação, bases de dados, diretórios de serviços e portais *Web* desenvolvidos para apoiar as atividades de Ciência, Tecnologia e Inovação (CT&I) do País. A Plataforma Lattes foi lançada pelo CNPq em 1999 através do Sistema de Currículos Lattes. Segundo a revista Plataforma Lattes (2002), sua concepção teve por finalidade promover a integração

dos sistemas de informação das agências federais de forma a racionalizar o processo de gestão de ciência e tecnologia tanto do ponto de vista do usuário quanto das agências de fomento e das instituições de ensino e pesquisa do País.

A Plataforma Lattes é composta de diversos sistemas distintos, porém integrados entre si. Sua construção foi realizada seguindo a metodologia para desenvolvimento de plataformas de governo proposta por Pacheco (2003). A Figura 13 apresenta a arquitetura da Plataforma Lattes, representada em forma de pirâmide e contendo quatro componentes metodológicos e tecnológicos: a) unidades de informação, definição de ontologias e padronização através da utilização da linguagem de marcação (XML); b) sistemas e fontes de informação; c) portais e serviços *Web*; e d) sistemas de conhecimento e agentes.



Figura 13 - Arquitetura conceitual proposta para projetos de *E-gov*
 Fonte: PACHECO, 2003

5.2.1 Unidades de informação, ontologias e padronização XML

As unidades de informação descrevem subdomínios da área-fim para a qual a plataforma é desenvolvida, sendo formadas por classes ou elementos do domínio da plataforma aos quais estão associados conteúdos, processos e serviços específicos (PACHECO, 2003). Na Plataforma Lattes, as unidades de informação são chamadas de “unidades de análise”, sendo as principais descritas a seguir.

- ✍ Currículo: conjunto de informações que descrevem a atividade profissional dos usuários da Plataforma, ou seja, de todas as pessoas relacionadas à CT&I, tais como pesquisadores, estudantes, docentes, gestores, técnicos de governo e administradores (PACHECO, 2003).
- ✍ Grupo de pesquisa: unidade que descreve os grupos de pesquisa, os quais podem ser definidos como conjunto de indivíduos organizados de acordo com hierarquia de experiência e liderança no terreno científico ou tecnológico em que atuam. Um grupo de pesquisa reúne profissionais envolvidos permanentemente com atividades de pesquisa, cujo trabalho é organizado em torno de linhas comuns de pesquisa, compartilhando instalações e equipamentos (GUIMARÃES et al., 1999).
- ✍ Projeto de Pesquisa: unidade que contempla a descrição das atividades de pesquisa, desenvolvimento ou extensão, realizadas por um pesquisador ou por uma equipe de pesquisa, tendo como base um tema ou objeto específico, com objetivos, metodologia e duração definidos, e tendo os pedidos de auxílios enviados ao CNPq (PACHECO, 2003).
- ✍ Instituição: unidade que representa as informações de organizações, institutos, empresas, universidades e demais organismos ligados à CT&I,

referenciados nas demais unidades como local de lotação profissional ou de pesquisa, ou representando atores institucionais que interagem com o CNPq (PACHECO, 2003).

As ontologias estabelecem compreensão compartilhada e comum de um domínio e podem ser trocadas entre pessoas e computadores (STUDER et al., 2000 apud PACHECO, 2003). Segundo Pacheco (2003), ao se estabelecerem ontologias para unidades de informação de uma plataforma, deve-se produzir padrões compartilháveis e intercambiáveis entre os interessados. Para a produção desses padrões, sugere-se a utilização da linguagem XML (*eXtended Markup Language*), que se constitui na formação de metadados específicos para cada uma das unidades de informação.

Na Plataforma Lattes, existe uma comunidade virtual específica para realizar os processos de construção e atualização de padrões para cada unidade de análise – a Comunidade LMPL (Linguagem de Marcação da Plataforma Lattes) (PACHECO, 2003). A definição de padrões Lattes e, principalmente, sua disseminação e inclusão nos sistemas de informação têm promovido pesquisas específicas da comunidade científica nacional na proposição de ontologias para a Plataforma Lattes (e.g., BONIFÁCIO, 2002 apud PACHECO, 2003).

5.2.2 Fontes e sistemas de informação

Fontes de informação são repositórios de dados referentes a cada unidade da plataforma cujos propósitos são o armazenamento e a gestão das informações públicas, produzidas pela interação com a comunidade usuária ou por enlaces dinâmicos com outras fontes de informação (PACHECO, 2003). As fontes de informação podem incluir todos os tipos de repositórios, inclusive *Data Warehouse*, o repositório utilizado para este trabalho.

A Plataforma Lattes possui três conjuntos de fontes de informação: (1) bases operacionais, (2) bases *warehouse* e (3) repositórios de informações *Web*. Como exemplo de base operacional, é possível citar a base Fomento do CNPq, que contém, entre outras, informações correspondentes aos bolsistas da agência, em todos os seus níveis, desde a iniciação científica até os bolsistas de produtividade, passando pelos bolsistas de mestrado e doutorado. Entre as bases *warehouse*, tem-se, por exemplo, o DMCurric, o *Data Mart* de Currículos Lattes, que contém todos os currículos Lattes enviados pelos usuários da Agência e sobre o qual se realizam as consultas textuais e a produção de indicadores.

Enquanto isso, os sistemas de informação, já vistos anteriormente, são componentes tecnológicos responsáveis pela captura e pelo tratamento da informação fornecida pela comunidade usuária à Plataforma. Seu objetivo é garantir que as informações incluídas nas bases de dados sigam as regras e especificações estabelecidas para a unidade correspondente (PACHECO, 2003).

A Plataforma Lattes dispõe de um conjunto de sistemas de informação on-line e off-line responsáveis pelo controle das informações referentes às suas unidades de análise. Como exemplo de sistema de informação da Plataforma tem-se o

Currículo Lattes (CV-Lattes), sistema de gestão de dados curriculares para a unidade de análise Currículo que oferece as versões on-line e off-line.

5.2.3 Portais e serviços *Web*

Portais e serviços *Web* são instrumentos desenvolvidos para apresentar informações na *Web* (*Websites*), publicar informações dinamicamente atualizadas com interação com a comunidade usuária (portais *Web*) e disponibilizar serviços de informação de governo na *Web* (*Web services*) (PACHECO, 2003).

Nesta camada, a Plataforma Lattes possui diversos sistemas de busca e apresentação de informações para as unidades de análise. Entre eles, é possível citar o site do projeto Diretório de Grupos de Pesquisa, um portal com acesso a diferentes sistemas que apresentam informações sobre a pesquisa nacional para a unidade de análise Grupo.

5.2.4 Sistemas de conhecimento e agentes

Sistemas de conhecimento são instrumentos projetados para gerar novos conhecimentos a partir das fontes de informação da plataforma e de sua operação por parte da comunidade usuária. Já agentes são alternativas tecnológicas e metodológicas que caracterizam uma das formas pelas quais os recursos desses sistemas podem ser disponibilizados à comunidade (PACHECO, 2003).

Os componentes da arquitetura representada pela pirâmide são integrados, sendo que, segundo Pacheco (2003), a camada de sistemas de conhecimento relaciona-se com todas as demais camadas da arquitetura. O autor afirma que os

sistemas de conhecimento encontram na camada de unidades de informação especificações e diretrizes sobre a ontologia dos domínios cobertos pela plataforma bem como especificações sobre os padrões de cada unidade de informação. Na segunda camada, os sistemas de conhecimento encontram os repositórios de dados e informações, fontes para as técnicas de extração de conhecimento. Os sistemas de conhecimento interagem com a terceira camada para tornar disponíveis seus serviços e para publicar seus resultados. Por fim, a configuração de agentes prevista na camada de sistemas de conhecimento indica uma das formas pelas quais esses sistemas poderão interagir com a comunidade usuária e com as demais camadas da plataforma (PACHECO, 2003).

Lattes Redes

Entre alguns exemplos de sistemas de conhecimento da Plataforma Lattes, é possível citar o Lattes Redes, uma aplicação de *Link Analysis* desenvolvida para extrair conhecimento em redes de relacionamento. De acordo com Balancieri (2004), o Sistema Lattes Redes tem o objetivo de investigar, induzir, avaliar e analisar as redes de pesquisa sob a ótica do sistema CT&I, através dos dados extraídos da Plataforma Lattes, segundo diversas possibilidades de visualização. Sendo assim, os atores do sistema CT&I passam da condição de fornecedores de dados a beneficiários da TI (BALANCIERI, 2004).

Lattes Egressos

Também construído a partir de conceitos de *Link Analysis*, tem-se o Lattes Egressos, outro sistema de conhecimento da Plataforma Lattes. O Lattes Egressos tem o objetivo de apresentar informações sobre egressos de cursos das Instituições

de Ensino Superior (IES) do País. Segundo Bovo (2004), as informações são apresentadas na forma de um grafo radial, no qual a instituição de estudo é representada por um vértice central, enquanto os vértices circundantes representam algum aspecto da informação sobre os egressos da instituição.

Lattes RedesGP

O mais novo sistema de conhecimento da Plataforma Lattes baseado na teoria do Lattes Redes é o Lattes RedesGP, que tem como finalidade revelar a estrutura e o conteúdo de um conjunto de informações que se relacionam a grupos de pesquisa por meio de unidades interconectadas entre si. Assim como no Lattes Egressos, as informações são apresentadas na forma de um grafo radial, podendo-se ter um grupo de pesquisa ou uma instituição como vértice central, sendo a lógica do sistema bastante similar. Uma vantagem do Lattes RedesGP é sua integração com o Diretório de Grupos de Pesquisa. Essa integração é apresentada com detalhes mais adiante, visto que o módulo da aplicação da metodologia em um caso real, apresentado neste trabalho, faz parte da integração.

Sistemas de geração automática de texto

Dois novos módulos do Sistema de Currículos Lattes, denominados CV-Resumé e CV-Perfil, foram propostos por Martins (2004) e também se encaixam no topo da pirâmide da arquitetura da Plataforma Lattes. Ambos os módulos foram criados por meio de uma tecnologia denominada InterLattes – a qual permite que recursos sejam inseridos na Plataforma Lattes sem que seja necessário gerar novas versões – e baseiam-se em regras que definem sentenças a serem geradas automaticamente. Esses dois sistemas permitem total integração com o currículo de

um indivíduo em um único ambiente, gerando novos conhecimentos a partir das fontes de informação oriundas da Plataforma Lattes e de sua operação por parte da comunidade usuária (MARTINS, 2004).

Plano Tabular – Diretório de Grupos de Pesquisa no Brasil

A metodologia proposta neste trabalho permite o desenvolvimento de sistemas de conhecimento, sendo um deles descrito neste capítulo: o Plano Tabular, site integrante do portal do Diretório de Grupos de Pesquisa no Brasil (DGP). O DGP é baseado em técnicas OLAP, que, segundo Pacheco (2003), encaixa-se no perfil do topo da pirâmide da arquitetura proposta. Tanto o Plano Tabular quanto os dois outros sistemas citados posteriormente e integrados a ele (Lattes RedesGP e Busca Textual) têm como unidade de análise os Grupos de Pesquisa e são baseados na mesma base *warehouse*, denominada DMGrupo.

Adiante, primeiramente será contextualizado o Plano Tabular, de forma que serão apresentadas algumas informações a respeito do DGP. Em seguida, será vista a metodologia aplicada no desenvolvimento do *Data Mart* dos Grupos de pesquisa e do Plano Tabular.

5.3 O Diretório de Grupos de Pesquisa no Brasil

O Diretório de Grupos de Pesquisa no Brasil (DGP) é um projeto do CNPq que se originou 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. O intuito era 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 projeto pretendia oferecer um suporte informacional atualizado sobre as atividades de pesquisa através de bases censitárias sobre os grupos de pesquisa em atividade no País. Atualmente, o DGP tem o claro objetivo de ser uma plataforma de informação sobre o parque científico e tecnológico brasileiro (CNPq, 2002). O esforço empreendido pelo Brasil após a Segunda Grande Guerra gerou o maior parque de C&T (Ciência e Tecnologia) da América Latina. Entretanto, ainda há carência de informação organizada a respeito, o que enfraquece e dificulta a tomada de decisão sobre os desígnios de C&T nacional. Tal fato transforma o DGP em um instrumento essencial para a gestão de Ciência e Tecnologia (MARTINS; GALVÃO, 1994).

No DGP destacam-se três finalidades principais (CNPq, 2002):

- ✍ serve como um eficiente instrumento para o intercâmbio de informações entre os integrantes da comunidade científica e tecnológica. O sistema responde com precisão quem é quem, onde se encontra, o que está fazendo e o que já produziu;
- ✍ no âmbito do planejamento e da gestão das atividades científicas e tecnológicas, o DGP é, talvez, a mais poderosa ferramenta já desenvolvida no Brasil. Seja no nível das instituições, seja no das sociedades científicas ou, ainda, no das várias instâncias de organização político-administrativa do País, a base de dados é uma fonte inesgotável de informação; e
- ✍ sua base de dados tem o importante papel de preservar a memória da atividade científico-tecnológica no Brasil.

Segundo o CNPq (2002),

[...] desde 1997, em sua versão 3.0, o DGP é capaz de descrever com precisão os limites e o perfil geral da atividade científico-tecnológica no Brasil. Igualmente, é capaz de fornecer aos interessados uma grande e diversificada massa de informação sobre detalhes de quem realiza as atividades, como e onde se realizam e sobre o que versam.

Desde a versão 4.0, de 2000, o DGP está integrado à Plataforma Lattes por meio do Sistema de Currículos Lattes. No ano de 2004 foi realizado o sexto Censo do Diretório de Grupos de Pesquisa, sendo aprofundado seu vínculo com outras fontes de informação. Dessa forma, o DGP inclui informações de fomento do CNPq (bolsistas e orientadores), informações da pós-graduação provenientes do Sistema Coleta da Capes e informações sobre o Currículo Lattes dos participantes em grupos de pesquisa.

Para um país que logrou desenvolver uma capacidade instalada de pesquisa com tamanha complexidade, é imprescindível a existência de instrumentos de acompanhamento e avaliação continuados e padronizados que ajudem a convencer o governo e a comunidade científico-tecnológica de que vale a pena investir nessa atividade. No estágio em que se encontra, o DGP é, sem sombra de dúvida, um desses instrumentos, um dos mais importantes de que dispõe o País (CNPq, 2002).

5.3.1 As fases do DGP

O DGP é formado por duas fases. A primeira é a fase de coleta de dados, que consiste em um conjunto de sistemas de captura de dados, utilizado pelos principais atores do sistema, que constitui a base de dados corrente do DGP. A segunda fase, chamada de fase censitária, é um conjunto de sistemas que constrói a base de

dados do censo e apresenta os resultados em diversos instrumentos de informação agrupados em um site.

Na fase de coleta de dados, os principais atores dos sistemas de captura são os dirigentes, os líderes de grupos de pesquisa, os pesquisadores e os estudantes. Cada um desses atores tem características e funções próprias no DGP, como apresentado a seguir.

- ✍ Dirigentes: têm a difícil função de reconhecer os líderes de grupos, validar e cancelar as informações cadastradas por eles para controlar o andamento da coleta de grupos de pesquisa de sua instituição. Tudo isso para que, na fase censitária, a instituição seja representada de forma adequada e reflita fielmente sua pesquisa no DGP.
- ✍ Líderes: têm a função de cadastrar e atualizar constantemente as informações do seu grupo, dos pesquisadores e estudantes que fazem parte do grupo, das linhas de pesquisa com seus objetivos, palavras-chaves, áreas do conhecimento e setores de aplicação, e dos vínculos do grupo com o setor empresarial.
- ✍ Pesquisadores e estudantes: têm como principal função manter seus currículos Lattes atualizados, pois os dados dos currículos complementam os dados do grupo na fase censitária. Eles compartilham com o líder o mesmo grupo de pesquisa, linha de pesquisa, instalações, laboratórios, recursos, e têm à disposição, através do DGP, informações sobre os grupos em que atuam.

Na fase censitária, outro conjunto de sistemas entra em operação. Esses sistemas têm duas funções principais: (1) complementar os dados da fase de coleta e (2) apresentar as informações à comunidade de C&T e ao público em geral.

A base censitária do DGP é construída da seguinte forma: os dados obtidos da fase de coleta são armazenados na base de dados corrente, depois são complementados os dados do Sistema de Currículos Lattes, da base Fomento do CNPq e da base DataCapes da Capes. O Sistema de Currículos Lattes armazena os currículos dos pesquisadores e estudantes integrantes dos grupos de pesquisa. Os dados da base Fomento do CNPq são correspondentes aos bolsistas de iniciação científica (IC), do Programa Institucional de Bolsas de Iniciação Científica (PIB IC), de mestrado e doutorado, de produtividade em pesquisa e aos dados de orientadores de IC, mestrado e doutorado. Por fim, os dados provenientes da base DataCapes da Capes são referentes aos cursos de pós-graduação nacionais em que os pesquisadores do DGP atuam.

5.3.2 O *Data Mart* do Diretório de Grupos de Pesquisa

Para armazenar os dados referentes aos grupos de pesquisa em uma base de dados, são necessárias 76 tabelas com uma complexa malha de interligações, apenas para o esquema do banco de dados operacional. Para utilizar essa base de dados na extração de informações gerenciais, a complexidade da estrutura de informações torna a consulta custosa e com tendência à lentidão. O alto grau de conexão entre os assuntos descritos no esquema do banco de dados operacional o faz inadequado à recuperação de informação gerencial. Portanto, há necessidade de um repositório que viabilize esse processamento.

Além disso, as análises realizadas do DGP não dependem somente dos dados que constam na base de dados operacional. Como já foi dito anteriormente, o DGP depende também dos dados armazenados em outras bases de dados. Esse é

o cenário que deu origem ao desenvolvimento de um *Data Mart* que armazene informação textual e que atenda às necessidades de informação gerencial da Agência.

A definição de um *Data Mart* para grupos de pesquisa permite a complementação das informações operacionais do DGP, sendo esta, conforme descrito no capítulo anterior, a primeira etapa da metodologia. Assim, a construção do *Data Mart* de Grupos de Pesquisa (DMGRUPO) seguiu basicamente as seguintes etapas: (a) seleção dos tópicos de interesse relacionados ao assunto definido; (b) verificação dos grupos de informações relevantes relacionados a esses tópicos e que serão apresentados aos usuários; e (c) compreensão da forma como os tópicos podem ser explorados e distribuídos. O Quadro 10 resume os elementos do *Data Mart* a serem definidos e sua comparação com as nomenclaturas conceituais e técnicas.

ELEMENTOS DO <i>DATA MART</i>	NOMENCLATURA CONCEITUAL	NOMENCLATURA TÉCNICA
Tópicos de interesse	Unidade de análise	Tabelas de fato
Grupos de informações relevantes a serem apresentadas	Unidade de conteúdo	Medidas das tabelas de fato ou relação entre elas
Formas de explorar e distribuir as informações	Unidade de filtro	Tabelas de dimensão

Quadro 10 - Comparação dos elementos do *Data Mart* com a nomenclatura conceitual e a técnica

Fonte: Da autora

Assim, foram identificadas as unidades da nomenclatura conceituada na metodologia proposta e implementadas as tabelas de acordo com essas unidades. O resultado é apresentado no Quadro 11. Neste quadro, como se pode observar,

para facilitar a identificação dos tipos de tabela foram utilizadas as iniciais FT para as tabelas de fato e DI para as tabelas de dimensão.

TÓPICO	TABELAS FATO	MEDIDAS	TABELAS DIMENSÃO	ATRIBUTOS DA TABELA DIMENSÃO
Grupos de pesquisa	FT_GRUPO	Ano do censo, ano de formação, nome do grupo, nº de identificação do grupo, total de estudantes, total de linhas de pesquisa, total de pesquisadores, total de técnicos, etc.	DI_CENSO, DI_AREA_CONHEC, DI_GEOGRAFIA, DI_INSTITUICAO, DI_PESSOA, DI_SETOR_ATIVIDADE	País, região, Unidade da Federação (UF), instituição, grande área, área, etc.
Linhas de pesquisa	FT_LINHA_PESQUISA FT_PESSOA_LINHA_PESQUISA	Ano do censo, nome da linha de pesquisa, total de pesquisadores, total de estudantes, etc.	DI_CENSO, DI_AREA_CONHEC, DI_GEOGRAFIA, DI_INSTITUICAO, DI_PESSOA,	País, região, UF, instituição, grande área, área, setor de aplicação, etc.
Integrantes de grupos (pesquisadores e estudantes)	FT_PESSOA_GRUPO	Ano do censo, campos que identificam se o integrante é pesquisador ou estudante, e seu status em relação a sexo, formação, nacionalidade, liderança, etc.	DI_CENSO, DI_AREA_CONHEC, DI_GEOGRAFIA, DI_INSTITUICAO, DI_PESSOA,	País, região, UF, instituição, grande área, área, setor de aplicação, nome da pessoa, data de nascimento, etc.
Técnicos	FT_TECNICO_GRUPO	Ano do censo, total de técnicos com níveis de formação: 1º e 2º graus, 3º grau, mestrado e doutorado, etc.	DI_CENSO, DI_AREA_CONHEC, DI_GEOGRAFIA, DI_INSTITUICAO, DI_PESSOA, DI_TECNICO_GRUPO, DI_ATIVIDADE	País, região, UF, instituição, grande área, área, atividade técnica, etc.
Produção CT&A (Científica, Tecnológica e Artística)	FT_PRODUCAO_GRUPOS	Total de produções bibliográficas, produções técnicas, orientações concluídas, produções artísticas/culturais, demais trabalhos relevantes, etc.	DI_CENSO, DI_AREA_CONHEC, DI_GEOGRAFIA, DI_INSTITUICAO, DI_PESSOA, DI_PRODUCAO	País, região, UF, instituição, grande área, área, etc.

Empresas	FT_EMPRESA A_GRUPO	Ano do censo	DI_CENSO, DI_AREA_CONHEC, DI_GEOGRAFIA, , DI_INSTITUICAO, DI_PESSOA, DI_EMPRESA, DI_RELACIONAMENTO, DI_REMUNERACAO	País, região, U F, instituição, grande área, nome da empresa, ramo de atividade, natureza jurídica, pessoal ocupado, etc.
----------	-----------------------	--------------	--	---

Quadro 11 - Unidades de análise, conteúdo e filtro do DMGRUPO

Fonte: Da autora

Definidas as unidades, o próximo passo foi modelar o DMGRUPO. A Figura 14 ilustra um fragmento do modelo – uma tabela de fato e suas respectivas dimensões (o DMGRUPO do DGP se encontra na íntegra no Apêndice A).

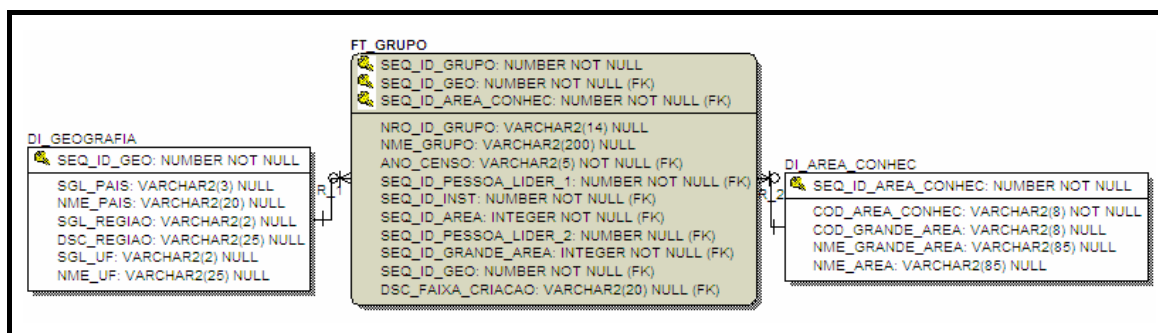


Figura 14 - Fragmento do DMGRUPO

Fonte: Da autora

A tabela de fato FT_GRUPO armazena as medidas relacionadas ao tópico ou à unidade de análise Grupo de Pesquisa, tendo como atributos, por exemplo, o nome do grupo e sua referência. As dimensões armazenam atributos que adicionam diferentes formas de explorar os grupos de pesquisa. Muitas das dimensões que estão ligadas ao FT_GRUPO são também ligadas a outras tabelas de fato. A dimensão DI_GEOGRAFIA, por exemplo, é ligada a várias outras tabelas de fatos,

pois existe a necessidade de distribuir as informações por país, região e UF para todas as outras unidades de análise definidas no modelo. Além disso, algumas tabelas de fato também estão ligadas entre si. A tabela FT_GRUPO, por exemplo, está ligada a FT_PESSOA_GRUPO (Figura 15). Ambas armazenam dados sobre a unidade de análise Grupos de Pesquisa, porém a tabela FT_PESSOA_GRUPO pode retornar informações sobre os integrantes dos grupos de pesquisa, como a formação de um pesquisador.

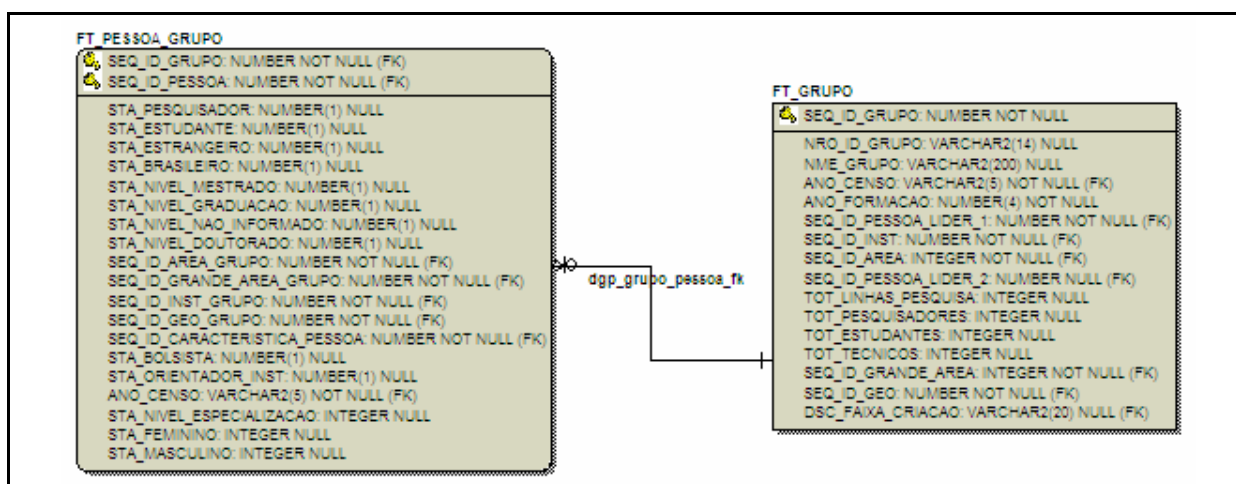


Figura 15 - Fragmento do DMGRUPO

Fonte: Da autora

5.3.3 Instrumentos de informação do DGP

Sobre o DMGRUPO foram desenvolvidos instrumentos de análise e recuperação textual das informações, apresentadas à comunidade através de um conjunto de sistemas de informação que constituem o site do Censo do Diretório de Grupos de Pesquisa. Segue uma breve definição desses instrumentos.

- ✍ **Séries Históricas:** instrumento que contém tabelas e gráficos com informações que sintetizam a evolução temporal e agregada do perfil das

principais unidades de informação do DGP ao longo dos censos realizados.

- ✍ **Súmula Estatística:** disponibiliza um conjunto de tabelas e gráficos selecionados, com informações que sintetizam o conteúdo da base de dados e fornecem um retrato bastante nítido da capacidade instalada de pesquisa no País. Embora de conteúdo predominantemente descritivo, fornece pistas analíticas sobre determinadas características desse retrato.
- ✍ **Estratificação:** possibilita a divulgação dos critérios e 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.
- ✍ **Plano Tabular:** permite visualizar quantitativamente o perfil da pesquisa no Brasil. O sistema oferece a possibilidade de cruzamento de variáveis capazes de gerar um grande número de tabelas, que podem ser salvas em outros formatos de arquivos para futuras consultas.
- ✍ **Busca Textual:** permite recuperar informações sobre os dados presentes na base DMGRUPO. Organiza-se a partir de três decisões a serem tomadas pelo usuário: qual informação deverá ser recuperada, onde essa informação deverá ser encontrada e segundo que critério deve ser procurada. Este instrumento possibilita visualizar detalhes textuais relacionados aos grupos de pesquisa do País.

As Séries Históricas, a Súmula Estatística e a Estratificação são geradas e apresentadas estaticamente a partir das informações do DMGRUPO, enquanto o Plano Tabular e a Busca Textual fazem o acesso ao DMGRUPO para a construção dinâmica dos resultados.

O instrumento a ser explorado nesta pesquisa é o Plano Tabular. Em virtude da necessidade de extrair e apresentar as informações relevantes do DMGRUPO à sociedade em geral, o Plano Tabular foi elaborado utilizando-se a metodologia descrita no capítulo anterior.

5.3.4 Plano Tabular

O Plano Tabular foi implementado pela primeira vez em 1995, na versão 2.0 do Diretório de Grupos de Pesquisa no Brasil. Nesta época, ele oferecia estatísticas limitadas com apenas algumas tabelas estáticas. Em 1997, em sua versão 3.0, o número de tabelas aumentou, porém as consultas continuavam estáticas. Somente em 2000, em sua versão 4.0, a forma de consulta do Plano Tabular foi modificada de modo a apresentar as informações dinamicamente. A versão explicitada neste trabalho é a utilizada no Censo de Grupos de Pesquisa 2004 (versão 6.0). A evolução do Plano Tabular pode ser visualizada das Figuras 23 a 28 do Apêndice C.

5.3.4.1 Aplicação da metodologia proposta

Como o DMGRUPO está pronto e disponível para ser utilizado, para aplicar a metodologia proposta no Plano Tabular, foi identificado cada item da nomenclatura conceitual (unidades de análise, de conteúdo e de filtro) e definida a melhor forma de apresentá-los de acordo com os seus objetivos e necessidades.

Baseando-se nas tabelas de fato do DMGRUPO (FT_GRUPO, FT_PESSOA_GRUPO, FT_LINHA_PESQUISA, FT_TECNICO_GRUPO, FT_PESSOA_LINHA_PESQUISA, FT_EMPRESA_GRUPO, FT_PRODUCAO e

FT_PESSOA_INSTITUICAO), foram definidas oito unidades de análise para o Plano Tabular, sendo elas: (1) Grupos de Pesquisa, (2) Linhas de Pesquisa, (3) Pesquisadores, (4) Estudantes, (5) Técnicos, (6) Produção CT&A, (7) Grupos/Empresas e (8) Empresas/Grupos

Para identificar as unidades de conteúdo do Plano Tabular, foi feito o espelho de algumas medidas das tabelas de fato do DMGRUPO. As unidades de conteúdo são as informações numéricas que o usuário visualiza em relação a uma unidade de análise específica. Por exemplo, para a unidade de análise Grupos de Pesquisa, o usuário visualiza Grupos, Pesquisadores, Estudantes, Linhas de Pesquisa e algumas relações numéricas entre esses itens.

Baseando-se nas tabelas de dimensão e hierarquias do DMGRUPO, foram identificadas as unidades de filtro do Plano Tabular. Da dimensão DI_GEOGRAFIA, por exemplo, foram extraídas três unidades de filtro: País, Região e UF. A partir das outras dimensões (DI_INSTITUICAO, DI_AREA_CONHEC, DI_ATIVIDADE_TECNICA, DI_PESSOA, DI_EMPRESA, DI_SETOR_ATIV_ECON, DI_SETOR_APLICACAO, etc.), foram extraídas outras unidades de filtro para serem utilizadas na distribuição das informações. Cada unidade de análise possui suas unidades de filtro específicas, de acordo com sua relevância e necessidade para visualização.

O Quadro 12 resume todas as unidades de filtro e de conteúdo referentes a cada unidade de análise do Plano Tabular.

UNIDADES DE ANÁLISE	UNIDADES DE FILTRO	UNIDADES DE CONTEÚDO
Grupos de pesquisa	<ul style="list-style-type: none"> - Geográfico/Institucional: Brasil, Região, Unidade da Federação, Instituição. - Grande área/Área: Grande área, Área. - Outras: Ano de formação. 	<ul style="list-style-type: none"> - Grupos de Pesquisa (G). - Linhas de Pesquisa (L). - Pesquisadores (P). - Estudantes (E). - Técnicos. - Relações: L/G; P/G; E/G; P/L.
Linhas de pesquisa	<ul style="list-style-type: none"> - Geográfico/Institucional: Brasil, Região, Unidade da Federação, Instituição. - Grande área/Área/Setor: Grande área, Área, Setor de aplicação. 	<ul style="list-style-type: none"> - Linhas de pesquisa (L). - Pesquisadores (P). - Grupos de pesquisa (G). - Relações: L/G; P/L; P/G.
Pesquisadores	<ul style="list-style-type: none"> - Geográfico/Institucional: Brasil, Região, Unidade da Federação e Instituição. - Grande área/Área/Setor: Grande área, Área, Setor de aplicação. - Outras: Faixa etária. 	<ul style="list-style-type: none"> - Formação. - Sexo. - Liderança. - Nacionalidade.
Estudantes	<ul style="list-style-type: none"> Geográfico/Institucional: Brasil, Região, Unidade da Federação, Instituição. - Grande área/Área: Grande área, Área. - Outras: Faixa etária. 	<ul style="list-style-type: none"> - Nível de Treinamento. - Sexo. - Nacionalidade.
Técnicos	<ul style="list-style-type: none"> - Geográfico/Institucional: Brasil, Região, Unidade da Federação, Instituição. - Grande área/Área: Grande área, Área. - Outras: Atividade técnica. 	<ul style="list-style-type: none"> - Total geral. - 1º e 2º graus. - 3º grau. - Mestrado. - Doutorado.
Produção CT&A	<ul style="list-style-type: none"> - Geográfico/Institucional: Brasil, Região, Unidade da Federação, Instituição - Grande área/Área: Grande área, Área - Outras: Ano da produção 	<ul style="list-style-type: none"> - Tipo de produção: Produção bibliográfica, Produção técnica, Orientação concluída, Produção artística/cultural e demais trabalhos relevantes, Todos - Autores: Pesquisadores, Estudantes, Pesquisadores doutores, Todos

Grupos/ Empresas	- Geográfico/Institucional: Brasil, Região, Unidade da Federação, Instituição - Grande área/Área: Grande área, Área	- Grupos - Empresas - Tipo de relacionamento - Tipo de remuneração
Empresas/ Grupos	- Geográfico: Brasil, Região, Unidade da Federação - Outras: Ramo de atividade, Natureza jurídica, Pessoal ocupado	- Empresas - Grupos - Tipo de relacionamento - Tipo de remuneração

Quadro 12 - Unidades de análise, de filtro e de conteúdo do Plano Tabular

Fonte: Da autora

Definidas as unidades, o próximo passo foi indicar a distribuição dessas unidades na interface gráfica, a forma e a ordem de navegação entre elas e que operações e cruzamentos entre elas são necessários para os usuários do sistema.

A Figura 16 mostra a interface gráfica do Plano Tabular e a disposição de todas as unidades de informação definidas. As unidades de análise estão demarcadas como número um (1), as unidades de filtro como número dois (2) e as unidades de conteúdo como número três (3).

Figura 16 - Plano Tabular e suas unidades de informação
Fonte: Da autora

A navegação do Plano Tabular segue basicamente estes passos: (1) seleção de uma unidade de análise; (2) seleção de ao menos uma unidade de filtro para fazer o detalhamento da pesquisa, dependendo da forma como o usuário preferir distribuir e agrupar as informações; e (3) seleção de ao menos uma unidade de conteúdo que, dependendo da unidade de análise, já vem preestabelecida.

Como resultado, obteve-se um sistema de extração automática de indicadores, com características de aplicação OLAP. Conforme relata o Quadro 12, o DMGRUPO armazena informações de forma hierarquizada, tornando possível a recuperação de informações de maneira hierarquizada e a realização de cruzamentos dos tópicos e de suas correspondentes visões.

Tanto os passos de navegação quanto as funcionalidades OLAP referentes ao Plano Tabular serão detalhados em seções específicas adiante.

5.3.4.2 Funcionamento

O Plano Tabular apresenta um ciclo padronizado de navegação, o que facilita sua utilização por parte de diferentes tipos de usuários. Esse ciclo segue os passos descritos no Quadro 13, ciclo este que já foi apresentado no Quadro 9 do capítulo anterior, no entanto, agora está adaptado a esse sistema.

PASSO	PERGUNTA-BASE	MÉTODO
1	O que consultar?	Escolha da unidade de análise (Grupos, Linhas, Pesquisadores, Produção, etc.).
2	Como é possível consultar?	Escolha da(s) unidade(s) de filtro (Região, Grande Área, Instituição, etc.).
3	O que apresentar?	(Eventual*) Seleção da opção para a primeira unidade de filtro (para Região, por exemplo, a região Sudeste).
4	O que é possível ver no resultado?	(Eventual**) Seleção da(s) unidade(s) de conteúdo (Formação, Sexo, Nacionalidade, etc.).
5	Como detalhar o resultado?	Matrizes com <i>links</i> dinâmicos nas células (operações OLAP, acesso ao sistema Busca Textual e ao sistema Lattes RedesGP).
6	O que é possível fazer com o resultado?	Salvamento de consultas.
7	O que mais?	Cabeçalhos; Notas Técnicas; Notas de Rodapé; Títulos Dinâmicos; Legendas; Saiba Mais; Exemplos; Voltar.
* Para casos em que é selecionada mais de uma unidade de filtro.		
** Opção não disponível para as seguintes unidades de análise: Grupos, Linhas de Pesquisa e Técnicos.		

Quadro 13 - Passos de navegação no Plano Tabular

Fonte: Da autora

A Figura 17 apresenta um exemplo de consulta para a unidade de análise “grupos”. Nessa consulta, foram selecionados a unidade de filtro “região” e o ano do

censo “2002”. O resultado mostra o número dos grupos de pesquisa distribuídos pelas regiões do País.

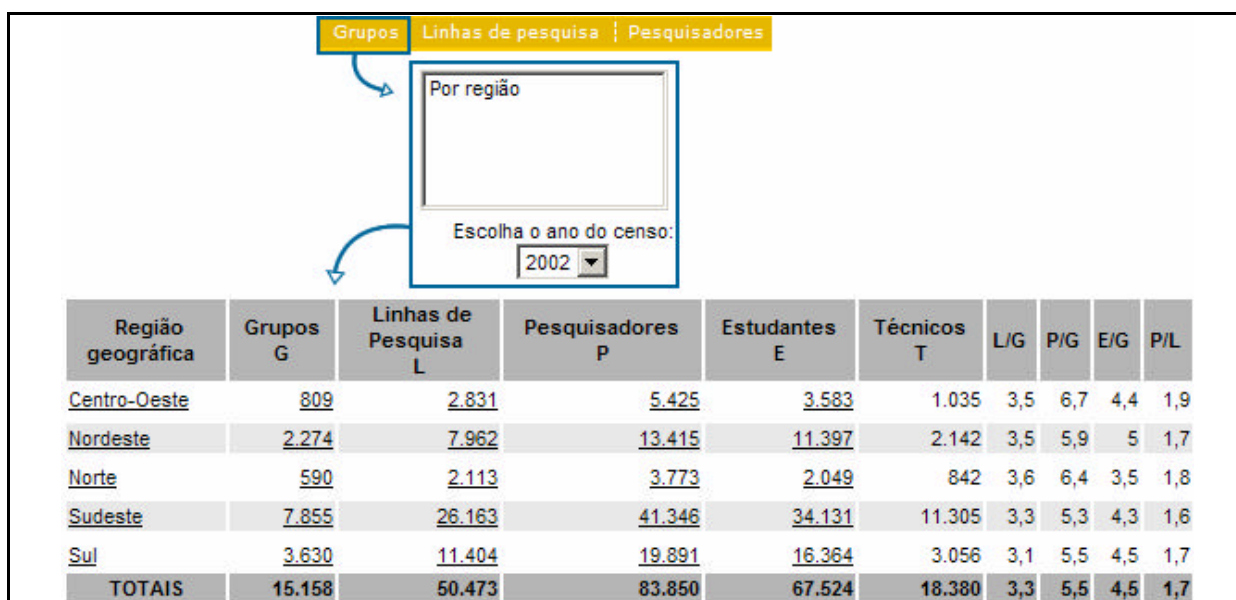


Figura 17 - Consulta da unidade de análise “grupos de pesquisa” pela unidade de filtro “região”

Fonte: Da autora

O usuário tem a opção de selecionar até quatro unidades de filtro, sendo esta uma limitação do sistema desenvolvido e não da metodologia utilizada. Caso ele selecione mais de uma unidade de filtro, deverá escolher uma opção de filtro para a primeira unidade. Supondo, por exemplo, que em uma consulta para a unidade de análise “grupos” sejam selecionadas as unidades de filtro “região” e “instituição”, antes do resultado ser apresentado, deve-se selecionar uma região do País (Nordeste, Sul, etc.). Caso tenha sido selecionada a região “Sul”, por exemplo, a consulta final apresentará o número dos grupos de pesquisa distribuídos para as instituições da região Sul.

Como já foi dito anteriormente, as formas de apresentação das unidades de conteúdo não estão preestabelecidas para todas as unidades de análise. Portanto,

em alguns casos o usuário deverá selecioná-las. Para a unidade de análise “pesquisadores”, por exemplo, o usuário poderá escolher se deseja agrupar os pesquisadores por “formação”, “sexo”, “liderança” ou “nacionalidade”, ou poderá agrupá-los selecionando algumas dessas opções concomitantemente. A Figura 18 mostra o número de pesquisadores agrupados de acordo com suas formações e distribuídos por região.

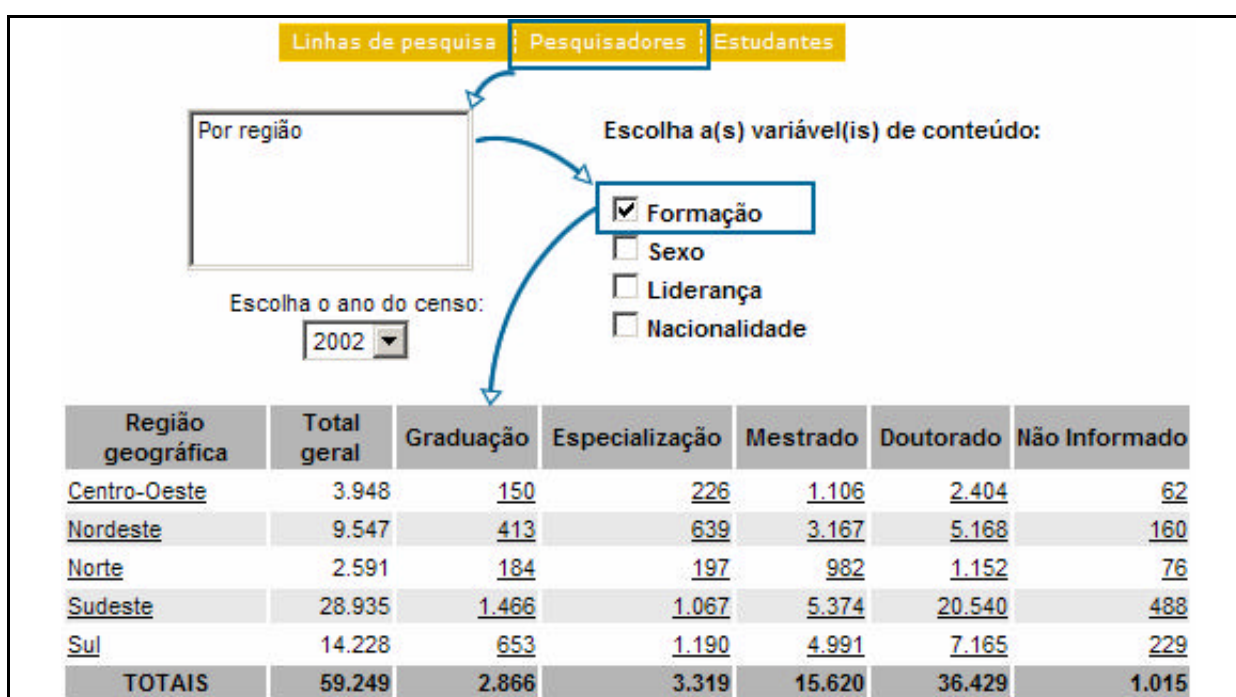


Figura 18 - Consulta da unidade de análise “pesquisadores” pela unidade de filtro “região” e unidade de conteúdo “formação”

Fonte: Da autora

Após o resultado de uma consulta de “grupos” por “região”, o usuário não precisa parar por aí, pode selecionar uma região específica e visualizar os valores referentes às “UFs” dessa região. Esse é um exemplo de operação OLAP que será visto na seção seguinte e ainda pode ser visualizado na Figura 29 no Apêndice C.

Outra opção é selecionar um número referente a “grupos de pesquisa” de uma UF específica para visualizar informações textuais relacionadas aos grupos dessa UF, como o nome do grupo, seus líderes e sua área de atuação. A Figura 19 apresenta os grupos de pesquisa para o Estado de “Santa Catarina”.

UF	Grupos G	Linhas de Pesquisa L	Pesquisadores P	Estudantes E	Técnicos T	L/G	P/G	E/G	P/L
<u>Paraná</u>	<u>1.070</u>	<u>3.279</u>	<u>6.463</u>	<u>3.854</u>	1.073	3,1	6	3,6	2
<u>Rio Grande do Sul</u>	<u>1.769</u>	<u>5.729</u>	<u>9.345</u>	<u>8.969</u>	1.494	3,2	5,3	5,1	1,6
Santa Catarina	791	2.396	4.083	3.541	489	3	5,2	4,5	1,7
TOTAIS	3.630	11.404	19.891	16.364	3.056	3,1	5,5	4,5	1,7

Grupos de pesquisa por Santa Catarina, Sul. Total: 791

- Grupo: A Condição Humana na Sociedade Contemporânea - [Lattes RedesGP](#)
 Líder: [Héctor Ricardo Leis](#)
 Líder: [Selvino Assmann](#)
 Área: Sociologia
- Grupo: A construção da identidade artística na contemporaneidade: paradigmas teóricos e práticos. - [Lattes RedesGP](#)
 Líder: [Antonio Carlos Vargas Sant'Anna](#)
 Líder: [Não se aplica](#)
 Área: Artes

Figura 19 - Apresentação das informações textuais para a unidade de análise “grupos”
 Fonte: Da autora

Em seguida, é possível ainda que o usuário visualize, por exemplo, o perfil do grupo de pesquisa e dos líderes do grupo através da Busca Textual e as redes de grupos de pesquisa através do Sistema Lattes RedesGP. A integração do Plano Tabular com esses sistemas será vista com mais detalhes em uma seção específica.

5.3.4.3 Operações OLAP

A escolha das operações OLAP a serem oferecidas no Plano Tabular foi baseada nas necessidades e nos objetivos do público-alvo do sistema. Essas operações podem ser visualizadas no Quadro 14, com exemplos.

OPERAÇÃO	EXEMPLOS
<i>Drill Down</i>	<p>Uma consulta por região é alterada para uma consulta por UF (Figura 29 do Apêndice C).</p> <p>Uma consulta pela região “Sul” é alterada para uma consulta pela UF “Santa Catarina”.</p>
<i>Drill Out</i>	<p>Apresentados os nomes dos grupos de pesquisa distribuídos pela UF “Santa Catarina”, selecionar o grupo “A Condição Humana na Sociedade Contemporânea” para visualizar seu perfil através do sistema de Busca Textual (Figura 30 do Apêndice C).</p>
<i>Slice</i>	<p>Apresentado o número de grupos de pesquisa por “ano de formação”, selecionar apenas o intervalo “2001 - 2004” para visualizar os nomes desses grupos, ou seja, a distribuição é feita na tabela de dimensão (Figura 31 do Apêndice C).</p>
<i>Filtering</i>	<p>Quando as informações são filtradas de forma a analisar somente os pesquisadores de uma determinada formação, segundo uma determinada característica, como, por exemplo, sua titulação máxima, a distribuição da informação é feita na tabela de fato (Figura 32 do Apêndice C).</p>
<i>Paging</i>	<p>Apresentado o número de grupos de pesquisa por UF e instituição, ao selecionar o número referente à instituição “UFG”, apenas os 25 primeiros resultados são apresentados na primeira página, sendo o restante acessível através do link “Próximo” (Figura 33 do Apêndice C).</p>
<i>Sort</i>	<p>Ao consultar os grupos de pesquisa por UF e instituição e selecionar a UF “Goiás” como opção para o filtro UF, as instituições (que representam as linhas) são ordenadas de forma crescente para facilitar, por exemplo, que o usuário encontre a instituição desejada rapidamente (Figura 34 do Apêndice C).</p>

Quadro 14 - Operações OLAP disponíveis no Plano Tabular

Fonte: Da autora

As operações OLAP oferecidas no Plano Tabular constituem uma limitação do sistema desenvolvido para a atual versão no DGP, e não da metodologia proposta neste trabalho.

5.3.4.4 Usabilidade

Ao se pensar no desenvolvimento do design de um site, os principais pontos a serem considerados são os objetivos e o público-alvo do sistema. É principalmente nesses pontos que o Plano Tabular se baseia.

O Plano Tabular possui um design consistente e utiliza uma nomenclatura própria que inclui as unidades de análise, unidades de filtro e unidades de conteúdo. Ele adota convenções consistentes para navegação, botões, títulos de páginas, *links*, notas, legendas, e as segue de modo consistente em todas as suas páginas.

O estado do sistema é apresentado de forma clara e visível. Existe uma homogeneidade de navegação e um padrão, que, uma vez assimilado pelo usuário, é repetido para todas as consultas que desejar realizar.

A motivação da proposta do Plano Tabular está na estrutura de navegação que foi elaborada com a possibilidade de se rever o caminho original por navegação simples (utilizando o link “Voltar”), por exemplo, ou por recomeço do ciclo 1-7, descrito no Quadro 14 (clicando em uma unidade de análise).

O sistema oferece um conjunto substancial de elementos que suportam a navegação em cada página do site, o que constitui um meio eficaz de manter o usuário constantemente orientado. A Figura 20 apresenta uma consulta básica do sistema, com alguns desses elementos demarcados.

Plano Tabular
Diretório de Grupos de Pesquisa no Brasil

BRASIL.GOV

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

GrupoStela

Grupos | Linhas de pesquisa | Pesquisadores | Estudantes | Técnicos | Produção C, T&A | Grupos/Empresas | Empresas/Grupos

Linhas de pesquisa

Escolha a variável de filtro:
Grande área/Área/Setor

Grande área
 Área
 Setor de Aplicação

Adicionar Excluir

Por região
Por grande área

Escolha o ano do censo:
2002

Consultar

Salvar consulta

Voltar

Número de linhas de pesquisa, pesquisadores, grupos e relações segundo região geográfica e grande área relacionada à linha de pesquisa.*

Região geográfica/ Grande área	Linhas de Pesquisa L	Pesquisadores P	Grupos de Pesquisa G	L/G	P/L	P/G
Sul						
Ciências Agrárias	1.676	1.938	432	3,9	1,2	4,5
Ciências Biológicas	1.466	1.720	436	3,4	1,2	3,9
Ciências da Saúde	1.725	2.304	550	3,1	1,3	4,2
Ciências Exatas e da Terra	1.382	1.464	374	3,7	1,1	3,9
Ciências Humanas	1.640	3.218	655	2,5	2	4,9
Ciências Sociais	1.070	1.839	468	2,3	1,7	3,9
Aplicadas						
Engenharias	2.037	2.392	521	3,9	1,2	4,6
Linguística, Letras e Artes	408	767	194	2,1	1,9	4
TOTAIS	11.404	15.642	3.630	3,1	1,4	4,3

* - Não há dupla contagem no número de pesquisadores na dimensão mais desagregada da tabela (ver nota técnica 2).

Figura 20 - Elementos de suporte de navegação do Plano Tabular

Fonte: Da autora

O Quadro 15 explica o objetivo de cada um desses elementos de suporte de navegação do sistema. Sua numeração é referente à figura anterior.

Nº	SUPORTE	DESCRIÇÃO
1	Grupo Stela	Link que envia o usuário para a página principal do Grupo Stela, responsável pela implementação do sistema.
2	CNPq	Link que envia o usuário para a página principal do CNPq.
3	Página Inicial	Link que envia o usuário para a página inicial do DGP.
4	Exemplos	Link que abre uma página que contém um conjunto de exemplos de como organizar em tabelas as informações de interesse do usuário.
5	Saiba Mais	Link que abre a página Saiba Mais, com informações adicionais sobre conceitos do sistema, seu funcionamento e sua estrutura.
6	Cabeçalho	Título da tabela que apresenta especificações da consulta realizada, ou seja, apresenta quais unidades de análise, unidades de filtro e unidades de conteúdo foram selecionadas.

7	Rodapé	Texto que apresenta observações técnicas com relação ao resultado da consulta. Por exemplo, diz se existe dupla contagem do número de estudantes.
8	Voltar	Link que leva o usuário para a página anterior à que ele está. Assim, o usuário pode realizar uma nova consulta com as mesmas unidades de análise e de filtro previamente selecionadas sem precisar reiniciar o ciclo 1-7.
9	Salvar consulta	Link que abre uma janela que ajuda o usuário a salvar a consulta realizada (de forma a visualizá-la posteriormente no modo off-line).
10	Notas Técnicas	Explica ao usuário especificações técnicas sobre o resultado da consulta. Cada unidade de análise possui uma nota técnica diferenciada.

Quadro 15 - Suportes de navegação do Plano Tabular

Fonte: Da autora

Além dos suportes indicados, o site ainda possui *links* para Legendas, e, ao entrar em Saiba mais ou Exemplos, oferece *links* para cada um de seus tópicos, inclusive para voltar ao modo de consulta do sistema e à página principal do DGP. Tudo isso ajuda o usuário a se manter no controle de sua pesquisa e com sentimento de liberdade e flexibilidade de navegação.

Para que o usuário saiba qual a forma de realizar sua pesquisa, o Plano Tabular oferece a ênfase no reconhecimento em vez da memorização. O desenho do sistema e as suas instruções de uso permitem tornar objetos e ações claros e visíveis. Assim, os usuários não são obrigados a memorizar informação entre vários componentes de um diálogo. Se os usuários reconhecem onde estão ao examinarem a página corrente, têm boas chances de não se perderem. O ciclo 1-7 e o layout do site garantem essa uniformidade de acesso, ou seja, as unidades de análise no topo da página, as unidades de filtro à esquerda e o resultado no centro para todas as combinações possíveis.

Um recurso técnico utilizado no Plano Tabular é o *JavaScript*. Ele pode ser utilizado, por exemplo, para verificar erros triviais no preenchimento de formulários. A informação de navegação é importante para prevenir erros ao selecionar a página de destino. O sistema de tratamento de erros obriga, por exemplo, o usuário a selecionar ao menos uma unidade de filtro, mas não permite selecionar mais de quatro.

Esses são alguns dos recursos mais importantes utilizados no Plano Tabular. Eles procuram facilitar a utilização do site e garantir que, quando o usuário o acessa, ele consiga adquirir a informação desejada com facilidade e conseqüentemente volte a utilizá-lo no futuro.

O Quadro 16 faz uma avaliação do Plano Tabular comparando-o com os critérios de usabilidade de Jakob Nielsen (1996, 1999a e 1999b).

CRITÉRIO	DESCRIÇÃO
<i>Frames</i>	Frames são utilizados no sistema com o intuito de facilitar a navegação do usuário para que ele possa sempre modificar sua consulta sem precisar reiniciá-la, modificando, por exemplo, a unidade de análise ou de filtro. No entanto, viu-se posteriormente que os <i>frames</i> não são necessários para fazer esse tipo de padronização, portanto, o Plano Tabular sem <i>frames</i> pode ser levado em consideração para um trabalho futuro.
Uso desnecessário de tecnologia	A tecnologia utilizada é confiável e consistente, não atrapalhando o usuário em sua consulta, além de proporcionar independência de plataforma e suporte aos mais variados navegadores <i>Web</i> .
Animações	Imagens animadas não são utilizadas para não sobrecarregar a visão periférica do usuário e não tirar sua concentração do conteúdo apresentado.
URLs complexas	A URL e os nomes de arquivos são legíveis e representam o conteúdo de sua informação.
Páginas órfãs	O sistema não apresenta páginas órfãs, e o usuário poderá sempre iniciar novamente sua consulta através do frame esquerdo ou superior.
Páginas longas	As opções essenciais são sempre visíveis. O usuário necessita utilizar a barra de rolagem somente para visualizar algumas consultas mais longas, que, no entanto, foram resultado de sua própria configuração.

Suporte à navegação	Para uma melhor compreensão da estrutura do espaço de informação, em qualquer página do site, o usuário tem acesso ao logo do site (que chama a página inicial) e às páginas de ajuda (Exemplos e Saiba Mais).
Cores dos <i>links</i>	Apesar de não utilizar as cores padrão de <i>links</i> (não visitados ficam azuis e visitados ficam vermelhos ou roxos), estes são sublinhados para melhor identificação, além de ser alterado o estilo ao se passar o mouse por cima dos <i>links</i> .
Informação desatualizada	A coleta dos dados do censo DGP é bienal, sendo o site atualizado também nesse período, pois esse é o propósito do site.
Tempo de <i>download</i>	As consultas mais longas foram otimizadas através da utilização de índices, tabelas de agregados e recursos disponíveis no banco de dados utilizado.
Botão “voltar”	O botão “voltar” não é desabilitado para facilitar a navegação.
Nova janela de <i>browser</i>	Só é utilizada a abertura de novas janelas nas seguintes situações: (a) <i>links</i> auxiliares ao site, como, por exemplo, “salvar consulta” e “legenda”; e (b) para acessar um sistema externo (Busca Textual e Lattes RedesGP).
Uso de GUI <i>widgets</i> não-padronizados	Não há uso de GUI <i>widgets</i> .
Biografias	Os autores envolvidos tanto com o desenvolvimento do site quanto com os grupos de pesquisa estão disponíveis no site.
Informação antiga	O usuário tem acesso desde o censo 2000 até o atual (2004).
Novas URLs	Cada versão do Plano Tabular é instalada em uma URL diferente. A partir da versão 2004, a URL será fixa, podendo-se acessar todos os censos pelo mesmo site.
Contexto de Títulos	Títulos, cabeçalhos e rodapés auxiliam na navegação do site.
Tempo de resposta do servidor	A arquitetura do sistema e os códigos foram revisados e otimizados para melhorar o tempo de resposta do servidor.
Design com aparência de propaganda	O design não possui aparência de propaganda.

Quadro 16 - Avaliação do sistema com relação aos critérios de usabilidade na *Web*

Fonte: Da autora

5.3.4.5 Integração do Plano Tabular com outros sistemas

Uma vantagem de se ter armazenada no mesmo *Data Mart* a informação textual e gerencial é a possibilidade de, a partir do resultado de uma consulta no Plano Tabular, visualizar o detalhamento textual de cada um dos elementos resultantes da consulta, sendo cada indicador visualizado minuciosamente de forma textual. Analisando-se essa forma de visualização detalhada de indicadores, é possível afirmar que esse modelo de *Data Mart* permite a busca textual de informação a partir da recuperação de indicadores. A funcionalidade de apresentação de informações textuais já está implementada no módulo Busca Textual do DGP, portanto o Plano Tabular é integrado a esse módulo para que ele também ofereça essa funcionalidade. Essa é uma das novas características do DGP em sua versão 2004. Um exemplo dessa integração pode ser visualizado na Figura 21.

Grupos de pesquisa por Santa Catarina, Sul. Total: 791

- Grupo: **A Condição Humana na Sociedade Contemporânea** [Lattes RedesGP](#)
 Líder: [Héctor Ricardo Leis](#)
 Líder: [Selvino Assmann](#)
 Área: Sociologia
- Grupo: **A construção**
 Líder: [Antonio Carlos V.](#)
 Líder: [Nôa de saica](#)
 Área: Artes



Diretório de Grupos de Pesquisa no Brasil (Censos)

Grupo de Pesquisa
A Condição Humana na Sociedade Contemporânea

Identificação	Recursos Humanos	Linhas de Pesquisa	Indicadores de RH	Indicadores de Produção
-------------------------------	----------------------------------	------------------------------------	-----------------------------------	---

Identificação

Dados básicos

Nome do grupo: A Condição Humana na Sociedade Contemporânea

Ano de formação: 1986

Data da última atualização: 27/06/2002 10:04

Líder(es) do grupo: [Héctor Ricardo Leis - hector.leis@brturbo.com](#)
[Selvino Assmann - selvinoa@terra.com.br](#)

Área predominante: Ciências Humanas; Sociologia

Instituição: Universidade Federal de Santa Catarina - UFSC

Figura 21 - Integração do Plano Tabular com a Busca Textual
 Fonte: Da autora

Através do Plano Tabular pode-se também selecionar um grupo de pesquisa ou uma instituição para visualizar sua rede de relacionamentos no contexto de grupos de pesquisa nacionais. Isso é possível devido à sua integração com o Lattes RedesGP, outro sistema de conhecimento da Plataforma Lattes já citado. A Figura 22 apresenta um exemplo dessa integração.

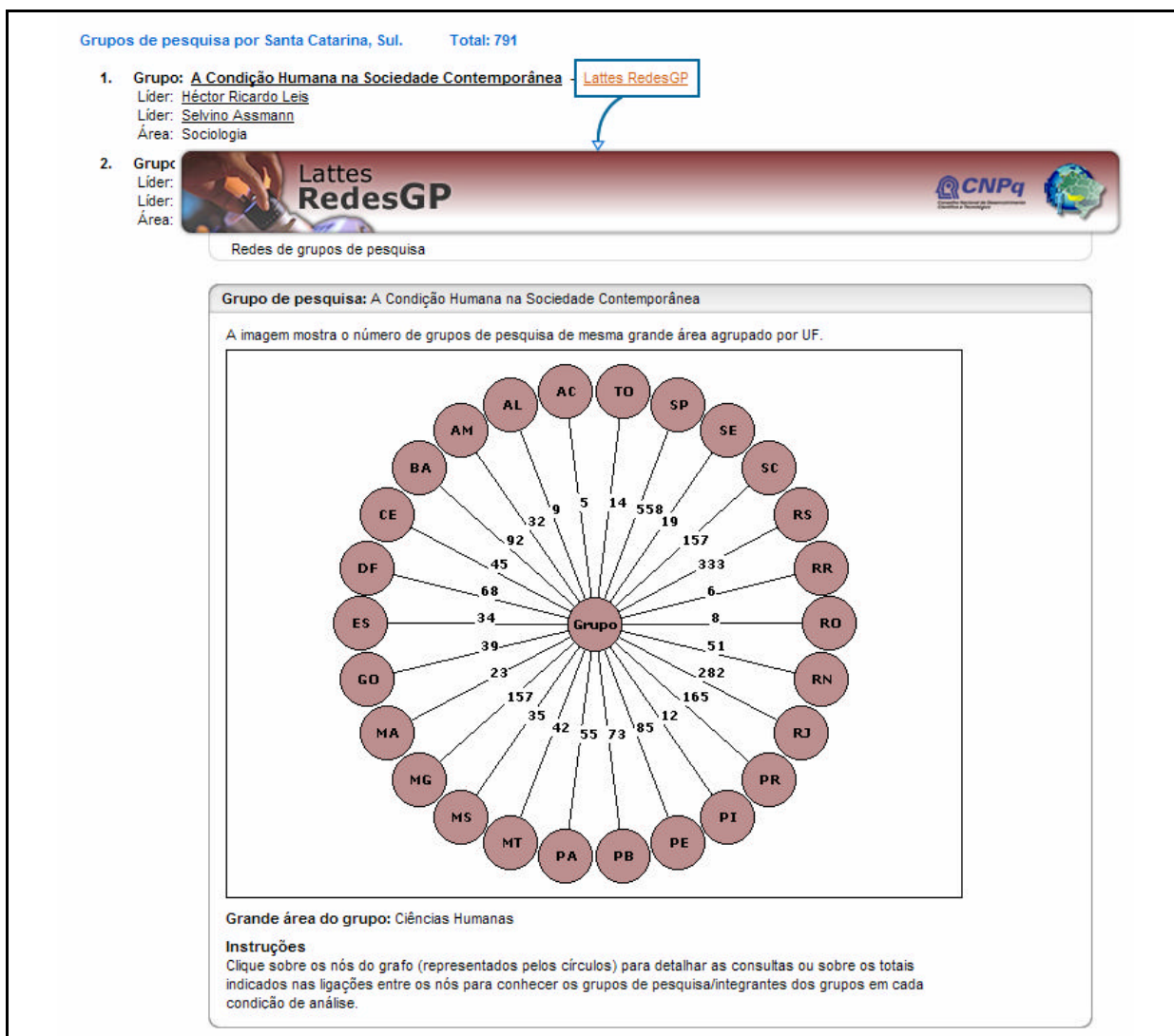


Figura 22 - Integração do Plano Tabular com o Lattes RedesGP

Fonte: Da autora

5.3.5 Vantagens da utilização da metodologia proposta no DGP

Foram muitas as alterações do Diretório de Grupos de Pesquisa ao se fazer uso da metodologia proposta e do *Data Mart* DMGRUPO. A utilização do DMGRUPO ofereceu um sistema (Plano Tabular) que pode ser utilizado em pesquisas relacionadas a todos os censos a partir de agora e possibilitou a integração de três instrumentos de informação importantes ligados a grupos de pesquisa no País (Plano Tabular, Busca Textual e Lattes RedesGP).

Com relação às características de construção e utilização do DGP, é possível fazer um comparativo do sistema em dois momentos: (1) na versão 3.0, com a construção manual de indicadores a partir de informações operacionais; e (2) na versão 2000 (4.0) em diante, com a construção de indicadores baseada no *Data Mart* DMGRUPO, como apresentado no Quadro 17.

CARACTERÍSTICA DE CONSTRUÇÃO E UTILIZAÇÃO	UTILIZANDO A BASE OPERACIONAL (v. 3.0)	UTILIZANDO A BASE DMGRUPO (v. 4.0, 5.0 e 6.0)
Forma de extração de informações	Consultas predefinidas, acessadas em páginas <i>Web</i> estáticas.	Consultas configuráveis pelo usuário, acessadas em páginas <i>Web</i> dinâmicas.
Cruzamento entre informações	Conjunto limitado de indicadores previamente definidos.	O limite é a quantidade de cruzamentos possíveis entre as unidades do <i>Data Mart</i> , sendo o usuário aquele que define o tipo de cruzamento que deseja.
Tipos de informações	Limite de informações em função do tempo gasto para geração de cada indicador.	O <i>Data Mart</i> compreende o conjunto total de informações sobre grupos de pesquisa, além de conter informações sobre a produção dos integrantes dos grupos.
Construção das consultas	Lenta, pois é necessária a intervenção de um especialista humano que gera individualmente as consultas a serem disponibilizadas.	Rápida, pois há sistemas que geram indicadores automaticamente a partir do <i>Data Mart</i> , não sendo necessário um enfoque individual para cada consulta.
Quantidade de consultas	Limitada somente às consultas que se projetou disponibilizar. Por exemplo, a unidade Grupo de Pesquisa disponibiliza somente 172 consultas.	Limitada à quantidade de cruzamentos possíveis disponibilizados pelo <i>Data Mart</i> . Por exemplo, a unidade de análise Grupo de Pesquisa disponibiliza 13.699 consultas, que representam as possibilidades de cruzamento entre as unidades relacionadas a grupos de pesquisa.
Alcance das consultas	Apenas os cruzamentos disponibilizados em páginas estáticas.	Além dos cruzamentos possíveis, se o conteúdo desejado não for encontrado no Plano Tabular, pode ser encontrado através de sua integração com a Busca Textual.

Integração entre ferramentas	Não há integração entre quaisquer ferramentas.	Há integração entre extração de indicadores e busca textual, pois o <i>Data Mart</i> compreende os dados para geração de indicadores e para busca textual. Além disso, existe a integração entre três sistemas distintos e que oferecem informações diferenciadas do <i>Data Mart</i> : Plano Tabular, Busca Textual e Lattes RedesGP.
Tempo gasto para a disponibilização dos resultados aos usuários	Cerca de um ano (tempo necessário para a análise individual de cada consulta e a construção das páginas estáticas).	Cerca de três meses (tempo necessário para a construção do <i>Data Mart</i> e das ferramentas para apresentação dos resultados). Com o <i>Data Mart</i> do DGP, o Plano Tabular estará pronto assim que os dados do censo atual forem armazenados no <i>Data Mart</i> .

Quadro 17 - Comparação de atividades executadas para apresentação de dados no DGP
 Fonte: Adaptado de GONZAGA et al. (2004)

5.4 Considerações finais

Este capítulo apresentou um conjunto de instrumentos para a extração de informações gerenciais, com enfoque no Plano Tabular. Foi evidenciada a importância da construção do *Data Mart* como repositório das informações a serem analisadas com a modelagem orientada ao assunto do negócio das análises, em comparação à modelagem, segundo o processo do negócio, que é a lógica de construção de sistemas de informação operacional.

Com relação à agilidade e eficiência que o *Data Mart* proporcionou, é possível citar as seguintes características particulares:

- ✍️ agilizou a construção de instrumentos para a extração de indicadores e informação textual;
- ✍️ agilizou o processamento desses instrumentos;
- ✍️ possibilitou a extração de informações com maior velocidade devido à estrutura de informações contidas no *Data Mart*;

✍ possibilitou uma rápida disponibilização dos instrumentos de informação para a comunidade a partir do término da coleta de dados em cada censo da pesquisa brasileira.

Com o uso do *Data Mart* no DGP, foi possível realizar a integração de suas informações com as de outras bases de dados. Essa integração favorece a qualidade da informação, uma vez que se eliminam dados redundantes e possivelmente inconsistentes, bem como favorece a oportunidade de criar novas funções, eliminar trabalho sem valor agregado e construir novas informações e conhecimentos.

Com relação aos instrumentos de recuperação e apresentação de informações resultantes do *Data Mart*, vale destacar a ferramenta de extração de indicadores das principais unidades de informação do DGP, o Plano Tabular. Essa ferramenta, além de proporcionar ao usuário final um sistema de consulta amigável, não técnico e eficiente, oferece milhares de cruzamentos de informações sobre os vários tópicos presentes na base do DGP. Além disso, sua integração com os sistemas Busca Textual e Lattes RedesGP permite que o usuário realize análises valiosas, consultando as unidades de informação da Plataforma Lattes através de diferentes pontos de vista.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este capítulo está organizado em duas seções. Na primeira são citadas as conclusões referentes à proposta de trabalho e aos resultados alcançados com sua aplicação. Na segunda seção são apresentadas as recomendações bem como sugestões relevantes para o prosseguimento de estudos e pesquisas na área abordada.

6.1 Conclusões

Este trabalho propôs uma metodologia para a construção de sistemas de análises multidimensionais através da tradução dos requisitos de análise do usuário de uma organização, utilizando-se de conceitos de sistemas de informação e de tecnologias de armazenamento, recuperação e apresentação de informações.

Para definição da metodologia, estudou-se a arquitetura de um sistema Business Intelligence e seus componentes – módulo ETL, repositórios de dados e técnicas de *front-end* –, já que a proposta de trabalho estava direcionada para o desenvolvimento de sistemas de recuperação e apresentação de informações utilizando as técnicas OLAP e a estrutura de armazenamento de dados baseada em repositório de dados de *Data Warehouse*.

Para traduzir os requisitos dos analistas de negócio bem como os requisitos de sistemas DW/OLAP, elaborou-se uma nomenclatura que fizesse a relação entre os conceitos técnicos e os de negócio. Essa nomenclatura conseguiu interligar conceitos divergentes, passíveis de serem compreendidos por parte de todos os

integrantes do sistema (usuários de negócio, analistas de sistema e sistemas DW/OLAP).

A grande vantagem da nomenclatura desenvolvida é a sua característica não técnica, própria, de fácil assimilação e adequada para todos os tipos de usuários, permitindo a tradução dos requisitos de análise do usuário em processos de recuperação da informação. Com a nomenclatura proposta, os analistas de negócios e de sistemas comunicam-se através da mesma linguagem, o que facilita o levantamento de requisitos. O fato de ser uma nomenclatura de fácil assimilação permite que qualquer usuário a entenda, podendo, portanto, ser utilizada em uma ferramenta de *front-end* de uso geral.

Assim, para o analista de sistemas, a identificação das unidades de informação fundamentais para a construção de indicadores, obtidos no levantamento de requisitos e utilizados no processo de decisão, é facilmente traduzida nos termos técnicos do domínio da técnica de *Data Warehouse*. Para o usuário de negócio, os sistemas de *front-end* que utilizam a nomenclatura proposta permitem moldar sua própria consulta de acordo com suas necessidades de análises.

A metodologia proposta foi aplicada no Diretório de Grupos de Pesquisa no Brasil (site que apresenta informações sobre os grupos de pesquisa do País) mais especificamente no desenvolvimento de um *Data Mart* e de um sistema de extração de indicadores denominado Plano Tabular. O *Data Mart* resultante foi utilizado não somente no Plano Tabular, como em outros instrumentos da Plataforma Lattes. Com o uso do *Data Mart* no DGP, foi possível fazer a integração entre alguns desses instrumentos (Plano Tabular, Busca Textual e Sistema Lattes RedesGP), havendo a apresentação de forma clara e objetiva das informações sobre a pesquisa nacional, o que vem a auxiliar o processo de tomada de decisão operacional e gerencial

relacionada a C&T no País. O rápido desenvolvimento, a fácil integração dos diversos instrumentos citados e a aceitação pelos usuários do sistema expressam claramente as vantagens da utilização da metodologia proposta.

Tal metodologia é aplicável em diferentes projetos de informação em que se utilizam as tecnologias DW/OLAP, sendo o caso real aqui apresentado um exemplo de sua aplicação.

6.2 *Trabalhos futuros*

Na linha de futuros desenvolvimentos, há a possibilidade de se aplicar a metodologia proposta em um sistema off-line, visto que a utilização das técnicas OLAP pode vir a ser mais flexível e oferecer mais funcionalidades.

Com relação aos conceitos de usabilidade na *Web*, aplicados a sistemas OLAP que façam uso da metodologia proposta, pode-se realizar um estudo aprofundado com o intuito de melhorar ainda mais a facilidade de utilização desses sistemas e de melhor explorar as potencialidades dos sistemas OLAP orientados a usuários de negócio.

Já para enriquecer a própria metodologia proposta e sua aplicação, sugere-se que o sistema resultante ofereça mais operações OLAP. No caso real descrito, as operações oferecidas são limitadas, pois não era de necessidade inicial, no entanto, novas operações podem se tornar relevantes nas análises dos usuários.

Outra sugestão é criar uma representação gráfica para a nomenclatura proposta. Ou seja, criar um modelo conceitual para a construção de uma linguagem gráfica para a nomenclatura proposta, como existe para o modelo Entidade-Relacionamento.

A proposta do trabalho foi elaborada pensando-se na aplicação da metodologia em sistemas OLAP baseados em estruturas de dados de *Data Warehouse*. Para ferramentas sem esse suporte informacional deve-se adaptar a metodologia à estrutura escolhida.

REFERÊNCIAS BIBLIOGRÁFICAS

AHMAD, A.; AHMAD, A.; YAHAYA, J. Web based on-line analytical processing (WOLAP) model. In: NATIONAL CONFERENCE ON COMPUTER GRAPHICS AND MULTIMEDIA, 2002, Malasia. Disponível em: <<http://staf.uum.edu.my/amran/CogRamm2002.pdf>>. Acesso em: 28 jan. 2004

ALALOUF, C. **Hybrid OLAP**: The best of both worlds. Canadá: Speedware Corporation Inc., 1997. Disponível em: <<http://www.dmreview.com>>. Acesso em: 9 maio 2003.

BALANCIERI, R. **Análise de redes de pesquisa em uma plataforma de gestão em ciência e tecnologia**: uma aplicação à Plataforma Lattes. 2004. 127 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção – Universidade Federal de Santa Catarina, Florianópolis, 2004. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/10749.pdf>>. Acesso em: 10 jan. 2005.

BONIFACIO, A. S. **Ontologias e consulta semântica**: uma aplicação ao caso Lattes. 2002. 85f. Dissertação (Mestrado em Computação) – Programa de Pós-Graduação em Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. Disponível em: <<http://www.uel.br/pessoal/ailton/Trabalhos/Disserta%C3%A7ao%20de%20Mestrado-Ailton-Final.pdf>>. Acesso em: 4 jan. 2003.

BOVO, B. A. **Um método de tradução de fontes de informação em um formato padrão que viabilize a extração de conhecimento por meio de Link Analysis e Teoria dos Grafos**. 2004. 96 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2004. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/10644.pdf>>. Acesso em: 10 jan. 2005.

CNPq. **Diretório de Grupos de Pesquisa no Brasil**: Censo 2002. Brasília. Disponível em: <<http://lattes.cnpq.br/censo2002/>>. Acesso em: 10 out. 2004.

CNPq. **Plataforma Lattes**: Diretório de Grupos de Pesquisa. Disponível em: <<http://www.cnpq.br/plataformalattes/dgp/index.htm>>. Acesso em: 5 jan. 2005.

CODD, E. F.; CODD, S. B.; SALLEY, C. T. **Providing OLAP (On-line Analytical Processing) to user-analysts**: An IT mandate. [Artigo de Codd & Associates Technical, para Arbour Software, hoje Hyperion Solutions]. EUA, 1993.

CÔRTEZ, S. **BI, Data Warehouse e Data Mining**: como a tecnologia aumenta a inteligência do negócio. PUC-Rio, 2002. Palestra “Casos, soluções e tendências em Inteligência de Negócios e Gestão do Conhecimento”.

Local: Auditório da FIRJAN - Rio de Janeiro - RJ). Disponível em:
<<http://www.tec.com.br/BI.asp>>. Acesso em: 15 dez. 2002.

CYBIS, W. de A. **Engenharia de usabilidade**: uma abordagem ergonômica. LABIUTIL - Laboratório de Utilizabilidade de Informática, UFSC, Florianópolis, 2003. Disponível em: <<http://www.labiutil.inf.ufsc.br/>>. Acesso em: 4 jun. 2003.

DAVENPORT, T.; PRUSAK, L. **Conhecimento empresarial**. Rio de Janeiro: Campus, 1999.

DMREVIEW. **Glossary: Business Intelligence (BI)**. Disponível em:
<<http://www.dmreview.com/resources/glossary.cfm?keywordId=B>>. Acesso em: 2 mar. 2005.

EARL, M. J. **Experiences in strategic information systems planning**. MIS QUARTELY, Mar. 1993.

GONZAGA, T et al. Instrumentos de Informação para extração de indicadores e de informação textual. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 24., 2004, Florianópolis. ABEPRO, 2004. p. 288-289.

GUIMARÃES, R. **Avaliação e fomento de C&T no Brasil**: propostas para os anos 90. Brasília: MCT/CNPq, 1994.

GUIMARÃES, R. et al. **A pesquisa no Brasil** – Perfil da pesquisa no Brasil e hierarquização dos grupos de pesquisa a partir dos dados do Diretório dos Grupos de Pesquisa no Brasil. 1999.

INMON, W. H. **Building the Data Warehouse**. 3rd ed, New York: John Wiley & Sons, 2002.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **Ergonomic requirements for office work with Visual Display Terminals (VDTs)** - Part II: Guidance on usability, ISO/DIS 9241-11,1995.

KIMBALL, R.; REEVES, L.; ROSS, M.; THORNTHWAITE, W. **The Data Warehouse lifecycle toolkit**: Expert Methods for Designing, Developing and Deploying Data Warehouses. New York: John Wiley & Sons, 1998.

KIMBALL, R.; ROSS, M. **The Data Warehouse toolkit**: The Complete Guide to Dimensional Modeling. New York: John Wiley & Sons, 2002.

KIMBALL, R. **Drilling Down, Up and Across**: understanding the vocabulary of navigating dimensions. DBMS on-line. San Francisco: Miller Freeman, Inc., Mar.1996. Disponível em: <<http://www.dbmsmag.com/9603d05.html>>. Acesso em: 02 mar. 2005.

KLINE, K.; GOULD, L.; ZANEVSKYA. **Transact-SQL programming** - Sample Chapter 1. DatabaseProNews. EUA: iEntry, Inc. May 2003. Disponível em: <<http://databasepronews.com/2003/0528.html>>. Acesso em: 27 dez. 2004.

LAUDON, K.; LAUDON, J. **Management Information Systems**: Organization and Technology. EUA: Macmillan Publishing Company, 1996.

_____. **Management information systems**: New Approaches to Organization & Technology. New Jersey: Prentice Hall, 1999.

MACHADO, F. N. R. **Projeto de Data Warehouse**: uma visão multidimensional. São Paulo: Érica, 2000.

MALHORTA, Y. **Tools@Work**: Deciphering the knowledge management hype. Journal of Quality and Participation, special issue on Learning and Information Management. v. 21, n. 4, p. 58-60, 1998.

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

MARTINS, S. R. **Geração automática de textos em plataformas de governo eletrônico**: um estudo de caso na Plataforma Lattes. 2004. 175 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2004. Disponível em: <<http://www.ppgep.ufsc.br/>>. Acesso em: 15 jan. 2005.

MOTTA, R. **Desenvolvimento de sistemática para análise de sistemas de informação**: uma aplicação para gestão da manutenção. 2001. 97 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal Santa Catarina, Florianópolis, 2001. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/4313.pdf>>. Acesso em: 10 fev. 2004.

NIELSEN, J. **Top ten mistakes in Web design**. Jakob Nielsen's Website, 1996. Disponível em: <<http://www.useit.com/alertbox/9605.html>>. Acesso em: 5 jun. 2003.

_____. **"Top ten mistakes" revisited three years later**. Jakob Nielsen's Website, 1999a. Disponível em: <<http://www.useit.com/alertbox/990502.html>>. Acesso em: 5 jun. 2003.

_____. **The top ten new mistakes of Web design**. Jakob Nielsen's Website, 1999b. Disponível em: <<http://www.useit.com/alertbox/990530.html>>. Acesso em: 5 jun. 2003.

O'BRIEN, J. A. **Sistemas de informação e as decisões gerenciais na era da Internet**. São Paulo: Saraiva, 2003.

PACHECO, R. C. dos S. **Uma metodologia de desenvolvimento de plataformas de governo para geração e divulgação de informações e de conhecimento**. Florianópolis: Grupo Stela, 2003.

PAN, S. L.; LEIDNER, D. E. **Bridging communities of practice with information technology in pursuit of global knowledge sharing**. The Journal of Strategic Information Systems. Article in Press, Corrected Proof. © 2002 Elsevier Science.

PARENTE, D. **OLAP**. DW Brasil, 2000. Disponível em: <<http://www.dwbrasil.com.br/html/olap.html>>. Acesso em: 25 abr. 2003.

PARRINI, E. **Gestão do conhecimento no suporte à decisão em ambiente OLAP**. 2002. 157 f. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2002. Disponível em: <<http://genesis.nce.ufrj.br/dataware/GestaoConhecimento/Teses/Esther/Tese-Esther.pdf>>. Acesso em: 26 abr. 2003.

POE, V.; KLAUER, P.; BROBST, S. **Building a Data Warehouse for decision support**. New Jersey: Prentice Hall, 1998.

REVISTA PLATAFORMA LATTES: revista de informações em C&T do Brasil. Florianópolis: Grupo Stela, 2002.

RODRIGUES, J. dos R. G.; NUNES, P. M. **O uso do Balanced Scorecard como núcleo de Business Intelligence**: um poderoso habilitador da excelência das decisões sobre temas qualitativos e estratégicos da organização moderna. UFMG, Belo Horizonte, 2001. Disponível em: <<http://www.fiber.com.br/literatura/artigos/artigo01.html>>. Acesso em: 25 abr. 2003.

SELL, D. **Uma arquitetura para distribuição de componentes tecnológicos de sistemas de informação baseados em Data Warehouse**. 2001. 100 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2001. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/3816.pdf>>. Acesso em: 10 fev. 2004.

STUDER, R.; DEKER, S.; FENSEL, D.; STAAB, S. Situation and prospective of knowledge engineering. In: CUENA, J. et al. (Ed.). **Knowledge Engineering and Agent Technology**. IOS Series on Frontiers in Artificial Intelligence and Applications. IOS Press, 2000. Disponível em: <http://wwwdb.stanford.edu/%7Estefan/paper/2000/ios_2000.pdf>. Acesso em: 2 jan. 2003.

TAIT, T. F. C. **Uma avaliação do processo de planejamento estratégico de sistemas de informação em empresas do mercado brasileiro e uma proposta simplificada de arquitetura de sistemas de informação**. 1994. Dissertação (Mestrado em Ciência da Computação), Universidade Federal de São Carlos, São Paulo, 1994.

TAIT, T. F. C.; PACHECO, R. C. S. An information systems architecture model for the public sector: from Data Processing State Enterprises. In: INTERNATIONAL CONFERENCE OF THE INFORMATION RESOURCES MANAGEMENT ASSOCIATION, 11., 2000, Alaska, EUA. Idea Group Publishing, 2000, p. 1196-1198. ISBN 1-878289-84-5.

THOMSEN, E. **OLAP**: construindo sistemas de informação multidimensionais. Rio de Janeiro: Campus, 2002.

TISSOT, H. **Proposta para documentação de requisitos em projetos de Data Warehouse**. 2004. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2004. Disponível em: <<http://www.ppgep.ufsc.br/>>. Acesso em: 15 jan. 2005.

VIEIRA, M. V. G. **Proposta de um modelo de sistema para monitoramento da educação a distância em cursos de nível superior utilizando ferramentas de tecnologia da informação**. 2001. 102 f. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2001. Disponível em: <<http://teses.eps.ufsc.br/defesa/pdf/7376.pdf>>. Acesso em: 10 fev. 2004.

WG Systems. **Arquitetura de um sistema de BI**. Disponível em: <http://www.wgsystems.com.br/bi/bi_arquitetura.htm>. Acesso em: 12 fev. 2003.

ZACHMAN, J. A. A framework for information systems architecture. **IBM System Journal**, v. 26, n. 3, p. 276-285, 1987.

APÊNDICE B – Evolução do Plano Tabular



Figura 23 - Plano Tabular em sua versão 2.0

Fonte: CNPq (2005)

Endereço http://www.cnpq.br/gpesq2/tab1/gd_area/1_gd1.htm Ir Links >>

CNPq - DIRETÓRIO DOS GRUPOS DE PESQUISA NO BRASIL - 1995
 TABELA 1.I - DISTRIBUIÇÃO DOS GRUPOS, LINHAS DE PESQUISA E RECURSOS HUMANOS ENVOLVIDOS, POR GRANDE ÁREA DO CONHECIMENTO, SEGUNDO AS GRANDES REGIÕES E UNIDADES DA FEDERAÇÃO
Grande Área: CIÊNCIAS AGRÁRIAS (1)

GRANDES REGIÕES UNIDADES DA FEDERAÇÃO	GRUPOS	LINHAS DE PESQUISA	RECURSOS HUMANOS			
			TOTAL	PESQUISADORES (2)	ESTUDANTES E ESTAGIÁRIOS (3)	TÉCNICOS (2)
CENTRO-OESTE	70	230	934	378	317	239
Distrito Federal	28	88	280	135	71	74
Goiás	22	61	345	100	166	79
Mato Grosso	11	41	140	88	34	18
Mato Grosso do Sul	9	40	169	55	46	68
NORDESTE	101	311	1519	536	690	293
Alagoas	2	5	20	4	14	2
Bahia	10	30	192	102	53	37
Ceará	28	95	349	96	177	76
Maranhão	1	2	3	3	-	-
Paraíba	16	54	312	85	173	54

Figura 24 - Exemplo de consulta no Plano Tabular versão 2.0

Fonte: CNPq (2005)

Endereço http://www.cnpq.br/gpesq3/plano_tabular/plano_tabular.htm Ir Links >>



fontes de informação

Diretório dos Grupos de Pesquisa no Brasil
Versão 3.0

Plano Tabular


página inicial página anterior

1. Informações Gerais
2. Súmula Estatística
3. Plano Tabular
4. Busca Textual
5. Hierarquização
6. Anexos

Variáveis	Total	Grandes áreas do conhecimento predominantes dos grupos	Áreas do Conhecimento predominantes dos grupos	Setores de atividade relacionados a pelo menos uma linha de pesquisa dos grupos
01 Grupos, linhas de pesquisa, pesquisadores 1/, estudantes e estagiários	Tabela 01	Tabela 02	Tabela 03	Tabela 04
02 Pesquisadores 2/ por sexo e titulação máxima	Tabela 05	Tabela 06	X	X
03 Estudantes e estagiários por nível de atividade	Tabela 07	Tabela 08	X	X

Figura 25 - Plano Tabular em sua versão 3.0
Fonte: CNPq (2005)

Endereço http://gpesq98.cnpq.br:9750/dcd_grpesq/owa/cons_total Ir Links >>



fontes de informação

Diretório dos Grupos de Pesquisa no Brasil
Versão 3.0

Plano Tabular

página inicial página anterior

1. Grupos, Linhas de Pesquisa, Pesquisadores, Estudantes e Estagiários

Regiões e Brasil

Região	Grupos	Linhas de Pesquisa	Pesquisadores	Estudantes e Estagiários	Técnicos
CENTRO-OESTE	349	1,041	2,104	5,740	558
NORDESTE	987	2,814	4,830	14,373	1,352
NORTE	153	378	722	1,661	343
SUDESTE	5,660	17,293	26,793	70,497	9,117
SUL	1,483	4,184	7,397	23,425	1,876
BRASIL	8,632	25,710	41,846	115,696	13,246

administrador do site

Figura 26 - Exemplo de consulta do Plano Tabular versão 3.0
Fonte: CNPq (2005)



Figura 27 - Plano Tabular em sua versão 4.0

Fonte: CNPq (2005)



Figura 28 - Plano Tabular em sua versão 5.0

Fonte: CNPq (2005)

APÊNDICE C – Operações OLAP do Plano Tabular

Região geográfica	Grupos G	Linhas de Pesquisa L	Pesquisadores P	Estudantes E	Técnicos T	L/G	P/G	E/G	P/L
Centro-Oeste	809	2.831	5.425	3.583	1.035	3,5	6,7	4,4	1,9
Nordeste	2.274	7.962	13.415	11.397	2.142	3,5	5,9	5	1,7
Norte	590	2.113	3.773	2.049	842	3,6	6,4	3,5	1,8
Sudeste	7.855	26.163	41.346	34.131	11.305	3,3	5,3	4,3	1,6
Sul	3.630	11.404	19.891	16.364	3.056	3,1	5,5	4,5	1,7
TOTAIS	15.158	50.473	83.850	67.524	18.380	3,3	5,5	4,5	1,7

UF	Grupos G	Linhas de Pesquisa L	Pesquisadores P	Estudantes E	Técnicos T	L/G	P/G	E/G	P/L
Paraná	1.070	3.279	6.463	3.854	1.073	3,1	6	3,6	2
Rio Grande do Sul	1.769	5.729	9.345	8.969	1.494	3,2	5,3	5,1	1,6
Santa Catarina	791	2.396	4.083	3.541	489	3	5,2	4,5	1,7
TOTAIS	3.630	11.404	19.891	16.364	3.056	3,1	5,5	4,5	1,7

Figura 29 - Drill Down

Fonte: Da autora

Grupos de pesquisa por Santa Catarina, Sul. Total: 791

- Grupo: [A Condição Humana na Sociedade Contemporânea - Lattes RedesGP](#)
Líder: Héctor Ricardo Leis - [A Condição Humana na Sociedade Contemporânea](#)
Líder: Selvino Assmann
Área: Sociologia
- Grupo: [A construção](#)
Líder: Antonio Carlos V.
Líder: Não se aplica
Área: Artes



Diretório de Grupos de Pesquisa no Brasil (Censos)

Grupo de Pesquisa
A Condição Humana na Sociedade Contemporânea

Identificação	Recursos Humanos	Linhas de Pesquisa	Indicadores de RH	Indicadores de Produção
Identificação				
Dados básicos				
Nome do grupo: A Condição Humana na Sociedade Contemporânea				
Ano de formação: 1996				
Data da última atualização: 27/06/2002 10:04				
Líder(es) do grupo: Héctor Ricardo Leis - hector.leis@brturbo.com Selvino Assmann - selvino@terra.com.br				
Área predominante: Ciências Humanas; Sociologia				
Instituição: Universidade Federal de Santa Catarina - UFSC				

Figura 30 - Drill Out

Fonte: Da autora

Ano de formação	Grupos G	Linhas de Pesquisa L	Pesquisadores P	Estudantes E	Técnicos T	L/G	P/G	E/G	P/L
[-,1980]	775	3.395	4.851	4.309	2.388	4,4	6,3	5,6	1,4
[1981,1984]	388	1.674	2.403	2.504	1.087	4,3	6,2	6,5	1,4
[1985,1988]	711	2.852	4.379	4.383	1.371	4	6,2	6,2	1,5
[1989,1992]	1.397	5.479	8.283	8.108	2.291	3,9	5,9	5,8	1,5
[1993,1996]	2.512	9.045	14.267	14.089	3.455	3,6	5,7	5,6	1,6
[1997,2000]	5.048	15.968	27.548	20.486	5.060	3,2	5,5	4,1	1,7
[2001,2004]	4.327	12.060	22.119	13.645	2.728	2,8	5,1	3,2	1,8
TOTAIS	15.158	50.473	83.850	67.524	18.380	3,3	5,5	4,5	1,7

Grupos de pesquisa por [2001,2004]. Total: 4.327

- Grupo: Patologia bucal** - [Lattes RedesGP](#)
Líder: [Marize Raquel Diniz da Rosa](#)
Líder: [Rita de Cássia Cavalcanti Gonçalves](#)
Área: Odontologia
- Grupo: Bioconversão, metabolismo e biossíntese** - [Lattes RedesGP](#)
Líder: [Valéria de Oliveira](#)
Líder: [Mariângela Fontes Santiago](#)
Área: Farmácia
- Grupo: Biodiversidade do Cerrado: Estudo da Ciência e Tecnologia de Alimentos** - [Lattes RedesGP](#)
Líder: [Eduardo Ramirez Asquieri](#)
Líder: [Elaine Meire de Assis](#)
Área: Ciência e Tecnologia de Alimentos

Figura 31 - Slice

Fonte: Da autora

Escolha a(s) variável(is) de conteúdo:

Formação
 Sexo
 Liderança
 Nacionalidade

Região geográfica	Total geral	Graduação	Especialização	Mestrado	Doutorado	Não Informado
<u>Centro-Oeste</u>	3.948	150	226	1.106	2.404	62
<u>Nordeste</u>	9.547	413	639	3.167	5.168	160
<u>Norte</u>	2.591	184	197	982	1.152	76
<u>Sudeste</u>	28.935	1.466	1.067	5.374	20.540	488
<u>Sul</u>	14.228	653	1.190	4.991	7.165	229
TOTAIS	59.249	2.866	3.319	15.620	36.429	1.015

Figura 32 - Filtering

Fonte: Da autora

Grupos de pesquisa por Goiás, UFG.		Total: 126
1.	Grupo: <u>Bioconversão, metabolismo e biossíntese</u> - Lattes RedesGP Líder: Valéria de Oliveira Líder: Mariângela Fontes Santiago Área: Farmácia	
2.	Grupo: <u>Biodiversidade do Cerrado: Estudo da Ciência e Tecnologia de Alimentos</u> - Lattes RedesGP Líder: Eduardo Ramirez Asquieri Líder: Elaine Meire de Assis Área: Ciência e Tecnologia de Alimentos	
⋮		
25.	Grupo: <u>Economia e Administração Rural</u> - Lattes RedesGP Líder: Jose Ferreira de Noronha Líder: Não se aplica Área: Agronomia	
Grupos de pesquisa: 1 a 25		Próximo

Figura 33 - Paging

Fonte: Da autora

UF/ Instituição	Grupos G	Linhas de Pesquisa L	Pesquisadores P	Estudantes E	Técnicos T	L/G	P/G	E/G	P/L
Goiás									
EMBRAPA	7	44	112	2	46	6,3	16	0,3	2,5
FURNAS	9	12	29	1	20	1,3	3,2	0,1	2,4
UCG	38	162	408	244	72	4,3	10,7	6,4	2,5
UEG	13	35	68	7	1	2,7	5,2	0,5	1,9
UFG	126	477	886	710	188	3,8	7	5,6	1,9
ULBRA	5	15	25	5	0	3	5	1	1,7
UNIVERSO	1	4	3	0	0	4	3	0	0,8
TOTAIS	199	749	1.531	969	327	3,8	7,7	4,9	2

Figura 34 - Sort

Fonte: Da autora