

Carlos Alberto Fugimura

**Desenvolvimento e Integração de Metadados**

Monografia apresentada à Escola Politécnica da Universidade de São Paulo para conclusão do Curso de Especialização em Engenharia de Software MBA-USP

Área de Concentração :

Engenharia de Software, Banco de Dados, Data Warehouse

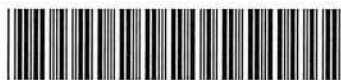
Orientador :

Prof. Doutor

Jorge Rady de Almeida Júnior

São Paulo

2005



31500020061

3

FICHA CATALOGRÁFICA

M2005 BQ

**Fugimura, Carlos Alberto**

**Desenvolvimento e Integração de Metadados**

Município de São Paulo, São Paulo. São Paulo, 2005. 50p

Dissertação (Monografia) – Escola Politécnica da Universidade de São Paulo. PECE –  
Programa de Educação Continuada em Engenharia de Software

1825989

Aos meus pais e à minha futura esposa que sempre tem me apoiado nos momentos mais difíceis e têm sido a grande razão e incentivo de meu aperfeiçoamento pessoal e técnico. Desejo que este trabalho seja prova e resultado do sucesso do apoio de todos.

## **AGRADECIMENTOS**

Ao orientador Prof. Dr. Jorge Rady de Almeida Júnior pelas diretrizes seguras e inigualável compreensão.

À Coordenadora do curso Selma Melnikof pelo estímulo e apoio.

A todos que, direta ou indiretamente, colaboraram na execução deste trabalho.

## RESUMO

A grande importância representada pelos metadados está sendo, novamente, reconhecida, devido à pressão imposta pelas necessidades das comunidades de negócios formadas por “trabalhadores do conhecimento”, envolvidos nas iniciativas de Data Warehouse.

No passado, metadados eram de interesse exclusivo dos técnicos envolvidos com a Tecnologia da Informação. Porém, com o advento do Data Warehouse, os metadados tornaram-se essenciais para os “trabalhadores do conhecimento”, pois eles permitem contextualizar a informação para otimizar a tomada de decisão, bem como entender dezenas ou centenas de terabytes de dados históricos armazenados no Data Warehouse, ao longo de vários anos.

No decorrer do trabalho são abordados os conceitos de metadados, o histórico da evolução da informação, o contexto dos metadados nas organizações, os padrões que existem no mercado, categorias de metadados, proposta de um processo de solução de metadados e gestão desta solução.

## **ABSTRACT**

The metadata rebirth is happening due to the pressure imposed by the businesses communities needs formed by knowledge workers, involved in Data Warehouse initiatives.

In the past, metadata was a matter of the IT technician's exclusive interest. However, with the introduction of the Data Warehouse Metadata, metadata became essential for the knowledge workers, because they provide the information background to optimize the decision process, as well as to understand dozens or hundreds of terabytes of historical data stored in the Data Warehouse, along several years.

In this work it is presented the main metadata concepts, a brief history of information evolution, the metadata context in organizations, the existing standards in market, metadata categories. Finally it is presented a proposal of a metadata solution process and its correspondent management.

## SUMÁRIO

|  |    |
|--|----|
| 1. Introdução .....  | 16 |
| 1.1. Objetivos.....  | 17 |
| 1.2. Estrutura do Trabalho .....                               | 17 |
| 2. Perspectiva da Informação .....                             | 19 |
| 2.1. Definição da Informação.....                              | 21 |
| 2.2. Evolução da Informação .....                              | 21 |
| 2.3. O papel da informação .....                               | 23 |
| 2.4. O Contexto das informações nas organizações atuais .....  | 23 |
| 2.5. Compartilhamento de informações e redundância.....        | 27 |
| 2.6. Suportando informação intra-organizacional.....           | 27 |
| 2.7. Modelagem de Dados .....                                  | 29 |
| 2.8. Data Warehouse.....                                       | 30 |
| 3. Metadados - Visão Geral .....                               | 33 |
| 3.1. Entendendo metadados através de exemplos.....             | 36 |
| 3.2. Tipos de Metadados .....                                  | 38 |
| 3.3. Metadados voltados para o Projeto do Data Warehouse ..... | 41 |
| 3.4. Padrões de Metadados.....                                 | 43 |
| 3.4.1. Meta-modelos .....                                      | 49 |
| 3.4.2. O Open Information Model[OIM] .....                     | 50 |
| 3.4.3. O Common Warehouse Model .....                          | 52 |

|         |  |    |
|---------|--|----|
| 3.4.4.  | Troca de Metadados.....  | 55 |
| 3.4.5.  | Processos orientados a metadados embutidos.....                    | 56 |
| 3.4.6.  | Linguagens baseadas em padrões .....                               | 56 |
| 3.4.7.  | XMI.....   | 57 |
| 3.4.8.  | Arquiteturas e Frameworks.....                                     | 59 |
| 3.4.9.  | CORBA.....   | 59 |
| 3.4.10. | Fornecedores .....   | 60 |
| 3.5.    | Fontes de metadados.....   | 61 |
| 3.6.    | Problemas encontrados devido à falta de Metadados .....            | 62 |
| 3.7.    | A Importância dos Metadados para a Organização .....               | 65 |
| 4.      | Proposta de um plano para criação de uma solução de metadados..... | 67 |
| 4.1.    | Tipos de Solução.....  | 67 |
| 4.1.1.  | Tipo 1 – Primeira vez armazenando Metadados.....                   | 67 |
| 4.1.2.  | Tipo 2 – Suporte ao DW .....                                       | 68 |
| 4.1.3.  | Tipo 3 – A solução de Metadados.....                               | 69 |
| 4.2.    | Identificar os beneficiários dos Metadados.....                    | 69 |
| 4.3.    | Listar os Requisitos de Metadados por Beneficiário. ....           | 69 |
| 4.4.    | Determinar a fonte de metadados. ....                              | 69 |
| 4.5.    | Planejar a Arquitetura Preliminar e o Escopo dos Metadados.....    | 71 |
| 4.6.    | Recategorizar os Metadados. ....                                   | 72 |
| 4.7.    | Construir um Meta-modelo.....                                      | 73 |
| 4.8.    | Considerar o acesso e apresentação dos metadados. ....             | 73 |
| 4.9.    | Antecipar o processamento de Metadados.....                        | 73 |

|   |    |
|---|----|
| 5. Gestão de Metadados em Data Warehouse.....       | 75 |
| 5.1. Metadados e as Mudanças em Data Warehouse..... | 79 |
| 5.2. Metadados para Múltiplos Data Warehouses ..... | 79 |
| 5.3. Versionamento. ....                            | 80 |
| 6. CONCLUSÕES .....                                 | 81 |

**LISTA DE TABELAS**

|   |    |
|---|----|
| Tabela 1 – Nível de Gerenciamento do Conhecimento .....                 | 34 |
| Tabela 2 – Questões e Metadados Afetados .....                          | 45 |
| Tabela 3 – Questões e Metadados de Processos Específicos Afetados ..... | 46 |

## LISTA DE FIGURAS

|  |    |
|--|----|
| Figura 1 – Processo de Transformação de Dados em Informação..... | 20 |
| Figura 2 – Exemplo de Modelo Estrela .....                       | 31 |
| Figura 3 – Arquitetura do Metamodelo CWM .....                   | 53 |
| Figura 4 - Arquitetura do Repositório de Metadados OMG .....     | 54 |
| Figura 5 – Compartilhamento de Metadados usando XMI.....         | 59 |
| Figura 6 – Fluxo de Metadados .....                              | 70 |
| Figura 7 – Abrangência dos Metadados em Data Warehouse .....     | 78 |

## LISTA DE ABREVIATURAS E SIGLAS

AD – Administrador de Dados

ANSI – American National Standards Institute

API – Application Program Interface

BI – Business Intelligence

CASE – Computer Aided System Engineering

CDIF – CASE Data Interexchange Format

CORBA – Common Object Request Broker Architecture

CWM – Common Warehouse Model

DBMS – DataBase Management System

DW – Data Warehouse

EIA – Eletronic Industries Association

ETL – Extraction Transformation Loading

IDL – Interface Definition Language

IEEE – Institute of Electrical and Electronics Engineers

ISO – International Standards Organization

IT – Information Technology

MDC – Meta Data Coalition

MOF – Meta Object Facility

ODS – Operational Data Store

OIM – Open Information Model

OMG – Object Management Group

RPC – Remote Procedure Call

SGBD – Sistema Gerenciador de Banco de Dados

SQL – Structured Query Language

TI – Tecnologia da Informação

UML – Unified Modeling Language

XMI – XML Metadata Interchange Specification

XML – Extensible Markup Language

W3C – World Wide Web Council

**LISTA DE SÍMBOLOS**

## 1. INTRODUÇÃO

As companhias estão sofrendo, atualmente, do complexo “Google”. Elas gastam milhões de dólares a cada ano comprando, construindo, implementando, e mantendo seus sistemas de processamento de dados e aplicações. Apesar disso, elas sempre esquecem de fornecer a última (e talvez a mais importante) peça do quebra-cabeça – um instrumento de “busca” que auxilie profissionais a pesquisar, encontrar, e acessar informações críticas (itens como relatórios de negócio, definições de dados corporativos, localização de arquivos, linhagem de dados e informações do tipo “onde foi usado/qual a última atualização”. Desta forma, um bom gerenciamento de metadados pode ser comparado a uma ferramenta de pesquisa na Web (“Google”, “Yahoo”, ou “Altavista”), auxiliando profissionais a localizar uma determinada informação mais fácil, e rapidamente, e de forma confiável, dentro do ambiente de informação corporativa.

A verdade é que, atualmente, a dependência do conhecimento é extremamente grande. Esse conhecimento, muitas vezes, concentra-se em algumas pessoas (fica na “cabeça” dessas pessoas), sem que se tenha um registro formal ou mesmo informal do mesmo. Se acontecer dessas pessoas se aposentarem ou mudarem para outra companhia corre-se o risco de perder esse conhecimento. Dessa forma, há a necessidade de buscar uma forma de documentar todos os processos e seu conhecimento associado, para que se possa compartilhá-los, tornando-os disponíveis a toda uma organização. Este tipo de raciocínio conduz aos assuntos relativos ao Gerenciamento do Conhecimento e do Gerenciamento da Informação. Este panorama leva ao conceito da Inteligência Empresarial (Business Intelligence), o qual acaba por alavancar a necessidade por uma boa Solução de Metadados. Por outro lado, pensando de uma maneira mais abrangente, convém voltar a atenção para uma solução que cubra outras áreas e necessidades, obtendo-se uma solução similar a uma ferramenta de pesquisa na Web.

Diferentemente do passado, metadados devem ser utilizados para prover a gestão dos recursos constituídos por dados, informação e conhecimento. Por conseguinte, metadados devem estar associados a tais recursos ao se construir soluções de metadados corporativas. Sabe-se que, uma informação precisa, porém, não contextualizada é

insuficiente para apoiar a melhor tomada de decisão, isto é, sem os metadados associados à informação é praticamente impossível mudar a estrutura mental do tomador de decisão em um processo decisório baseado em fatos, tal como ocorre no contexto do Data Warehouse. Descartar visões ultrapassadas de metadados é um dos fatores críticos de sucesso de soluções de metadados voltadas, por exemplo, para um Data Warehouse corporativo, onde ocorre geração de informação sob demanda.

### **1.1. Objetivos**

O objetivo deste trabalho é o de destacar a grande importância que uma representação eficiente e adequada de metadados pode proporcionar às empresas, notadamente aquelas que fazem uso intensivo de Sistemas de Informação.

Portanto, um objetivo inicial é o de destacar a importância que a informação representa para os negócios atuais, apontando-se a grande participação dos metadados nesse ambiente.

A partir dessa apresentação, apresenta-se uma proposta de um plano para a criação de uma solução de metadados, que contemple os principais aspectos envolvidos no processamento de metadados, notadamente aqueles ligados aos Data Warehouses.

### **1.2. Estrutura do Trabalho**

No capítulo 2 apresenta-se mostrado uma visão geral da informação sob diversas perspectivas, sua evolução dentro das empresas, problemas que surgiram com a má utilização dessas informações e conceitos que ajudam a compreender melhor a importância da informação.

No capítulo 3 apresentam-se os principais conceitos ligados à teoria de metadados, destacando-se seus tipos principais e alguns modelos utilizados.

No capítulo 4 serão abordadas as possíveis fontes de metadados nas empresas.

No capítulo 5 serão abordados os problemas que podem ser encontrados devido à falta de soluções de metadados.

No capítulo 6 será abordada a importância dos metadados para as empresas

No capítulo 7 será abordado o processo de Gestão de Metadados

## 2. PERSPECTIVA DA INFORMAÇÃO

Metadados é um assunto que é tratado há muito tempo em sistemas de processamento da informação. O que tem mudado é o escopo de sua atuação conforme a evolução da Tecnologia da Informação.

Antigamente os metadados somente estavam vinculados internamente aos programas. Com o surgimento dos Sistemas Gerenciadores de Banco de Dados (SGBD), muitas organizações tentaram implementar alguma forma de dicionário de dados, a partir dos anos 70. Depois vieram as ferramentas CASE e os ambientes de desenvolvimento com seus repositórios proprietários. Nos anos 80, houve uma iniciativa frustrada da IBM de criar um repositório global para troca de metadados através do seu AD/Cycle.

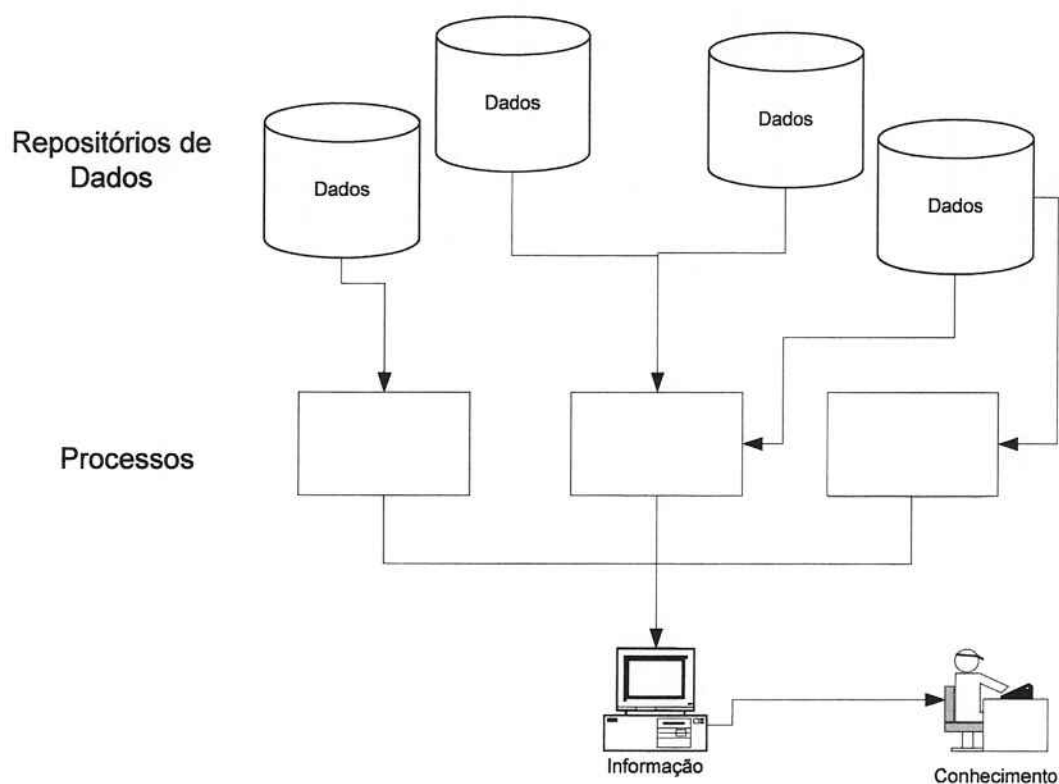
Através dos tempos, metadados têm sido apresentados para a comunidade de informática como um problema intratável. E como as organizações cresceram, os problemas com metadados se multiplicaram. Entretanto, os próprios problemas com metadados representam boas oportunidades para os fabricantes de software. Porém, isto é mais evidente em ambientes de Data Warehouse corporativos.

Por mais que se tenha informações suficientes, dificilmente há a certeza de que estas informações estão totalmente corretas. E a repercussão de uma informação incorreta pode ser catastrófica.

Nas organizações é comum passar mais tempo procurando a informação correta, validando-a e tentando provar sua correção, do que talvez executar a tarefa mais importante : analisar a informação e neste momento a área de Tecnologia da Informação passa a recuperar a sua responsabilidade sobre a informação.

Dados que não são capturados podem se tornar informação, mas apenas baseados em fatores externos, ou seja, algumas informações dependem de uma contextualização como por exemplo a experiência de um indivíduo em executar certas tarefas que não são documentadas mas fazem parte importante de um processo. Se o dado é capturado, sua função como informação depende de como ele é identificado no contexto de dados

relacionados e processos. Informação coletada e processada freqüentemente torna-se conhecimento, conforme mostra a figura 1.



**Figura 1 – Processo de Transformação de Dados em Informação**

Como a informação é definida em uma organização freqüentemente leva a questões como:

- Quanto tempo e recursos foram gastos localizando o dado?
- Quanto tempo adicional poderia ter sido gasto para se ter certeza de que o dado localizado estava pronto para suportar requisições adicionais de informação?
- Qual é o custo da informação em sua primeira obtenção vez e com qual freqüência repetem-se os mesmos gastos?

## **2.1. Definição da Informação**

A informação pode ser definida como o conjunto de dados disponíveis, que podem ser internos ou externos, numéricos, textuais, gráficos, visuais, áudio, etc., que quando combinados com qualquer conjunto de processos manuais ou mecanizados e integrados com informações preexistentes, auxiliam a tomada de decisão.

## **2.2. Evolução da Informação**

No início a informação tinha sua definição atrelada intrinsecamente à aplicação e, raramente, informações geradas por uma aplicação poderiam ser usadas para análise devido a esta dependência.

No princípio os dados eram transacionais e os processos computadorizados voltados para desempenho, eficiência e precisão. Aí apareceram os bancos de dados hierárquicos, possibilitando a separação e processamento de dados, mas ainda mantendo o vínculo com as aplicações transacionais e sem a intenção de serem compartilhados.

Quando o número de aplicações transacionais começou a crescer, criando múltiplos processos para acessar os mesmos dados houve a necessidade de se criar uma administração desses dados com o intuito de gerenciar e padronizar nomes e valores corporativos.

Com o advento dos gerenciadores de bancos de dados relacionais na década de 70, os dados passaram a associar-se com a informação final (resultante). Desta forma Processamento de Dados passou a ter significado de Sistemas de Informação e os dados passaram a ser “modelados” baseados no relacionamento de negócio. Linguagens de consulta padrão como Structured Query Language [SQL] permitiram a diferentes aplicações acessar os mesmos dados. Iniciou-se a separação entre dado e informação onde os dados eram organizados e relacionados conforme a sua função “potencial” como informação.

As novas aplicações da década de 70 e 80 usaram bancos de dados relacionais como seus alicerces e os softwares de sistemas de gerenciamento de banco de dados (SGBD) tornaram-se cada vez mais eficientes, as transações poderiam então ser organizadas e relacionadas para posterior análise, formando os dados que poderiam tornar-se informação.

No início a modelagem de dados era freqüentemente implementada seguindo-se rigorosamente as regras, o que sacrificou a flexibilidade ao acesso aos dados. O acesso aos bancos dados se tornou mais fácil e os usuários começaram a criar seus próprios “banco de dados privados”, muitas vezes usando dados extraídos dos bancos de dados corporativos. O advento do computador pessoal permitiu aos usuários de negócios mais experientes a construir seus próprios bancos de dados isolados extraíndo informações de diversas aplicações transacionais. A área de Sistemas de Informação passou a ser chamada de Tecnologia da Informação e passou a se concentrar na infra-estrutura tecnológica completa.

Com o surgimento das aplicações cliente/servidor o processamento foi transferido para os PCs (Personal Computers-Clientes) e até mesmo os dados, seguindo a filosofia de se ter os dados onde eles eram necessários, em detrimento de seu gerenciamento. Silos e mais silos de informação começaram a proliferar por todos os departamentos, cada um com seu servidor para um propósito específico.

Nos início da década de 90 a área de TI começava a sofrer com sua própria ignorância. As pessoas de TI pareciam ter esquecido como o dado, eventualmente, torna-se informação e que o negócio de TI é a informação. Usuários começaram a assumir o objetivo e destino da informação. Administradores de dados foram alocados nas áreas de negócio na tentativa de padronizar os dados antes que eles se tornassem informação. Redes de dados proliferavam em muitas organizações e a área de TI controlava os dados principais, mas não tinha o controle de onde ela estava sendo utilizado, e ainda pior, perdendo o controle da atualização das informações baseadas em fontes fora do domínio de TI.

### **2.3. O papel da informação**

Dados e Processos caminham juntos para formar a informação. Para que a informação possa ser utilizada, ou seja, sua exerce sua função, depende principalmente da forma como as empresas adotam regras básicas para a informação.

Regra 1: A informação pode, virtualmente, tomar qualquer formato, desde que seja identificável, armazenável e recuperável por seus identificadores.

Regra 2: A criação da informação não deve ser restringida pela sua função dentro do framework do processo de sua criação.

Regra 3: Nem todos os dados são criados, inicialmente, para se transformarem em informação, mas eles podem se tornar facilmente informação se o ambiente de negócios da empresa se alterar.

Regra 4: O processo de derivar informação de outras informações deve ser desencorajado, isto é, a informação deverá sempre ser derivada de dados para evitar a perda de controle sobre os mesmos.

Regra 5: A informação está em toda parte, mas sua usabilidade está relacionada diretamente com seu relacionamento com os dados básicos, todas as informações devem estar padronizadas, disponíveis e compartilhadas.

Regra 6: Quando a informação é necessária, seu usuário comum pode ter de trilhar um longo caminho para encontrá-la, obtê-la e aplicá-la.. Um bom ambiente de gerenciamento da informação elimina a necessidade de um árduo esforço.

Regra 7: A menos que o usuário da informação seja informado da existência da mesma, é certo que mesmo a melhor informação será usada opcionalmente. O usuário deve ter orientação e fácil acesso à informação.

### **2.4. O Contexto das informações nas organizações atuais**

É difícil admitir mais ainda há uma perspectiva limitada da informação, apesar da reengenharia e da evolução da tecnologia, dificilmente considera-se que localizar e

identificar uma informação seja mais importante do que interpretá-la. Se não se pode encontrar e identificar a informação, não se pode utilizá-la.

Normalmente criam-se túneis de informação, que são subconjuntos de informações orientadas ao um processo, devido aos requisitos de processamento que direcionam a informação para assuntos específicos que podem ser necessários em outros processos, mas que são difíceis de se compartilhar devido ao reflexo de um conjunto limitado de requisitos.

Muitas empresas gastam milhões na implantação de novos processos, novas aplicações, mas se esquecem do grande alicerce para tudo isso, a informação organizada, o que segundo Adrienne Tannembaum nos leva a 5 questões primordiais:

- Que dados a empresa possui?
- O que esse dados significam?
- Onde estão esses dados?
- Como localizar esses dados?
- Como obter esses dados?

A maioria das empresas poderá responder a algumas dessas questões, mas de maneira insatisfatória, e a forma como essas empresas respondem a essas perguntas podem levar a indicar a perspectiva limitada que se está tendo sobre a informação.

Geralmente as respostas para as 5 questões compartilham qualidades similares. Elas têm foco limitado e atendem à perspectiva do indivíduo ou grupo que forneceu as respostas. Mesmo dentro de organizações de gerenciamento de dados, as respostas para estas questões refletem um escopo reduzido baseado, normalmente, na incapacidade de atender fisicamente ilhas de dados específicas (plataforma física, pacotes de aplicações). Ironicamente, mesmo organizações de gerenciamento de dados que implementaram repositórios de metadados, dicionários de dados, ou diretórios de informações restringiram os limites de abrangência em muitos casos. Estes esforços, originados no

departamento de Tecnologia da Informação não são, normalmente, direcionados pelos requisitos de negócios dos usuários, acesso ou relevância.

Uma visão de perspectiva limitada é óbvia quando se atende às respostas para cada questão do ponto de vista das implementações de informação atuais.

**a) Que dados a empresa possui?**

A verdadeira resposta para esta pergunta deveria resultar em um índice de catálogo dos dados da empresa, organizado por áreas de assunto principais. Mas o que normalmente ocorre é que, geralmente, há uma grande dependência de processos manuais ou de pessoas chave para que se possa obter uma resposta satisfatória.

**b) O que esses dados significam?**

Na realidade, muitas organizações não têm bancos de dados corporativos, o que as leva a criar ilhas de informações redundantes e com significados diferentes, cada uma das quais representando perspectivas específicas do significado do dado, e nenhuma delas mantida por uma estrutura central ou global.

**c) Onde estão esses dados?**

Está diretamente ligado à definição do que estamos procurando e qual o esforço necessário para se obter esta definição o que pode variar dependendo do tipo, localização e nome do dado, nas organizações atuais onde não existem repositórios centralizados ou padronização de conceitos mesmo com um sistema sofisticado de busca fica difícil definir qual é o dado correto.

**d) Tipos de Dados**

Dados não operacionais são fáceis de se encontrar como documentos, aplicações, apresentações assumindo-se que seus nomes e conteúdos tenham um qualificador principal, mas os dados operacionais gerados por transações são mais difíceis de se localizar, pois muitas vezes estão orientados para o processo e para aplicação e não para o usuário final.

**e) Como localizar os dados?**

Determina o quão fácil o dado pode ser encontrado através de bases de dados, repositórios, diretórios, aplicações proprietárias.

**f) Associação dos Dados/Nomeação dos Dados**

Padronização de Nomes, apelidos, categorias de áreas de assunto ajudam a identificar dados que não têm nomes óbvios.

**g) Qual a origem dos dados?**

Mostra a verdadeira fonte da informação e como a informação foi, originalmente, gerada.

**h) Como obter os dados?**

A forma de acesso à informação correta pode ser um dos aspectos mais fáceis de ser atendido, mas algumas vezes quando se utilizam ferramentas ou aplicações, não se tem certeza de que processos e métodos foram utilizados para obter essas informações. Portanto, para se ter uma solução satisfatória deve-se ter uma forma de rastrear esses dados.

## **2.5. Compartilhamento de informações e redundância**

Em um cenário bem planejado a informação a ser compartilhada não deveria ser formada apenas pelos valores de dados, mas também pelas características e qualidades destes valores; as regras que envolvem a criação, manutenção e exclusão de valores; e os meios de acessar esses valores. Para estabelecer um compartilhamento padronizado há a necessidade de planejamento, análise e visão futura.

Muitos esforços de padronização iniciam-se pela reorganização dos dados, definindo-se características em comum baseadas em relacionamentos e perspectivas.

## **2.6. Suportando informação intra-organizacional**

A informação sustentada deve ser desejada.

Deveria ser mais fácil utilizar a informação suportada do que criar uma nova informação.

Acessar as informações existentes através de rotinas de acesso a dados existentes deveria facilitar o trabalho dos desenvolvedores .

Algumas das informações necessárias ao pessoal de Tecnologia da Informação são

- Os dados desejáveis;
- Os identificadores associados ao dado desejado;
- As definições, significado de negócios, e regras que permeiam os dados desejados;
- As regras que cercam a criação e/ou duplicação do dado desejado;
- Procedimentos de acesso necessário para obter o dado desejado (software).

## **Informações fora da organização**

Muitas organizações utilizam dados externos para validar, complementar ou verificar informações internas. Antes da disseminação da Internet havia empresas focadas

unicamente no fornecimento da informação, mas com o advento da Internet muita informação ficou disponível gratuitamente.

### **Os Fornecedores de dados**

O monopólio no fornecimento de informações levou muitas empresas a aceitar essas informações como verdade única, pois era única forma de obtê-las.

Informações da própria empresa passaram a ser processadas por fornecedores de dados (externos) para depois serem utilizadas internamente.

Com o advento da Internet e o crescimento na oferta de informações houve a necessidade da criação de padrões.

### **Troca de Informações**

O princípio da troca de informações é a criação de formatos padrões que podem ser agrupados em três categorias, considerando-se os padrões de troca atuais:

- Formato: a ordem, tamanho e estilo da entrada de dados;
- Conteúdo: que dados devem ser incluídos e que valores deverão ser manipulados;
- Processamento: o que deveria ocorrer com o dado e como

A troca de informações envolve ainda três outros conceitos:

- Formato Padrão: a ordem, tamanho, e estilo das entradas de dados

Um formato padrão deveria apontar que não importa o que está sendo descrito, o código é sempre de um tamanho fixo e consiste de componentes identificáveis específicos em uma ordem específica.

- Conteúdo Padrão: que dados devem ser incluídos, e valores.

Tipos comuns definidos dentro de um domínio específico, criados e aprovados por entidades reguladoras e esforços de padronização.

- Processamento Padrão: o que deveria acontecer com os dados e como.

Quando um padrão é criado deve-se garantir que o processo padrão seja aplicado igualmente a todo domínio.

## **2.7. Modelagem de Dados**

A modelagem de dados foi criada, tendo como finalidade original representar requisitos de dados de negócios de forma gráfica em termos de entidades, relacionamentos e atributos geralmente orientados a um projeto específico causando redundância nas informações entre projetos e em alguns casos a falta de padrões.

Qualquer tipo de projeto envolvendo desenvolvimento de banco de dados deve envolver três tipos de modelos de dados:

### **Modelo de Dados corporativo**

Modelo de dados de alto nível organizado por áreas de assunto, contendo entidades, relacionamentos e características descritivas, este modelo fornece uma descrição geral como ponto de partida para aplicações transacionais ou modelo de data warehouse.

### **Modelo de Dados da aplicação**

As especificações de dados da aplicação orientam o desenvolvimento deste modelo lógico. Identificando entidades, relacionamentos e atributos de um produto específico segundo padrões da organização.

### **Modelo de Dados Físico**

O modelo de dados físico é orientado à tecnologia de banco de dados a ser adotada em relação aos tipos de dados, definição de tabelas, campos e chaves de relacionamentos

(chave primária, chave estrangeira), determina como o modelo lógico será implementado fisicamente.

### **A organização do gerenciamento de dados**

A dificuldade no acesso aos dados gerou a demanda por uma área de gerenciamento de dados que por si só não poderia resolver o problema, pois as empresas deveriam passar por um processo de mudança na abordagem do desenvolvimento de aplicações. Os benefícios dessa visão são:

- Dados Corporativos padronizados para as áreas de assunto principais;
- Acesso comum aos dados que permitam a qualquer um chegar aos dados corporativos usando rotinas de acesso padrão;
- Eliminação de conflitos de dados pela lenta e estável infiltração de dados comuns;
- Ciclos de informe mais curtos através do incentivo do acesso aos dados comum.; e
- Entendimento claro dos dados de negócio através do acesso das definições comuns de dados.

### **2.8. Data Warehouse**

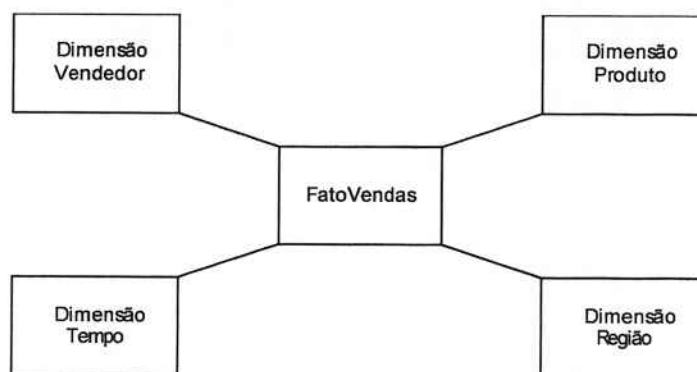
A técnica de Data Warehouse constitui-se em uma das maiores tendências identificadas nos anos 90, sendo idealmente focado em fornecer aos usuários finais dados precisos e integrados em um formato desejado, permitindo consultas e construção de relatórios conforme a necessidade sem a necessidade de customização.

O banco de dados contido em um Data Warehouse contém dados sumarizados e agregados em termos de fatos (por ex., transações de vendas, números, valores) e dimensões (produto, região, tempo) sob os quais os fatos são normalmente analisados (conforme exemplo da figura 2). Estes fatos e dimensões podem ser combinados de duas

formas principais: em formato estrela (star schema-exemplificado na figura 2) e em formato floco de neve (snowflake schema). Um Data Warehouse difere de um banco de dados comum, pois contém informações sobre o aspecto dos negócios da organização ao invés das operações transacionais do cotidiano. Os dados em um Data Warehouse geralmente são provenientes de processos já executados em sistemas transacionais.

Normalmente, o Data Warehouse é inicialmente definido e modelado em termos de fontes de dados . Os dados de um Data Warehouse provêm de muitas fontes, portanto uma intensa análise dos dados é necessária na fase de requisitos. Processos de Extração, Transformação e Carga (ETL) são projetados e uma arquitetura é definida em termos de ferramentas de interface e suas partes neste processo.

O ciclo de desenvolvimento de um Data Warehouse é iterativo, o que significa que ele está continuamente sendo atualizado . O ciclo de desenvolvimento de um Data Warehouse se repete em um padrão de aprimoramentos enquanto as mudanças de condições de negócio afetam a demanda por dados de suporte à decisão.



**Figura 2 – Exemplo de Modelo Estrela**

### **Abordagens para Data Warehouse**

Organizações executam o processo de Data Warehouse de diversas formas. Algumas organizações seguem a abordagem clássica de Data Warehouse (definida anteriormente). Algumas organizações simplesmente descarregam os dados do sistema fonte para o banco de dados de destino sem se preocupar com a qualidade da fonte, com as

agregações apropriadas ou com os dados sumarizados. Outras organizações desenvolvem Repositórios de Dados Operacionais (ODS-Operational Data Stores), os quais representam dados limpos trazidos de vários sistemas transacionais e armazenados em áreas de assunto principais em uma “área de staging” que alimenta vários data warehouses.

### **Vantagens e Desvantagens do Uso de Data Warehouse**

O resultado obtido a partir de um Data Warehouse bem planejado constitui-se em dados integrados e de qualidade, devidamente formatados e definidos para consulta e recuperação por usuários finais. Deve haver confiança suficiente de que os dados são precisos para se tomar decisões através deles. Porém a realidade pode ser bem diferente e como resultado os metadados passam a ter uma importância crucial.

Dados não se tornam precisos simplesmente por terem sido carregados em um Data Warehouse. Se as aplicações possuírem diferentes definições e fórmulas associadas com a criação dos dados, o processo de integrar múltiplas fontes não é, normalmente, simples e direto.

O sucesso de um Data Warehouse depende, em muito, do sucesso em se conseguir executar o processo de transformação adequado dos dados operacionais, tornando os dados inseridos em um Data Warehouse “limpos”, ou seja, sem problemas de consistência entre si.

### 3. METADADOS - VISÃO GERAL

Profissionais que têm algum contato com alguma ferramenta de Tecnologia da Informação utilizam, provavelmente, alguma forma de metadados, mesmo sem ter conhecimento que as usa ou mesmo o seu significado. Isto é normal até porque a sua própria definição não é um consenso, gerando alguma confusão. A definição mais comum de metadados é dados sobre dados. Porém esta definição não diz muita coisa. Ao pesquisar sobre metadados provavelmente serão obtidas diferentes interpretações. Algumas delas são aqui descritas, somente para ilustrar esta questão:

- Metadados são dados que descrevem atributos de um recurso. Eles suportam um número de funções: localização, descoberta, documentação, avaliação, seleção, etc.;
- Metadados fornecem o contexto para entender os dados através do tempo;
- Metadados consistem em dados associados com objetos que auxilia seus usuários potenciais a ter vantagem completa do conhecimento da sua existência ou características; e
- Metadados representam o instrumental para transformar dados brutos em conhecimento.

Segundo Terry Moriarty, estas diferentes interpretações estão relacionadas ao estágio da organização dentro da hierarquia evolucionária de gestão do conhecimento (ver figura 1). Organizações no nível mais baixo da hierarquia gerenciam dados brutos. Organizações mais avançadas são capazes de administrar seus recursos de informação no nível de Informação, Conhecimento ou Sabedoria. No nível de informação, o foco se concentra nos relacionamentos entre todos os componentes do sistema e os papéis individuais que eles assumem no sistema. Uma organização que está envolvida ao ponto que ela pode mostrar explicitamente as regras de negócio, governando seu comportamento está no nível de conhecimento. Uma organização atinge o nível mais avançado da hierarquia quando ela monitora ativamente seus sistemas para garantir que

seu comportamento esteja em conformidade com o planejado. Tal que uma organização pode detectar e diagnosticar qualquer comportamento anormal do sistema.

| Estágio      | Recurso a ser administrado  | Definição de Metadados   |
|--------------|---|--|
| Dados        | Valores dos dados   | Informação necessária para administrar o recurso dos dados   |
| Informação   | Valores dos dados e o contexto da informação  | Informação necessária para administrar o recurso da informação   |
| Conhecimento | Valores dos dados, contexto da informação e instruções das regras de negócio  | Informação necessária para administrar as regras e políticas de negócio da organização                                   |
| Sabedoria    | Valores de dados, contexto da informação, regras de negócio executáveis, monitoração das regras de negócio e regras e métricas de avaliação | Informação necessária para administrar o comportamento da organização, de acordo com suas regras e políticas de negócio. |

**Tabela 1 – Nível de Gerenciamento do Conhecimento**

A finalidade principal dos metadados é documentar e organizar, de forma estruturada, os dados das organizações, com o objetivo de minimizar duplicação de esforços e facilitar a manutenção dos dados.

Para se ter uma idéia de sua importância, vale a pena fazer uma analogia no mundo real. Os dados sem os metadados comportam-se como se fossem um turista em uma cidade desconhecida sem qualquer informação sobre a mesma. O usuário fica sem uma orientação para obter a informação desejada.

Atualmente o interesse sobre o assunto de metadados vem crescendo pelos seguintes motivos:

- As pessoas necessitam melhores formas de encontrar e avaliar informações na Internet e nas Intranets;
- Os sistemas de gerenciamento de conhecimento, integrando informações de múltiplas fontes e aplicações precisam oferecer maior facilidade de pesquisa e manutenção.

Pode-se, simplesmente, definir metadados como sendo os “dados sobre os dados”, ou ainda, como sendo a “informação dos dados”. Mas esta realmente seria uma definição muito simplista diante de tudo aquilo que os metadados podem representar. Na verdade tudo possui metadados, até os metadados possuem metadados. A grande preocupação não deve ser em torno do que são metadados, e sim de quais são os metadados existentes, como classificá-los, documentá-los, capturá-los e, principalmente, como armazená-los.

Metadado é a descrição de cada componente de um sistema de informação corporativo, é a base fundamental para se obter qualidade nos dados, maior integridade e unicidade entre as aplicações existentes, conhecimento e um gerenciamento mais efetivo do ambiente de Tecnologia da Informação.

Talvez uma definição mais exata seja a de que metadado é uma abstração dos dados, ou ainda, dados de mais alto nível que descrevem dados de um nível inferior. Sem metadados, os dados não têm significado. São exemplos de metadados as descrições de registros em um programa de aplicação ou o esquema de um banco de dados descrito em seu catálogo ou ainda as informações contidas em um dicionário de dados.

Em um ambiente operacional, os metadados são especialmente valiosos para os desenvolvedores de aplicações e para os administradores de bancos de dados. Os bancos de dados operacionais são, usualmente, utilizados por aplicações, que já contêm as definições embutidas de dados. Seus usuários, simplesmente, interagem com as telas do sistema, sem precisar conhecer como os dados são mantidos pelo banco de dados. O ambiente de suporte à decisão, por sua vez, é muito distinto. Nele, analistas de dados e executivos procuram por fatos não usuais e correlações que serão reconhecidas, apenas

quando encontradas. Aplicações rotineiras e predefinidas não fazem sentido nesse ambiente. Os usuários de um Data Warehouse precisam examinar seus dados e, para tal, conhecer sua estrutura e significado.

### **3.1. Entendendo metadados através de exemplos**

Os metadados estão presentes em todos os lugares no percurso dado-conhecimento. O problema está no fato de aquilo que alguns entendem como dados, para outros podem ser metadados. Além das perspectivas de metadados distintas consideradas, o que advém do fato de dado ser um termo genérico, devendo, portanto, ser renomeado, e precisamente definido como “dado instância”, ficando a precisão da definição de metadados dependente da precisão da definição do termo “dado instância”.

Dado Instância: Descrito pelo valor do Metadado, é aquele que é entrada para uma ferramenta de recepção, aplicação de banco de dados/arquivo, ou um mecanismo de processamento elementar. Se o valor de metadados é Total de Vendas então o dado instância associado poderia ser R\$ 1.000,00

É relevante que os responsáveis e envolvidos em soluções de metadados entendam a maneira pela qual todas as instâncias de metadados relacionam-se com o dado instância, que é sempre dissociada dos processos. Se adotada a definição de dado instância, então uma definição flexível de metadado seria:

Metadado: É a descrição detalhada do dado instância, o formato e as características do dado instância populado através de instâncias e valores.

Considerem-se alguns exemplos que referentes a situações bastante comuns, e que ilustram perspectivas de metadados:

a) Não é recomendável se aplicar ou ingerir um medicamento sem primeiro, por exemplo, verificar o prazo de validade, que é um metadado explícito, no caso, de valor

inquestionável. A inexistência ou falta de precisão deste metadado pode gerar desconfiança em relação ao conteúdo – o medicamento – por parte de quem o utilizaria e, conseqüentemente, retardar ou descartar o seu uso;

b) No processo de montagem de um PC, quando as instruções – metadados explícitos – não são precisas e sim conflitantes, incorretas ou mesmo inexistentes, acaba por gerar desconfiança, sendo tal processo interrompido, pois inibe a montagem do equipamento;

c) Se alguém está circulando por entre as gôndolas de um supermercado e percebe que o preço na etiqueta de marcação – metadado explícito – de um determinado produto não está correto, então, imediatamente, tende a rejeitar aquele produto. Entretanto, se um especialista em precificação deste supermercado estiver presente e contextualizar a informação de preço, explicando que acabou o período de promoção – metadado tácito – então, tendo modificado a sua estrutura mental, a pessoa reverte a sua decisão, podendo adquirir o produto.

Transportando estas metáforas do cotidiano para o ambiente de um Data Warehouse corporativo, no qual armazenam-se dados históricos de um período de alguns anos, construído sem metadados, pode-se verificar um panorama bastante complicado.

Segundo [Inmon] os metadados englobam o Data Warehouse e mantêm as informações sobre o que está armazenado e em qual local. Ele ainda define quais informações os metadados mantêm:

- A estrutura dos dados segundo a visão do programador;
- A estrutura dos dados segundo a visão dos analistas de Sistemas de Apoio à Decisão;
- A fonte de dados que alimenta o Data Warehouse;
- A transformação sofrida pelos dados no momento de sua migração para o Data Warehouse;
- O modelo de dados;
- O relacionamento entre o modelo de dados e o Data Warehouse;

- O histórico das extrações de dados;
- Informações referentes a relatórios gerenciais;
- Informações referentes às camadas semânticas; e
- Informações referentes aos processos de Carga.

### 3.2. Tipos de Metadados

Metadados técnicos fornecem suporte a analistas, desenvolvedores e administradores, durante desenvolvimento, manutenção e gerenciamento de um ambiente de Tecnologia da Informação. Esses metadados constituem uma espécie de “cola” técnica que liga as ferramentas, aplicações e sistemas que, em conjunto, constituem uma solução. Por exemplo, metadados técnicos podem referir-se à estrutura de bancos de dados, a aplicações instaladas ou a sistemas, dentre outros.

Metadados técnicos representam a descrição dos dados necessários pelas várias ferramentas para armazenar, processar ou movimentar dados. Estas ferramentas incluem banco de dados relacionais, ferramentas de desenvolvimento de aplicações, ferramentas de modelagem, ferramentas de pesquisa em banco de dados, ferramentas OLAP, etc.

Metadados de negócios, por outro lado, tornam os serviços do ambiente empresarial mais compreensível para os usuários finais. Por exemplo, tais metadados fornecem explicações dos objetos de negócios e processos, visando fácil consulta, navegação, e pesquisa de dados.

Metadados de negócio constituem a descrição de dados necessários pelos usuários de negócio, para entender o contexto do negócio e o significado dos dados. Atualmente existem ferramentas só para efeito de documentação.

Quando metadados for rotineiramente usado para gerar regras de negócio executáveis, a definição de metadados será a representação de instruções de regra de negócio de acordo com o esquema de classificação que pode ser transformado em sistemas de informação do negócio.

Metadados operacionais (ao nível das aplicações): definem a estrutura dos dados mantidos pelos bancos operacionais, usados pelas aplicações de produção da empresa. São, freqüentemente, complexos e tendem a serem orientados à aplicação ou ao departamento;

Metadados centrais do Data Warehouse (Administrativo): representam a forma como os dados são tratados nos sistemas de Data Warehouse. Mantidos no catálogo do Data Warehouse, distinguem-se por serem orientados por assunto e, portanto, mais voltados ao mundo real, definem como os dados são transformados e como devem ser interpretados, incluindo definições de agregação e campos calculados, assim como a forma com que as visões sobre cruzamentos de assuntos serão criadas;

Podem estar divididos em informações estáticas e informações dinâmicas.

Das informações estáticas constam o nome do Data Warehouse, sua descrição, além de informações sobre a qualidade dos dados que ele contém.

Nas informações dinâmicas estão relacionadas a operacionalidade do Data Warehouse, como por exemplo, controle de cargas, ou seja, se a carga de todas as lojas ocorreu durante a noite ou não. Informações como essas são valiosas para os analistas que estão trabalhando com as informações.

Este metadado é visível pelo usuário final, quando, ao acessar o sistema, verifica qual o critério de cálculo utilizado para definir as quotas de vendas de produto, por exemplo.

Metadados do nível do usuário (de Negócio): mapeiam os metadados do Data Warehouse para conceitos que sejam familiares e adequados aos usuários finais.

Unindo essas duas informações observa-se que o metadado ao nível do usuário final é utilizado, como o próprio nome diz, pelo analista de negócio, que deseja analisar os dados de sua empresa, de acordo com as regras de negócio, além de poder analisar os cálculos utilizados para a geração de dados derivados. Este nível de metadado também é

utilizado para a geração de relatórios e gráficos nas ferramentas de acesso ao Data Warehouse.

O metadado Core Warehouse (metadados com informações detalhadas de onde serão extraídas as informações, estruturas de bancos de dados, modelos de dados é utilizado pelo administrador de dados e por desenvolvedores de ferramentas de acesso para a geração de programas, consultas, controles e procedures para a gerência e processamento do Data Warehouse.

O Metadado ao nível de aplicação é utilizado para acessar as informações que comporão os Data Warehouses.

Os metadados associados a um Data Warehouse podem ser classificados conforme a classe de seus componentes, dentre elas:

- Mapeamento: descreve como os dados de sistemas operacionais são transformados antes de entrarem no Data Warehouse. Exemplos desta classe de metadados podem ser os que identificam campos fontes, mapeamentos entre atributos, conversões, codificações, padrões, etc.;
- Histórico: com a evolução dos sistemas operacionais, as regras de negócio de uma empresa podem mudar, cabendo a estes metadados manter o histórico de mudanças destas regras, pois regras certas devem ser aplicadas aos dados certos;
- Miscelânea: esta classe define diversos tipos de metadados, informações da situação sobre estágios de desenvolvimento de partes do Data Warehouse, informações sobre volume dos dados para estimativas de tempo e recursos, etc.;
- Algoritmos de sumarização: mostram a relação entre os diferentes níveis de detalhes dos dados, indicando, inclusive, que nível de sumarização é mais adequado para um dado objetivo;

- Padrões de acesso: mantêm informações sobre frequência e tipo de acesso aos dados.

Conforme visto os dados sobre desempenho e monitoramento também se qualificam como metadados. Os processos que monitoram o ambiente de um Data Warehouse (tais como extração, carga e uso) criam metadados que são usados para determinar como o sistema vem atuando em termos de desempenho. Da mesma forma, dados que identificam questões relativas à qualidade dos dados, detectados durante os processos de extração e carga, devem também estar disponíveis para os usuários, para que estes possam julgar a precisão de suas análises.

### **3.3. Metadados voltados para o Projeto do Data Warehouse**

A seguir são apresentados quatro tipos de metadados apresentados por Barquin. Esses metadados são orientados para projetos iterativos e incrementais, incluindo-se a sua implementação e manutenção de Data Warehouse.

#### **Metadados que capturam o ambiente corrente**

Neste tipo de modelo de metadados o Data Warehouse deve manter os seguintes tipos de metadados:

- Registro e definição de elementos de dados; e
- Os inter- e intra- relacionamentos dos bancos de dados.

Os inter-relacionamentos refletiriam o mapeamento entre sistemas operativos e os atributos existentes no Data Warehouse e poderiam incluir a representação das regras de negócios.

Uma definição para cada interface (programa) é utilizada para construir ou atualizar um warehouse. Neste caso a definição deveria incluir:

- Quais junções entre o banco de dados são utilizadas (entrada e saída);
- Tempo e a direção da execução;
- Parâmetros de execução;
- Dependências entre interfaces; e
- O nome e a localização dos arquivos que contenham os códigos dos programas de interface e os nomes de ferramentas e sessões, no caso dos programas de interface terem sido gerados, automaticamente.

### **Metadados para reduzir os custos de erros**

No processo de análise de dados e teste de interfaces, erros associados aos valores dos dados devem ser detectados.

O modelo de metadados deve permitir incluir informações como a faixa de valores aceitos.

Poder-se-ia incluir uma lista de valores aceitos ou uma regra de negócios (expressão) que descreva as características dos dados válidos, ou o nome de um arquivo ou tabela que os contenha.

Pode ainda constar uma lista de exceções válidas com as quais a interface poderia apresentar ao se defrontar com um valor ilegal.

Esses tipos de erros estão relacionados, basicamente, a mudanças de esquema. Um exemplo clássico é o tipo "F" e "M", que pode passar a "Fem" e "Mas", que poderia ter sido mapeado para o Data Warehouse na forma de 1 e 2 reduzindo as chances de erro. A grande preocupação seria da carga da informação, mas qualquer mudança não afetaria a estrutura do metadado.

### **Metadados para reduzir o custo de complexidade**

Este tipo de metadados se preocupa com a forma como o usuário verá a informação. Esta preocupação ocorre em virtude da dificuldade de se entender os nomes dos campos nos esquemas atuais.

Para reduzir a complexidade, o metadado deve suportar a definição de *aliases* (um rótulo, um nome legível) para os nomes de campos e definir sub-esquemas.

### **Metadados de Fator Tempo**

O modelo de metadados deve apresentar dois tipos de informações relacionadas a tempo:

- Informações que possam afetar o tempo de execução sobre cada banco de dados, como tamanho e volatilidade (periodicidade de alteração do dado) do banco, o tempo durante o qual cada banco pode ser acessado e o mecanismo que pode ser utilizado para capturar dados alterados.
- Versionamento: Selecionar a estratégia de versionamento adequada pode fazer uma significativa diferença na redução dos custos relacionados à manutenção e iterações. A forma mais fácil de se lidar com diferentes versões de metadados é salvá-los sempre que uma mudança for realizada. Entretanto, as abordagens existentes podem não suportar a análise de impacto.

### **3.4. Padrões de Metadados**

Para os fabricantes atingirem um nível de integração de metadados há várias estratégias: construir uma interface proprietária entre vários produtos, fornecer uma solução completa de Tecnologia de Informação ou criar um padrão internacional de troca de metadados.

As interfaces oferecem uma alta funcionalidade e compartilhamento de metadados. Mas manter estas interfaces consome muitos recursos, especialmente se o número de pontes cresce. Vários fabricantes estão adquirindo ou construindo os componentes necessários

para oferecer um ambiente integrado de uma solução completa. Esta estratégia requer que o usuário abandone as melhores abordagens e se sujeite aos produtos de um único fabricante. Os usuários têm que pesar as desvantagens de ficarem presos a uma linha de produtos de um fabricante contra as funcionalidades que um ambiente proprietário pode oferecer.

A outra alternativa é a utilização de um meio comum para a troca de metadados através dos padrões internacionais.

Existem padrões diferentes de metadados para finalidades distintas de informações. Para se ter uma idéia da variedade de esforços, relacionam-se, a seguir, vamos relacionar alguns destes padrões:

### **Padrões internos**

A maioria das empresas inicia pela padronização dos dados, resultando na integração de aplicações e dados, ao passo que esses padrões de dados evoluem, o primeiro tipo de padrão orientado para os metadados normalmente foca em nomes para identificação de dados – nomes de programas, nomes de elementos de dados, nomes de campos, nomes de tarefas. Estes padrões de nomenclatura iniciais sempre incluem padrões de abreviação, primariamente porque linguagens de programação anteriores possuíam restrições físicas (limite de caracteres). Mas o advento das ferramentas que perderam as restrições físicas fez com que as abreviações, muitas vezes se tornassem obsoletas e por vezes ser um processo executado manualmente, sendo que em pouco tempo deixariam de ser padrões. Desta forma, os padrões internos evoluíram para incluir e afetar metadados além dos nomes. Padrões de metadados internos deveriam incluir todos os metadados associados com a habilidade de responder às questões da tabela 2.

| <b>Questão</b>         | <b>Metadados Afetados</b>   |
|------------------------|---|
| Que dados nós temos?   | Nome do Elemento de Dados   |
| O que eles significam? | Definição do Elemento de Dados  |
| Onde eles estão?       | Localização do Elemento de Dados(Nome do repositório de dados), Identificador do elemento de dados                        |
| Como ele chegou lá?    | Origem do Elemento de Dados, Procedimento(s) de transição do elemento de dados, Regra(s) tradução dos elementos de dados  |
| Como obtê-los?         | Nome do Procedimento de Acesso ao Elemento de Dados, Passo(s) da Execução do Procedimento de Acesso ao Elemento de Dados. |

**Tabela 2 – Questões e Metadados Afetados**

De fato, é essencial que enquanto se esteja criando os dados, o metadados associados também sejam criados. E enquanto esses metadados são criados eles estão populando diferentes partes de um meta-modelo organizacional. Estes padrões de metadados são orientados aos dados, mas eles são essenciais para fornecer informação por demanda. Se processos específicos de metadados são também de importância e preocupação, outros elementos de metadados provavelmente tornaram-se requisitos padrões que refletiram um ponto de vista das 5 questões como na tabela 3 a seguir.

| <b>Questão</b>                       | <b>Metadados Afetados</b>   |
|--------------------------------------|---|
| Que processos nós temos (programas)? | Nome da Aplicação, Nome Programa, Nome Modelo Processos   |
| Que funções eles executam?           | Nome Função, Processo/Programa/Matriz de Funções/Relacionamento   |
| Onde eles estão?                     | Localização do Programa(Nome da Biblioteca de Programas)  |
| Como eles chegaram lá?               | Desenvolvedor/Nome do Analista, Situação do Programa (Desenvolvimento/ Teste/ Produção)                           |
| Como obtê-los                        | Aplicação/Nome do Procedimento de Acesso à Biblioteca do Programa, Passo(s) da Execução do Procedimento de Acesso |

**Tabela 3 – Questões e Metadados de Processos Específicos Afetados**

Muitas organizações já requerem uma quantidade substancial de metadados como na tabela acima enquanto as aplicações são desenvolvidas. Contudo, a acessibilidade dos metadados varia, dependendo de onde os programas correntes estão armazenados e a acessibilidade.

### **Padrões Externos**

A palavra metadados, acompanhando meta-modelos, e soluções de metadados existem desde o início dos anos 90. Naquela época, fornecedores principais tentaram usar um conjunto de “meta-modelos padrões” como um meio de intercambiar modelos focados na aplicação de suas melhores ferramentas de modelagem. Antes dos anos 90, projetistas de dicionários de dados independentes não estavam preocupados com a padronização dos meta-modelos com o propósito de conectividade para instâncias de outros produtos. A solução da IBM AD/Cycle não teve sucesso, baseado nesta falha, ferramentas CASE (Computer Aided System Engineering) concordaram com um formato de intercâmbio

padrão, conhecido como CDIF (Case Data Interexchange Format), o que permitiu que alguns modelos e componentes fossem trocados entre ferramentas, enquanto fossem compatíveis com o CDIF e compartilhassem construções de modelagem comuns. Este padrão, fiscalizado pela Associação de Indústrias de Eletrônicos (EIA), foi a primeira a enfatizar semelhança e intercâmbio ao nível de metadados. Contudo o padrão CDIF foi incorporado a padrões mais influentes), seu projeto e execução ainda são comuns.

Nos padrões de troca de metadados originais, um arquivo batch era formatado com um cabeçalho padrão (header). O cabeçalho continha informação relacionada com as características físicas do restante do arquivo. Este padrão não necessitava de um tipo específico de processamento ou tratamento do conteúdo do arquivo de exportação pela ferramenta de recebimento. Tornou o processo de transferência um pouco mais fácil para os fornecedores provendo um formato que eles poderiam seguir em suas próprias ferramentas utilizando os recursos de importação e exportação.

Um padrão tão simples como este era desprovido dos seguintes componentes principais:

- Meta-modelos padronizados;
- Processamento das instâncias de meta-modelos padronizados;
- Componentes de meta-modelo padronizados que encorajariam o compartilhamento de tipos específicos de metadados entre as ferramentas; e
- Instâncias de componentes de meta-modelo padrão que encorajariam o compartilhamento de metadados dentro das indústrias, sem contar com a ferramenta disponibilizada ou aplicação.

A padronização de metadados teve seu foco, inicialmente, no intercâmbio, enfatizando o formato dos arquivos batch de exportação. Como a demanda por iniciativas baseadas em metadados cresceu, padrões começaram a se tornar influentes. Aproximadamente 10 anos haviam se passado desde o primeiro padrão de troca de metadados. Durante este período, muitas iniciativas de padrões conflitantes uniram-se ou foram eliminados, e hoje resta um pequeno grupo influente. Cada um com um ramo de atividade e filosofia,

apesar disso a ênfase da política dos fornecedores está se tornando mais orientada ao usuário. Cada padrão é agora representado por indivíduos dos seguintes campos:

- Fornecedores de Software;
- Representantes das Indústrias de consumo, focando na necessidade de compartilhar e trocar dados e metadados dentro de seu próprio conjunto de organizações; e
- Defensores de padrões internacionais, primariamente de organizações de pesquisa, todos defendendo a necessidade por um mecanismo de transferência baseado em metadados padronizado entre indústrias e independente de fornecedor.

Além disso, todas elas pretendem conseguir uma integração externa real. Os resultados destes esforços permitiram que os usuários de software compartilhem dados entre produtos de diferentes fornecedores, baseado no uso de metadados padronizados.

Atualmente as empresas de metadados principais se preocupado com três aspectos dos mundos dos metadados :

- Meta-modelos, como um meio de padronização de tipos de metadados usados e rastreados por funções de TI e seus produtos suportados entre as indústrias de consumo;
- Troca de Metadados, focando na transferência das instâncias desses meta-modelos (metadados padrão) entre implementações de software; e
- Arquiteturas e Estruturas de Metadados, como um meio de criar um pacote de metadados padrão que não apenas inclui o meta-modelo e os formatos de trocas, mas também processos associados e componentes de ferramenta que podem ser usados para garantir a aderência a um padrão de metadados definido.

Grandes iniciativas de padronização têm sido combinadas na maioria das vezes no Object Management Group (OMG). O OMG recentemente adotou o MDC (Meta Data Coalition) em uma “fusão de padrão de metadados” e é agora a força principal nos três aspectos dos requisitos de componentes padrão de metadados.

O American National Standards Institute (ANSI) também tem um meta-modelo padrão ajustado para compartilhar metadados. Finalmente o Institute for Electrical and Eletronics Engineers (IEEE) tem sido conhecido por expandir gradualmente este meta-modelo para suportar indústrias baseadas em engenharia específicas. E por último o XML (Extensible Markup Language) é o resultado de um padrão ANSI (W3C) que juntou forças com uma corporação internacional de padrões (International Standards Organization - ISO).

A melhor forma de analisar o impacto de padrões no mundo dos metadados é avaliar algumas perspectivas importantes, em termos de seus meta-modelos, padrões de troca, e arquitetura geral.

#### **3.4.1. Meta-modelos**

Muitos esforços em padronização não parecem inicialmente focar em meta-modelos, mesmo assim esta é sempre a área que os usuários de padrões tendem a se prejudicar durante a avaliação de padrão. Quando meta-modelos são padronizados, o intuito é uma organização uniforme de metadados e seus inter-relacionamentos. Esta uniformidade é tipicamente centrada em uma função específica, coleção arquitetural, requisitos físicos de TI (SGBD, linguagem de programação, pacote de software ou ferramenta), ou focado em metodologia. Infelizmente quando o desenvolvimento de um meta-modelo padrão não é baseado em um conjunto bem definido de metadados existentes, a definição de padrões é freqüentemente usada apenas por pessoas responsáveis por sua definição.

Meta-modelos padrão almejam uma integração de metadados sem traumas, apesar de sua ferramenta originária, aplicação ou localização física do software. Quando o modelo é uniforme, não é necessário que todos os fornecedores de metadados unam-se a ele,

mas preferencialmente que cada fornecedor de metadado esteja ciente do meta-modelo padrão para popular interfaces residentes e pontos de transição.

Houve várias iniciativas ao longo do tempo para padronização dos meta-modelos. Iniciando com o AD/Cycle, meta-modelos anteriores focaram em aspectos distintos do mundo de TI, em oposição ao mundo de negócios. O modelo de informação original AD/Cycle consistiu de larga variedade de sub-modelos, incluindo um sub-modelo corporativo que identificava e organizava a corporação empresarial em termos de sua informação, objetivos, recursos, modelos de negócios e modelos de aplicação lógicos. Outros sub-modelos então detalharam estes aspectos da empresa, e foram classificados como submodelos de tecnologia ou configuração de aplicação. As construções comuns eram modeladas separadamente em um sub-modelo global. Com a falta de aceitação do AD/Cycle, uma pausa despencou sobre o mundo de padrões de meta-modelagem por aproximadamente 5 anos, nos quais houve o desenvolvimento do Open Information Model (OIM) que iniciou como uma junção entre Microsoft e Texas Instruments.

### **3.4.2. O Open Information Model [OIM]**

O [OIM] refletia uma filosofia um pouco diferente do AD/Cycle. Ao invés de pegar os metadados de uma perspectiva empresarial de cima para baixo, o OIM iniciou de baixo e trabalhou o seu caminho para cima. O OIM começou a abranger não somente funções de TI (análise e projeto), mas também começou a estabelecer elementos genéricos como um meio de padronização de tipos de metadados que poderiam ser rastreados por todas essas funções., sem levar em consideração as ferramentas que estavam sendo disponibilizadas como implementações de meta-modelo. Esta abordagem independente de ferramenta foi a primeira de seu tipo, e começou pouco depois que os modelos iniciais foram liberados. Naquela época, incorporação da abordagem de Unified Modeling Language (UML), originária de outra iniciativa de padronização (OMG) padronizou-o. Neste ponto da história da OIM, muitos fornecedores juntaram-se à iniciativa e foram responsáveis por prover a abordagem “muitos fornecedores, ferramenta independente”.

O Objetivo desses modelos é representar uma especificação para todos os fornecedores de ferramentas que suportam análise e projeto de TI. Aderência a estas especificações não apenas facilitariam a troca de metadados associados com análise e projeto, mas também permitiriam o armazenamento deste metadado em um outro repositório de metadados além do da própria ferramenta, dessa forma encorajando uma abordagem de solução de metadados para compartilhamento de troca de metadados. O OIM finalmente incluiu modelos que representavam virtualmente todos os aspectos de desenvolvimento de software e suas implementações associadas. Categorias de Modelo principais eram:

- Análise e Projeto;
- Objetos e Componentes;
- Bancos de Dados e Warehousing;
- Engenharia de negócios; e
- Gerenciamento do Conhecimento

Todos os modelos nestas categorias teriam que ser representados em UML, a nova Linguagem de Modelagem Unificada.

Como o desenvolvimento do OIM continuou, mesmo assim houve outras iniciativas de padronização, desta vez focando exclusivamente em metadados para Data Warehouse. Esta iniciativa, o Meta Data Coalition (MDC) , foi liderada por um fornecedor de uma ferramenta específica de Data Warehouse, com o objetivo de padronizar os metadados necessários para suportar o Data Warehouse, do início do desenvolvimento até o suporte à decisão e análise do usuário final que resultou no final das contas. O MDC teve impacto mínimo no início, especialmente devido ao resultado de sua declaração pública da necessidade de suporte de empresas privadas e contribuições. Seus padrões focaram na troca de metadados em batch (MDIS) semelhante ao CDIF (CASE Data Interexchange Format) sem qualquer requisito de conformidade do meta-modelo.

À medida que o Data Warehouse tornava-se mais popular, o OIM foi estendido em várias áreas para suportar o Data Warehouse. Mais de 200 fornecedores participaram

desta iniciativa, incluindo aqueles que faziam parte do MDC inicial, e o contraditório ainda que complementar esforço foram unificados no final de 1998. O OIM, neste ponto inicialmente apoiador do Data Warehouse, foi entregue ao MDC.

Em conjunto com o MDC, outra grande iniciativa surgiu na área de Data Warehouse. O OMG (Object Management Group), o criador do UML, que foi adotado durante o desenvolvimento do OIM, e criou o Common Warehouse Model (CWM). Esta foi uma iniciativa posterior do OMG, pois seu trabalho original estava centrado no nível de meta-meta-modelo e a necessidade pela padronização da troca de metadados, em oposição às suas instância e relacionamentos. Muitos profissionais de TI não se identificaram com as iniciativas da OMG no início.

### **3.4.3. O Common Warehouse Model**

Quando o Data Warehouse se tornou a primeira aplicação com custo/benefício real para troca de metadados, muitas iniciativas de padrões teóricos começaram a se voltar para realidade e oferecer o primeiro meio, aparentemente prático e disponibilizável, de justificar o intercâmbio de metadados comuns, ou abordagem de solução de metadados ao Data Warehouse. O OMG, em existência desde 1989, tem sempre se esforçado para desenvolver especificações independentes de fornecedor para indústria de software. Além disso, sua abordagem ao Data Warehouse introduziu ainda uma outra abordagem filosófica. Diferente do Open Information Model, no qual o objetivo primário era o intercâmbio e relacionamento de metadados, o OMG tem sempre se focado na padronização com o intuito de reuso. De fato, ele tem sempre se esforçado para facilitar o desenvolvimento de software baseado em componentes pela definição de padrões e componentes independentes de ferramenta. O Common Warehouse Model é um exemplo desta filosofia, ele foi construído exclusivamente usando a UML da OMG. Por que criar mais componentes de modelo para Data Warehouse do que os já existentes? Por que redefinir requisitos de intercâmbio quando recursos de mercado, como o XML,

e/ou outras APIs já definidas (como o MOF-Meta Object Facility) já foram desenvolvidas ?

Baseado na filosofia orientada a componentes da OMG, o CWM, contudo nomeou e considerou dentro do escopo de Data Warehouse, reflete o fato de que Data Warehouse é uma reorganização de recursos empresariais existentes. Estes recursos incluem dados assim como processos que quando aplicados aos dados, resultam em repositórios de dados reestruturados com conclusões orientadas à análise baseada em dados. Seguindo esta abordagem, o CWM consiste de vários pacotes, cada um contendo um conjunto de outros pacotes (podemos considerá-los meta-modelos). Muitos dos pacotes que formam o CWM não são específicos para o Data Warehouse. Elementos de modelo são definitivamente comuns e reutilizáveis para suportar a abordagem baseada em componentes da OMG.

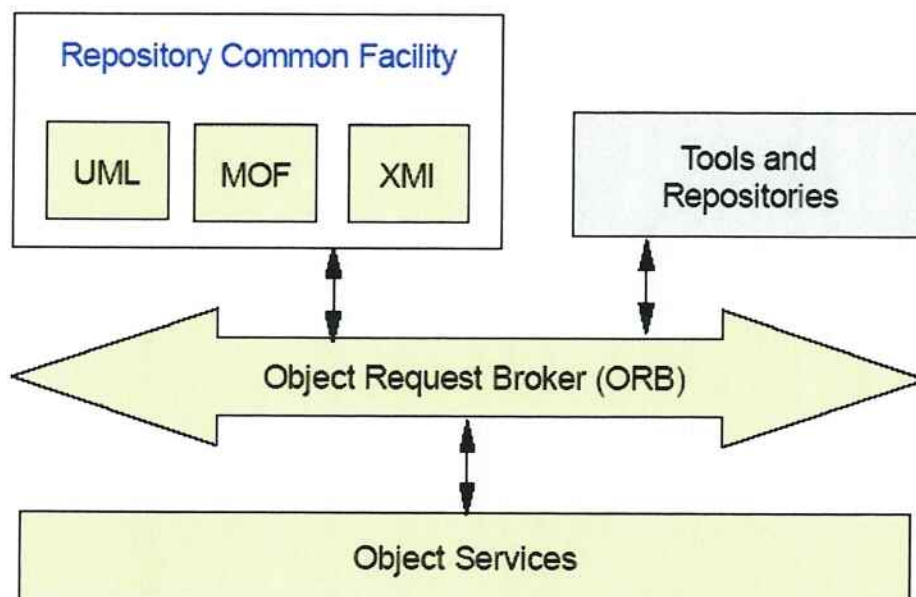
|              |                      |            |            |                     |                           |                       |
|--------------|----------------------|------------|------------|---------------------|---------------------------|-----------------------|
| Management   | Warehouse Process    |            |            | Warehouse Operation |                           |                       |
| Analysis     | Transformation       |            | OLAP       | Data Mining         | Information Visualization | Business Nomenclature |
| Resource     | Object Model         | Relational | Record     | Multidimensional    |                           | XML                   |
| Foundation   | Business Information | Data Types | Expression | Keys and Indexes    | Type Mapping              | Software Deployment   |
| Object Model |                      |            |            |                     |                           |                       |

**Figura 3 – Arquitetura do Metamodelo CWM**

O CWM consiste de 4 pacotes principais(conforme figura 3):

- O Foundation Package identifica elementos comuns de metadados (informações de negócios, Tipos de Dados, Expressões, Chaves e Índices, Disponibilização de software, Mapeamento de Tipos) como definido no UML e reutilizável por todos CWM.
- O Resource Package organiza os metadados relacionados com os repositórios de dados classificando-os com XML (Relacional, Registro, Multidimensional, XML).
- O Analysis Package foca nos metadados associados com análise de dados(Transformação, OLAP, Data Mining, Visualização da Informação, Nomenclatura de negócios).
- O Management Package organiza os metadados que envolvem o suporte e coordenação dos data warehouses (Processos Warehouse, Operações Warehouse).

## OMG Metadata Repository Architecture



**Figura 4 - Arquitetura do Repositório de Metadados OMG**

#### 3.4.4. Troca de Metadados

Um meta-modelo padrão sozinho não garante a troca e utilização padrão de metadados. Muitas iniciativas em padrões têm focado nos meios de acessar e transportar metadados do meta-modelo populado. Padrões de troca variam substancialmente em suas implementações. Além dos padrões de troca batch (CDIF, MDIS) existem outros meios de acessar e trocar metadados como:

- Application Program Interfaces (APIs), rotinas padrão que podem ser chamadas de dentro de programas de acesso a metadados, dentro do controle deste programa;
- Remote Procedure Calls (RPCs), rotinas padrão que são executadas por programas de acesso a metadados, fora do controle deste programa;
- Embedded Processes, que são associados com componentes distintos de um meta-modelo padrão; e
- Standards-Based Languages, que são definidas por órgão de padronização e disponíveis para uso por programas de acesso a metadados. Em muitos cenários, estas linguagens são combinadas com interfaces e arquiteturas necessárias.

Iniciativas de padronização normalmente tentam focar na troca de metadados de forma reutilizável, baseado normalmente no tipo de metadado que está sendo trocado, ou em alguns casos, na origem do metadado. Padrões de foco restrito, como discutido, englobam apenas formatos de troca de metadados, sem se preocupar muito com o que está sendo trocado, como o metadado deveria ser processado, ou onde ele deveria ser colocado dentro do repositório. Recentemente, os problemas de troca de metadados têm sido considerados mínimos em conjunção com meta-modelos padrões. As implementações tem resultado em conjuntos de APIs e/ou RPCs padrão, que são associadas com tipos de metadados, baseados em seus relacionamentos no meta-meta-modelo.

### **3.4.5. Processos orientados a metadados embutidos**

O OMG tem uma abordagem arquitetural para seus padrões de metadados, focando principalmente no potencial reuso de muitos de seus componentes de meta-modelo. O OMG tem defendido há muito tempo o Common Object Request Broker (CORBA), dos quais o Meta Object Facility (MOF) é um componente. No nível de meta-meta-modelo (o Pacote de Modelo é o modelo dos modelos), é importante notar a colocação de características comportamentais, especificamente, as Operações, que estão indiretamente associadas com Elementos de Modelo. Estas operações são, tipicamente, uma idéia do ponto de vista de uma instância de meta-modelo, mas implementadas no MOF como um tipo de comportamento. O MOF define operações específicas, e algumas (findRequiredElements) são direcionadas para troca de metadados entre implementações em conformidade com o MOF. Considere o MOF um exemplo de um conjunto de processos orientados a metadados embutidos.

À parte das operações embutidas que nós vemos no nível meta-meta do MOF, cada pacote (meta-modelo) tem suas interfaces definidas, as quais são tipicamente direcionadas às instâncias do próprio meta-modelo, portanto nosso metadado.

### **3.4.6. Linguagens baseadas em padrões**

O OMG esforça-se para padronizar as linguagens usadas para definir novas interfaces, conforme elas são necessárias. De fato, desde que todos os pacotes (meta-modelos) são baseados no MOF, qualquer modelo definido em termos do MOF pode ser mapeado no IDL, uma Interface Definition Language padrão CORBA. IDL é destinada para ser usada de aplicações de chamada, fornecendo assim uma fácil interface semelhante à API para repositórios de metadados compatíveis com o MOF.

Mais recentemente, padrões de metadados têm incorporado a Internet com um meio de trocar metadados padrão. Metadados tem sido a muito tempo um requisito de busca e

recuperação baseado em Web, recentemente iniciativas de padronização de metadados e meta-modelos uniram forças com a natureza experiente da utilização metadados baseado em Web.XML tem sido visto por muitos como uma boa forma de conexão entre o mundo dos metadados padrão e a acessibilidade da World Wide Web. E talvez pela primeira vez, a conexão poderia ser feita através de “tags” padrão, cada uma representando um aspecto de um meta-modelo padrão. Para uma verdadeira padronização, os valores das tags (comentários em html-`<tab name></tab name>`) também devem refletir instâncias destas construções de meta-modelo padrão. Levando a questão das tags padrão para troca padrão um passo a diante, o comitê de padrões W3C (desenvolvedores do XML) criou o XML Metadata Interchange specification ou XMI.

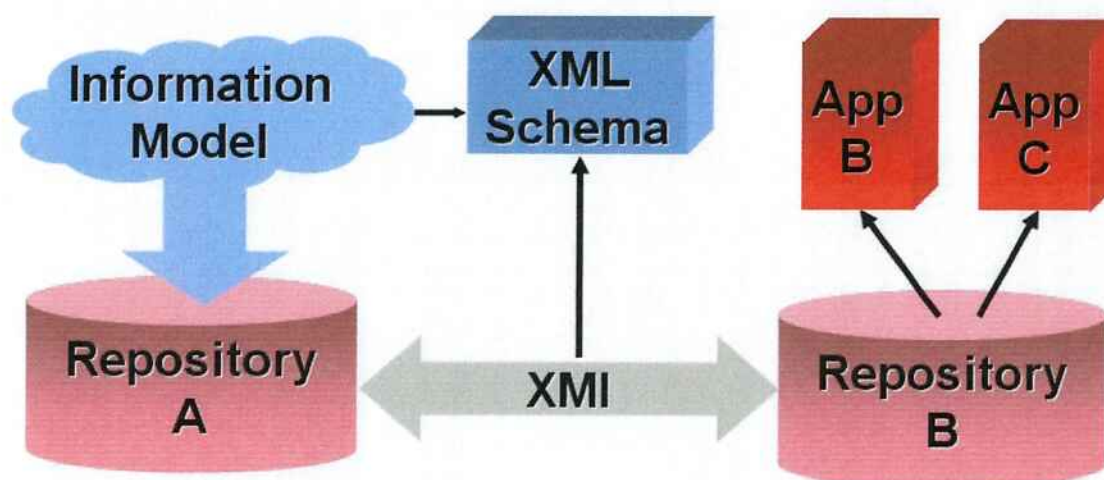
#### **3.4.7. XMI**

O propósito do XMI é a integração do XML, UML e MOF para resultar em um Intercâmbio Internet verdadeiro de não somente modelos compatíveis com o UML, mas também as instâncias destes modelos, em muitos casos metadados. Em princípio, XMI suporta a troca de qualquer tipo de metadados que podem ser expressos utilizando-se do MOF. O formato dos fluxos de metadados do XMI (normalmente arquivos) é baseado em um conjunto padrão de tags XML. Este formato proposto foi originalmente chamado de SMIF (Stream-based Metadata Interchange Format). Contudo, o advento do Data Warehouse tem melhorado o SMIF intensivamente, e o intuito do XMI é o intercâmbio de metadado sem qualquer formato, enquanto o XML for usado para representar os metadados em termos de uma instância da especificação MOF conforme figura 5. O XMI deveria ser considerado como outro formato de intercâmbio de metadados comum, e talvez o único formato traduzível. Qualquer solução de metadados ou fornecedor que pode codificar e decodificar um fluxo XMI pode trocar metadados. Além disso, se estes fluxos representam meta-modelos MOF ou qualquer porção disso, o advento de metadados padronizados poderia não estar longe.

A importância do objetivo do XMI não pode ser subestimada. O XMI está em conformidade com o MOF e cada modelo e cada modelo construído com uma instância de um MOF pode incluir suas próprias definições e relacionamentos em termos do MOF, o esforço de padronização está atualmente finalizando a habilidade de gerar XMI de um modelo MOF compatível. O resultado seriam instâncias das tags XML padrão já definidas, dessa forma garantindo de troca e interface. Exceções a estes padrões poderiam ser admissíveis, contanto que eles se conectem ao padrão em algum ponto.

Qualquer metadado que é compartilhado ou acessado de dentro da uma organização( ou entre organizações) precisaria ser representada consistentemente. Comparando identificadores implicaria em equivalência no nível meta-meta (mesmo tipo de construção de metadado), e o software de processamento saberia imediatamente que o valor era, por exemplo, uma definição.

É importante perceber que o XMI é talvez o primeiro padrão com potencial de viabilidade. Ele assume conformidade com regras arquiteturais OMG (base de modelo compatível com o MOF), mas pode ser disponibilizado contanto que o usuário leigo do MOF possa usar software que pode traduzir instâncias de metadados individuais em instâncias MOF. Com o progresso dos fornecedores, suas escolhas serão tornar o metadados de suas ferramentas compatível com o modelo MOF ou traduzir os metadados não compatíveis com o MOF em algum que possa ser relacionado a algum padrão. Este é talvez o primeiro esforço de padronização na história que requer este nível de compromisso.



**Figura 5 – Compartilhamento de Metadados usando XMI**

#### **3.4.8. Arquiteturas e Frameworks**

O OMG é uma das poucas corporações de padrões externos a abordar cada um destes componentes como parte de um framework geral. Este framework, conhecido como CORBA, é a cola que sela o relacionamento entre os modelos compatíveis com o UML, o MOF, e a especificação XMI proposta. Funciona como um componente de comunicação e permite componentes compatíveis conectem-se independente de plataforma técnica.

#### **3.4.9. CORBA**

Na arquitetura CORBA, um negociador de requisições funciona com um mecanismo para recebimento e direcionamento de requisições. A Interface Definition Language (IDL) é a linguagem padrão necessária para estas requisições, e é independente da linguagem de programação, mas contém mapeamentos para linguagens de programação populares para que as aplicação possam enviar e receber requisições para as funcionalidades CORBA. A arquitetura geral define a estrutura, interfaces, serviços e objetos. Como tal, todos os componentes da solução de metadados que formam o padrão

OMG são definidos sem uma visão restrita e representam um aspecto do padrão CORBA.

#### **3.4.10. Fornecedores**

O impacto dos maiores fornecedores de software nos padrões não pode ser subestimado. Apesar da necessidade de haver neutralidade nos padrões, a economia de software é provavelmente a maior fonte de efetividade política quando se trata de disponibilizar padrões. Soluções de Metadados eram originalmente controladas pelos fornecedores, os maiores fornecedores sendo os únicos no mercado a beneficiar-se da integração de produtos baseados em metadados. Estes fornecedores, todos com produtos distintos e desconectados, estavam interessados em usar metadados como uma forma fácil de integrar seus próprios produtos. Era uma teoria original de que um fornecedor com tanta influência tornar-se-ia o primeiro a instituir um meta-modelo embutido no produto. Como resultado, o aviso dos fornecedores e os usuários não teriam escolha a não ser aderir a ele para interfacear com o meta-modelo que se domina o mercado.

#### **Alguém realmente está seguindo estes padrões ?**

Padrões são sempre impressionantes em termos de objetivo e potencial. Contudo, quando viabilidade não é a característica principal, a chance de aderência cai substancialmente. Especialmente na área de metadados onde sabemos o potencial de um padrão de metadados bem implementado.

Avaliar a viabilidade de padrões do ponto de vista da análise de informações diárias. A facilidade de conformidade é sempre pesada contra os benefícios esperados. Se é mais fácil chegar à informação necessária, sob demanda, sem prestar atenção a qualquer padrão, então é provável que ninguém esteja realmente seguindo os padrões, tanto externo como interno. Se de outra forma, é impossível processar informação sem seguir qualquer tipo de padrão ou a quantidade de processamento customizado e o tempo necessário associado para processar conjuntos diferentes ou similares de informação

repetidamente ultrapassa a quantidade de tempo que ele leva para compatibilizar apenas uma vez.

### 3.5. Fontes de metadados

Os metadados podem ser "encontrados" em vários locais durante o desenvolvimento de um DW. Em [ADE97] alguns tipos de metadados são classificados conforme suas fontes, essas fontes e o tipo de metadados que pode ser obtido através delas são:

- Repositórios de ferramentas CASE: Normalmente os dados contidos em ferramentas CASE são estruturados o que facilita a integração automática entre a origem dos metadados e o repositório do ambiente de DW. Pode-se extrair informações sobre a origem dos dados, o fluxo dos dados (os processos que utilizam e transportam os dados), o formato dos dados e as definições de negócios;
- Documentação do desenvolvimento dos sistemas operacionais: O tipo de metadados potencialmente disponível é idêntico ao item acima. A diferença é que normalmente a documentação de desenvolvimento dos sistemas não está estruturada o que pode dificultar o entendimento das origens e fluxos dos dados;
- Código fonte dos sistemas operacionais: Quando não existe uma documentação eficiente dos sistemas operacionais, é possível extrair as informações sobre eles através dos programas fontes. Como vasculhar todos os programas de um ou vários sistemas operacionais a procura de regras é um trabalho demorado e oneroso é possível simplesmente utilizá-los como forma de esclarecer dúvidas que a documentação não contempla, também cobre os mesmos tipos de informações das fontes anteriores; e
- Entrevistas: Apesar de não ser uma fonte estruturada de informações, entrevistar profissionais da empresa que entendam do negócio, como gerentes e analistas, é de vital importância. Destas entrevistas pode-se obter regras e informações que não estão explícitas na documentação dos sistemas como, requisitos para teste dos dados e indicadores de qualidade dos dados; e

- O próprio ambiente do DW: Informações tais como frequência de acesso às informações, em que nível de agregação, tempo de resposta de cada consulta, auditoria de acesso de informações por usuários, são informações interessantes de se manter, que podem ser geradas pelo próprio sistema ao longo de sua utilização, podendo ser usadas, dentre outros propósitos, para a criação de estruturas de metadados.

### **3.6. Problemas encontrados devido à falta de Metadados**

Grandes corporações sofrem com a ausência de um ambiente de metadados. Dentre os problemas pode-se os citar:

#### **1 - Os metadados são inacessíveis ao usuário de negócio**

Na maioria dos casos, os usuários de negócio não possuem meios de acessar os metadados sobre o ambiente de informações que utilizam. Meta informações contidas no modelo de dados e na ferramenta OLAP são pouco utilizadas e, de certa maneira inacessíveis aos usuários finais.

#### **2 - Os metadados estão armazenados em ambientes distintos**

Os metadados são armazenados em repositórios de dados e em arquivos proprietários das diversas ferramentas que compõem o Warehouse. Informações que poderiam ser integradas e relacionadas, ficam pulverizadas no ambiente informacional da empresa.

#### **3 - Usuários de Negócio são dependentes da área de Tecnologia**

Por não conhecer em sua totalidade a camada semântica da ferramenta OLAP e a origem e estruturação dos dados do DW, constantemente os usuários de negócios interpelem os

técnicos e analistas da área de desenvolvimento sobre o conteúdo, a origem, a fórmula de cálculo de determinados objetos gerenciais.

#### **4 - Inexistência de Interfaces de Entrada de Metadados de ETL**

Em empresas que optaram por não adquirir ferramentas de ETL, partindo para o desenvolvimento de processos de carga proprietários, não existe interface ou aplicativo que permita o vínculo de metadados aos processos de carga., o ambiente de ETL não está documentado e, portanto, torna-se uma "área cega" do ambiente de Warehouse.

#### **5 - Dificuldade de manutenção de Relatórios e Camadas Semânticas**

A alteração de tabelas e atributos no modelo de dados, reflete nos Universos e relatórios. Quando isso ocorre, relatórios e camadas semânticas que funcionavam corretamente passam a não funcionar, ou a trazer dados incorretos para a máquina do usuário de Negócios. A necessária iteração entre as equipes de Modelagem e da ferramenta OLAP não existe, e manutenções preventivas no ambiente para prever erros deste tipo não ocorrem.

#### **6 - Dificuldade de mapeamento de dados entre o Data Warehouse e os Sistemas Operacionais**

As alterações, filtragens, resumos, conversões e alterações estruturais, que ocorrem sobre os dados dos sistemas operacionais quando são transferidos para o DW não são disponibilizados aos usuários de negócio e técnicos. Quando um gerente necessita rastrear dados do Data Warehouse em sua fonte operacional, leva-se muito tempo para se obter a resposta.

#### **7 - Sub-utilização do Ambiente de Informações Gerenciais**

Analistas de suporte à decisão, técnicos, analistas de sistemas e usuários de negócio, tem dificuldade para conhecer todo o dado armazenado no warehouse. Este conhecimento é vital para que informações valiosas ao negócio da empresa sejam extraídas do ambiente. Desta forma, um ambiente que poderia ser frutífero para a organização em termos de geração de informações, fica sub-utilizado.

#### **8 - Alterações das Estruturas e Definições de Dados não são acompanhadas**

Ao contrário dos sistemas operacionais, que não têm como preocupação o armazenamento histórico de dados e que têm como pressuposto que só há uma definição correta para os dados, o Data Warehouse necessita armazenar séries históricas que variam de 5 a 10 anos. Durante este período, a dinâmica do negócio da empresa impõe uma série de modificações organizacionais que devem ser refletidas em seus modelos dados. Estas modificações precisam ser armazenadas, e gerenciadas de maneira correta.

#### **9 - Impossibilidade de se mensurar a utilização do ambiente**

Uma das características de um Data Warehouse é a exigência de vasto espaço em disco e de janelas amplas de produção. Estes recursos são extremamente caros e escassos na empresa e, por isso, precisam ser utilizados com inteligência. Em diversas empresas não se consegue determinar se todos os atributos e tabelas do modelo estão sendo utilizados e, como consequência, os recursos de produção acima citados são utilizados em partes do modelo que não são utilizadas.

#### **10 - Visão não integrada da empresa**

Por incrível que possa parecer, grandes empresas não tem uma visão unificada sobre seus objetos corporativos. Cada unidade analisa a estrutura organizacional, o cliente e os produtos de forma diferenciada. Isso ocorre devido às periódicas mudanças dessas estruturas em razão de forças políticas ou mercadológicas. Dessa maneira, a relação

entre sistemas e unidades organizacionais distintas fica muito comprometida, acarretando muitas vezes na geração de informações incorretas para tomada de decisões, uma vez que não existem conceitos corporativos sobre objetos utilizados por toda corporação.

### **3.7. A Importância dos Metadados para a Organização**

Metadados existem em todas as funções da Tecnologia da Informação (TI). Porém uma das razões porque os esforços de metadados falham é porque metadados é considerado apenas um recurso técnico.

Em *Post Capitalist Society*, Peter Drucker escreve "o recurso básico da economia não é capital, nem recursos naturais ou trabalho. Ele será o conhecimento. O valor agora é criado pela produtividade e inovação, ambas aplicações do conhecimento para o trabalho".

Através de afirmações como esta se pode dizer que a informação é o patrimônio mais valioso da organização? Pode-se afirmar que sim, mas as pessoas ainda encontram bastante dificuldade para justificar esta afirmativa em suas organizações.

Muitas organizações não fazem o melhor uso de suas informações porque elas não são bem gerenciadas. Uma das razões porque a informação é tão mal administrada pode ser porque ela não é bem entendida. A informação não obedece às mesmas leis econômicas de outros patrimônios. Ela tem propriedades únicas que devem ser entendidas para administrá-la efetivamente. A seguir, são discutidas algumas destas propriedades.

A informação é compartilhável infinitamente. Pode-se combinar as informações de várias formas para beneficiar a organização. Porém, conhecimento é poder, e como resultado, as pessoas não compartilham informações facilmente. A restrição de informação representa uma perda de oportunidade de negócio, tanto quanto custos adicionais para a organização, com várias áreas de negócio controlando a mesma informação.

O valor da informação aumenta com o uso. Em muitas organizações existem muitas informações que poderiam ser usadas para adquirir vantagem competitiva, entretanto as oportunidades não são realizadas porque as pessoas não sabem que elas existem ou não podem acessá-las.

O valor da informação diminui com o tempo, porém ela varia conforme o tipo da informação: informações para tomada de decisão têm uma vida útil maior que informações operacionais.

O valor da informação aumenta quando combinada com outra informação. A informação geralmente é mais útil quando ela pode ser comparada e combinada com outra informação.

A informação gera outra informação. O processo de usá-la tende a resultar em mais informação.

Até hoje a transformação de dados em informação tem sido conduzida pela tecnologia ( $\text{Dados} + \text{Contexto} = \text{Informação}$ ). Estes avanços tecnológicos estão sendo integrados pela organização. Esta integração ajuda a gerenciar não só a informação, mas também formas em que as pessoas aprenderão e influenciarão a informação e a compartilharão com outros. ( $\text{Informação} + \text{Experiência} = \text{Conhecimento}$ ).

A tecnologia de metadados está surgindo em função das necessidades das organizações de conhecer melhor os dados que elas mantêm e conhecer com mais detalhes os dados de outras organizações através de Intranets e Extranets. A catalogação dos dados propiciará a maior utilização deles por usuários com múltiplos interesses. Sem uma documentação eficiente dos dados é dificultada aos usuários a localização de dados necessários para suas aplicações.

Organizações que não documentam seus dados, freqüentemente, com o decorrer do tempo, ficam sujeitas à superposição de esforços de coleta e manutenção de seus dados, vulneráveis a problemas de inconsistências e, principalmente, pagarão um alto custo pelo não uso ou uso impróprio dessa informação.

#### **4. PROPOSTA DE UM PLANO PARA CRIAÇÃO DE UMA SOLUÇÃO DE METADADOS**

Neste capítulo apresenta-se uma proposta para a criação de uma solução de metadados que pode servir de base para construção de uma solução de metadados completa. As atividades a serem desenvolvidas nessa proposta são descritas a seguir.

##### **4.1. Tipos de Solução**

A solução de metadados é sempre um benefício mal entendido em organizações de grande porte. Tipicamente, a necessidade e valor do metadado é igualmente tão mal entendida quanto, tanto a necessidade ou requisição por uma justificativa de custo de qualquer despesa que contenha a palavra “metadado”.

Para que se possa medir o retorno no investimento de qualquer solução de metadados, é essencial que a profundidade e intervalo de cobertura de solução de metadados seja claro desde o início. Muitas soluções de metadados caem em um dos seguintes 3 tipos a seguir.

##### **4.1.1. Tipo 1 – Primeira vez armazenando Metadados**

Originalmente chamado de dicionário de dados, agora expandido para cobrir metadados além das definições e nomes dos elementos armazenados fisicamente.

Estas soluções de metadados são tipicamente iniciadas pela administração de dados central, funções de administração de dados. Originalmente, eles eram comuns apenas em grandes organizações, hoje, eles estão iniciando a existir em qualquer organização que não pode encontrar ou interpretar dados existentes. Porque eles baseiam-se no conjunto de objetivos da corporação, eles estão mais voltados para requisitar justificção de custo, a não ser é claro que a obrigação de iniciar uma venha de cima.

As características distintas deste tipo de solução de metadado são o seu intuito e localização dentro da organização. Estabelecida como um significado da padronização em muitos casos, eles esforçam-se pelo uso de nomes padrões, definições, e características, todas as quais são rastreadas dentro desta solução de metadados

normalmente centralizada. Estes criadores têm controle, e são tipicamente os únicos com permissão para estabelecer e manter estes padrões para acesso de leitura através da organização.

#### **4.1.2. Tipo 2 – Suporte ao DW**

DWs tem resultado da inabilidade comum de encontrar, acessar, e interpretar dados corporativos em tempo hábil. Baseado nestas características comuns de pré-requisito, estas organizações com previsão insistem na entrega de metadados como parte da arquitetura do DW. Aquelas organizações sem previsão normalmente descobrem a necessidade por metadados quando o DW é colocado em produção.

Organizações com insucesso ou que não possuem o Tipo 1 de Solução de metadados freqüentemente criam outra solução de metadados, focada exclusivamente em metadados para suportar o DW apenas. Em muitos casos, os metadados tornam-se rastreáveis dentro de uma documentação de DW adquirida, ETL, ou produto de apoio à decisão.

Soluções do Tipo 2 de Metadados são freqüentemente a mais fácil de implementar em que eles tipicamente requerem a menor quantidade de overhead e custo adicional. Em muitos esforços do DW, a compra de ferramenta de ETL ou programa de documentação é um ganho, e todos os vendedores oferecem áreas internas para o armazenamento de metadados o qual uma vez populado são normalmente tão acessíveis quanto os dados a serem analisados. O escopo raramente se expande além da cobertura e explicação do armazém dos dados que residem dentro do armazém desde então, mesmo quando planejado, o programa de documentação do Data Warehouse é normalmente incapaz de suportar uma larga intervalo dos metadados da organização.

#### **4.1.3. Tipo 3 – A solução de Metadados**

Muito recente na intenção, esta categoria de solução de metadados é normalmente subestimada por aqueles que já tenham subestimado iniciativas das 2 primeiras categorias, e ainda não estão experimentando sucesso nos metadados, a seguir serão detalhadas as etapas para se construir esta solução que parece ser a mais bem estruturada e completa.

#### **4.2. Identificar os beneficiários dos Metadados.**

Neste passo, todos os usuários potenciais e/ou fornecedores de metadados são identificados. Beneficiários podem ser indivíduos (usuários), ferramentas, aplicações ou repositórios de metadados. Questionários e sessões de JAD (Joint Application Development) devem ser utilizados no levantamento dos requisitos de metadados.

#### **4.3. Listar os Requisitos de Metadados por Beneficiário.**

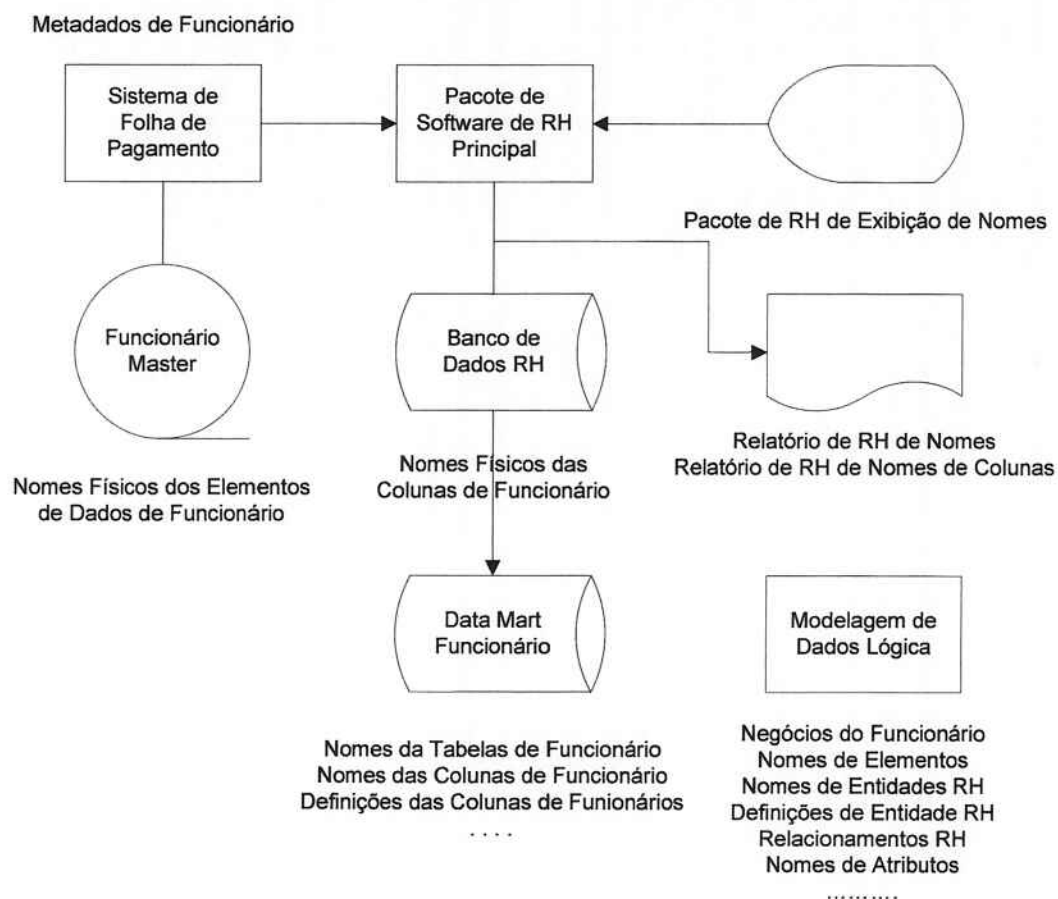
Este é o início da verdadeira análise de Requisitos, pois as instâncias de metadados são organizadas por beneficiário identificado ou categoria de beneficiário. Considere as funções dos indivíduos e como elas se relacionam a uma informação específica utilizada para completar os resultados desejados, como exemplos podemos ter :

- Desenvolvedores
- Gerentes de projetos
- Usuários finais
- Catálogo do Banco de dados

#### **4.4. Determinar a fonte de metadados.**

Determinando a fonte de cada requisito de metadados identificado, o “metadado de registro” que pode ser frequentemente nomeado(identificado). Muitos problemas de

requisitos são normalmente identificados nesta etapa, e eles são normalmente finalizados durante o escopo e esclarecimento do ambiente técnico da solução de metadados. Considere onde existe a mais precisa fonte de metadados, e considere que perspectiva é representada por suas instâncias. Isto normalmente não pode ser determinado sem pesquisar as próprias instâncias de metadados. É freqüentemente necessário desenhar um fluxo de metadados conforme a figura 6.



**Figura 6 – Fluxo de Metadados**

#### **4.5. Planejar a Arquitetura Preliminar e o Escopo dos Metadados.**

Baseado nos resultados dos passos anteriores, uma caixa é desenhada ao redor dos metadados que serão incluídos na solução inicial de metadados. A “caixa” preliminar será finalizada durante o esclarecimento do ambiente técnico. Determinar uma arquitetura preliminar neste ponto na processo de levantamento de requisitos garante que os seguintes elementos não serão ignorados:

- Ferramentas, aplicações, repositórios de metadados, ou outros pacotes de software que poderiam ter um impacto nas instâncias de requisitos de metadados identificados
- Processos que dependem da existência de instâncias de requisitos de metadados identificados
- Beneficiários de metadados, que precisam acessar instâncias de metadados.

Uma arquitetura de solução de metadados descreve os metadados da perspectiva de todos os beneficiários de metadados. Isto é, ela reflete a ferramenta/aplicação/pacote/interfaces de repositório que são necessários para manter um conjunto preciso de metadados por toda a execução contínua dos processos dependentes de metadados.

O princípio das arquiteturas de solução de metadados são os resultados dos seguintes tipos de análise:

- Fluxo de Processos. Processos relevantes são descritos, focando nos metadados identificados como necessários e em sua viagem entre processos.(De que metadados eu preciso? Que metadados eu possuo? Onde eles estão?)
- Fluxo de Metadados. Uma perspectiva de metadados descreve onde metadados necessários originariam em um cenário totalmente arquitetado tal qual onde ele se origina atualmente.(Como ele chegou lá?)

- **Metadados da Determinação de Registro.** Determinando, quando necessário, um metadado ou registro reduz a redundância de metadados em uma arquitetura de solução de metadados bem planejada. O metadado de registro serve como uma versão oficial de uma parte dos metadados. Semelhante aos dados de registro, ele deverá ser padronizado e controlado para maior abrangência possível.
- **Categorização de metadados.** Agrupar metadados por categoria melhora a eficiência de uma arquitetura distribuída de metadados. Este passo é o início da meta-modelagem. Este esforço cria uma visão organizada dos requisitos de metadados, da perspectiva da solução de metadados e sua arquitetura envolvida. Esta última categorização depende do relacionamento de cada requisito de metadados não apenas para sua fonte, mas também grupo beneficiário desejado. De fato, a demanda e disponibilidade de requisitos de metadados que ambos contribuem diretamente para a arquitetura de uma eventual solução de metadados. Categorias individuais confirmam a localização e função das instâncias de metadados no resultado arquitetural.

#### **4.6. Recategorizar os Metadados.**

Refinar os requisitos de metadados para incluir categorias convencionadas baseadas na demanda das categorias de beneficiários ajuda a categorizar todos os metadados identificados como específico, único ou comum. Esta categorização inicia o processo de meta-modelagem.

- **Metadados específicos.** Qualquer metadado que é preferido e necessário por apenas uma categoria beneficiária, é tão preciso que ele se origina de uma fonte equivalente à própria categoria beneficiária; todo metadado que é funcionalmente separado de ambas as perspectivas de arquitetura e utilização.
- **Metadados Único.** Metadados que é procurado, necessário e usado por uma categoria específica de beneficiário (como no metadados específico), mas que não se origina do mesmo componente arquitetural, ou seja, pode se originar de múltiplas fontes ou de uma fonte desconhecida, mas sempre de uma fonte

diferente.

- **Metadados Comum.** Metadados que requerem centralização e padronização com um certo controle corporativo e são utilizados por todas as categorias de beneficiários e não se encaixam nas categorias acima.

#### **4.7. Construir um Meta-modelo.**

Organizando-se os metadados identificados e categorizados forma-se a base para o alicerce do repositório de metadados. Este processo, similar ao processo de modelagem de dados, se relacionará com a arquitetura da solução de metadados inteira e ajudará a determinar onde os metadados deverão ser armazenados.

#### **4.8. Considerar o acesso e apresentação dos metadados.**

Baseado nos metadados validados (confirmados/aprovados) e nas categorias de beneficiários, tal qual a arquitetura de componentes identificada na solução de metadados geral, esta fase determina como os metadados serão acessados (inseridos, consultados, atualizados, excluídos), como eles ficarão sincronizados (se distribuídos ou múltiplos repositórios de metadados estiverem planejados), e que mecanismos de interface/apresentação serão necessários.

#### **4.9. Antecipar o processamento de Metadados.**

Neste ponto do processo, o tipo de solução ajuda a determinar os requisitos que estão faltando, ou seja, tudo que não foi contemplado ou considerado no início, mas que se torna importante devido ao processamento de metadados não humanos. Processamento específico de metadados é avaliado e uma customização da arquitetura ao nível meta-meta é especificado se necessário. Se um repositório estiver bem planejado, meta-meta-modelos devem ser avaliados segundo a extensibilidade. Se soluções baseadas em web estiverem planejadas, comentários padronizados e requisitos de troca são resolvidos nesta fase. Se soluções customizadas estiverem planejadas, fluxos de processos e requisitos são finalizados. Se pacotes de soluções estão sendo considerados, seu meta-

modelo do processo baseado em componentes deve ser avaliado contra requisitos específicos.

Considere Requisitos não diretamente ligados aos metadados. Neste ponto a solução oficialmente com o escopo determinado. Ferramentas são avaliadas no ambiente de acordo com suas funções, se houver. A arquitetura completa é aprovada e toda apresentação, acesso e processamento de metadados é revisado. Administração de Repositório, suporte técnico, e a função da nova solução de metadados nos ambientes de desenvolvimento e usuário final são finalizados.

## 5. GESTÃO DE METADADOS EM DATA WAREHOUSE

No capítulo anterior foram descritas as etapas para a criação de uma solução de metadados, mas para que uma solução de metadados seja completa ela deve ser acompanhada de um processo de administração destes metadados, neste capítulo descrevem-se as principais etapas a serem realizadas para uma gestão eficiente de metadados em um Data Warehouse.

Para iniciar o gerenciamento de metadados é necessário, em primeiro lugar, delimitar seu escopo de atuação. É muito difícil decidir quais metadados devem ser coletados e mantidos. Uma arquitetura de informação deve ser flexível para permitir um acréscimo ou decréscimo na quantidade de metadados à medida que novas necessidades surgem.

Há vários problemas ao lidar com metadados:

- Metadados tomam uma variedade de formas, especializadas ou gerais;
- Novos conjuntos de metadados irão ser criados à medida que a informação da rede fique mais madura;
- Diferentes comunidades irão propor, projetar e ser responsáveis por diferentes tipos de metadados;
- Existem muitos usuários de metadados; e
- Adoção de diferentes vocabulários de metadados significa aumento de buscas usando vocabulários de metadados que não são familiares.

As tarefas para criar um ambiente de gestão de metadados são:

- Definir requisitos para metadados que devem estar disponíveis para os usuários de Data Warehouse;
- Desenvolver a arquitetura de gestão dos metadados do Data Warehouse;
- Selecionar que ferramentas deveriam ser incorporadas na infra-estrutura de gestão de metadados;

- Desenvolver programas que integram e customizam as ferramentas selecionadas para atender as necessidades de gestão de metadados específicos da organização; e
- Desenvolver e executar um programa de treinamento para os usuários de metadados do Data Warehouse.

O projeto de gestão de metadados é mais um problema de integração de sistemas do que um esforço de desenvolvimento de aplicações. Poucas organizações constroem suas próprias ferramentas de gestão de metadados.

Por incrível que possa parecer, a maior parte das grandes empresas não têm uma visão unificada sobre seus objetos corporativos. Cada unidade analisa a estrutura organizacional, o cliente e os produtos de forma diferenciada. Isso ocorre devido às periódicas mudanças dessas estruturas em razão de forças políticas ou mercadológicas. Dessa maneira, as relações entre sistemas e unidades organizacionais distintas ficam muito comprometidas, acarretando muitas vezes na geração de informações incorretas para tomada de decisões, uma vez que não existem conceitos corporativos sobre objetos utilizados por toda corporação.

Um papel importante do Metadado é auxiliar o usuário a entender as diferenças dos dados ao longo do tempo, os quais podem ser resultados de mudanças em definições ou mesmo no negócio da empresa.

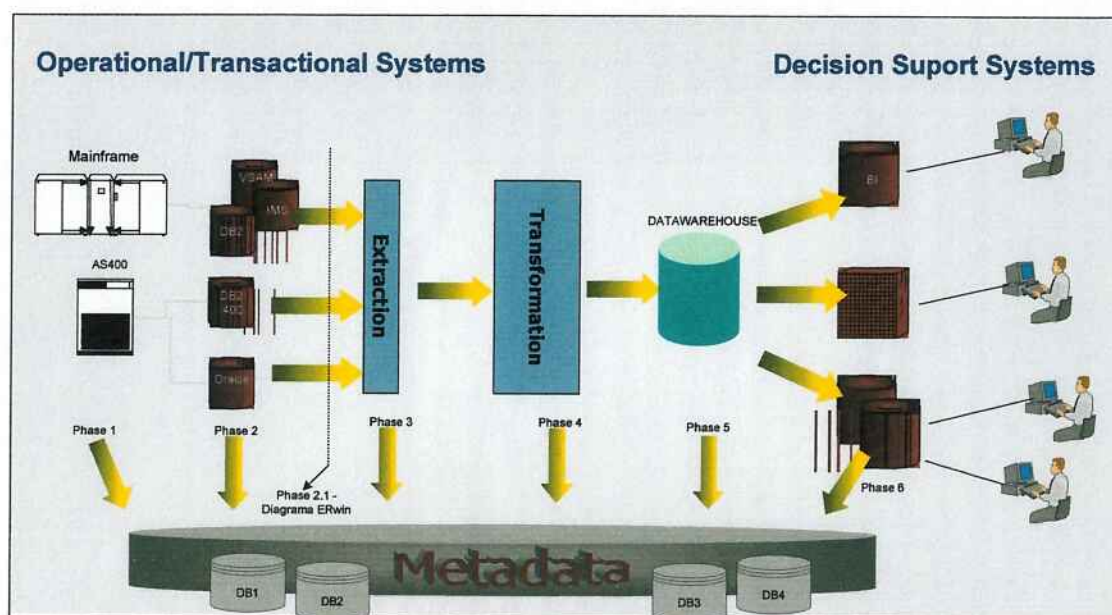
É fácil entender a necessidade do metadado para um DW, mas como avaliar quais requisitos o metadado deve procurar enfatizar, bem como quais informações devem ser armazenadas para garantir o mapeamento de metadados dos sistemas operativos (legados) para o Data Warehouse (conforme figura 7)? De acordo com a [COMPUTERWIRE], uma nova abordagem para os metadados de um Data Warehouse deve envolver os seguintes requisitos:

- Quais tabelas, atributos e chaves o Data Warehouse contém?

- Qual a origem de cada conjunto de dados? De onde eles vêm?
- Qual a transformação lógica aplicada na carga do dado?
- Como tem sido as mudanças de metadados ao longo do tempo?
- Quais aliases existem e como eles estão relacionados entre si?
- Quais as referências cruzadas entre os termos técnicos e de negócios?
- Qual a frequência com que o dado deve ser carregado?
- Qual o volume de dados existentes? Isto auxilia o usuário final a evitar submeter uma consulta irreal, por exemplo, uma consulta onde o volume de dados seja muito grande e demasiadamente demorado.

Com relação ao mapeamento entre os dados dos sistemas operativos e os dados do DW normalmente a preocupação está voltada para:

- Identificação da fonte do campo;
- Mapeamento simples atributo-atributo;
- Conversões de atributos;
- Conversões de características físicas;
- Codificação/referências das tabelas de conversões;
- Mudanças de nomes;
- Mudanças de chaves;
- Valores defaults;
- Lógica para escolher entre fontes múltiplas; e
- Mudanças de algoritmos;



**Figura 7 – Abrangência dos Metadados em Data Warehouse**

Como os Data Warehouses são, normalmente, construídos por partes é interessante observar a existência de metadados [ZACHMAN] que tratam as informações sobre o próprio Data Warehouse. Essas informações podem ser divididas em dois grandes grupos: informações estáticas e informações dinâmicas.

Em informações estáticas estão o nome do Data Warehouse, sua descrição e informações sobre a qualidade dos dados que ele contém.

Em informações dinâmicas estão relacionadas a operacionalidade do data warehouse, como por exemplo, controle de cargas, ou seja, se a carga de todas as lojas ocorreu durante a noite ou não. Informações como essas são valiosas para os analista que estão trabalhando com as informações.

Este metadado é visível pelo usuário final, quando ao acessar o sistema, verifica qual o critério de cálculo utilizado para definir as quotas de vendas de produto, por exemplo.

### **5.1. Metadados e as Mudanças em Data Warehouse**

Os usuários de negócios estão continuamente buscando por novos padrões interessantes, questionando sobre novos conjuntos de dados e comparando informações de diversas partes da empresa. Essas buscas levam a necessidade de mudanças no DW.

A chave para se lidar com as mudanças também é o Metadado. É ele que se preocupa com o que é feito a cada estágio (cada incremento) no DW e como isso se relaciona com os metadados descritos para outros sistemas.

Mudanças podem ser introduzidas em virtude da utilização de uma metodologia iterativa e incremental, mudanças nos negócios da empresa, evolução por motivos de utilização e mesmo por tempo, erros , entre outros.

Para trabalhar bem com mudanças seria necessário manter uma estratégia de versionamento e um modelo de metadados que minimizasse o impacto de mudanças.

Segundo [BARQUIN] não há nenhum conjunto de ferramentas integradas que permita unificar os processos de projeto, recuperação de dados e transformação, e o relacionamento entre sistemas origens e finais, e muito menos a habilidade de lidar com diferentes versões. A boa notícia é que muitas ferramentas usadas para implementar DW são orientadas a Metadados. A má notícia é que usam diferentes tipos de metadados e podem ou não ser capazes de mudar seus metadados através de outras ferramentas de organização.

### **5.2. Metadados para Múltiplos Data Warehouses**

Utilizar repositórios é uma forma de centralizar os metadados de uma empresa.

A forma ideal seria trabalhar no desenvolvimento do DW global (empresa) centralizado garantindo dessa forma a integração dos dados e a partir deste global tratar informações para datamarts (caso necessário).

### **5.3. Versionamento.**

Selecionar a estratégia de versionamento adequada pode fazer uma significativa diferença na redução dos custos relacionados à manutenção e iterações.

A forma mais fácil de se lidar com diferentes versões de metadados é salvá-los sempre que uma mudança for realizada. Entretanto, as abordagens existentes podem não suportar a análise de impacto.

## 6. CONCLUSÕES

A necessidade de integração de dados devido à intensa utilização de Sistemas de Informação, notadamente aqueles que fazem uso de Data Warehouse, tem alavancado a necessidade de contextualização e administração da informação juntamente com a documentação de procedimentos e processos.

Os metadados exercem papel fundamental no emprego de tais sistemas, servindo de base para integração de dados e homogeneização de conceitos técnicos e de negócios que elevaram o acesso e a disponibilização da informação a um outro patamar.

Desta forma, a proposta do plano para a criação de um modelo de metadados aqui apresentada visa proporcionar uma documentação adequada a todo o processo de criação e manutenção de dados, notadamente aqueles dados contidos em um ambiente de Data Warehouse.

As atividades contidas nessa proposta têm a capacidade de organizar o ambiente técnico representado pelos Data Warehouses, no que se refere ao detalhamento de todo o processo de obtenção dos dados, tarefa esta vital aos Sistemas de Informação atualmente utilizados.

## REFERÊNCIAS BIBLIOGRÁFICAS

MORIATY, Terry. What is Metadata? Database Programming and Design, San Mateo, v. 10, n. 7, p. 57-59, July 1997.

Inmon, W. H.; Imhoff, Claudia; Sousa, Rayan. Corporate Information Factory. New York: Wiley Computer Publishing, 1998. 274p.

Marco, David Building and Managing the Meta Data Repository. New York: Wiley Computer Publishing, 2001. 392p.

Tannembaum, Adrienne, Metadata Solutions. Boston: ADDISON-WESLEY Professional, 2002. 490p.

[BARQUIN] - Barquin, Ramon e Edelstein, Herb – Building, Using and Managing the Data Warehouse - Prentice Hall PTR - 1997.

[ZACHMAN] - Inmon, W.H., Zachman, John A. – Data Stores Data Warehousing and the Zachman Framework – Managing Enterprise Knowledge - Editora Campus - 1996.

[COMPUTERWIRE] – What is Metadata – Computerwire – 1996

[OIM] - MDC Open Information Model, Version 1.0, 1999

[SQL] - ISO/IEC 9075-2:1999, *Information technology - Database languages - SQL - Part 2: Foundation (SQL/Foundation)*, 1999

Apresentações e Materiais em powerpoint de autoria da Profa. Maria Luiza M.Campos