

MODELAGEM DA INFORMAÇÃO EM SISTEMAS DE INFORMAÇÕES COMPUTADORIZADOS PARA O REÚSO DA INFORMAÇÃO NAS ORGANIZAÇÕES

Marcio de Carvalho Victorino*
Marisa Bräscher Basílio Medeiros**

RESUMO

Este trabalho apresenta um estudo interdisciplinar de doutorado em andamento que propõe a utilização dos conceitos, métodos e técnicas da Ciência da Informação para modelar informações geradas no ambiente externo e interno às organizações. O objetivo é contribuir para que essas informações sejam compartilhadas entre diferentes sistemas transacionais e utilizadas de forma consistente pelos sistemas de apoio à decisão, proporcionando movimentos rápidos da informação entre unidades ou entre níveis diferentes da mesma unidade funcional nas organizações. A modelagem da informação proposta é fundamentada nos conceitos, métodos e técnicas de Organização da Informação preconizados pela Ciência da Informação.

Palavras-chave: Sistemas de Informação. Organização da Informação. Engenharia de Software.

1 INTRODUÇÃO

O presente trabalho propõe um processo para organizar as informações de interesse da organização, geradas interna e externamente, para que possam ser compartilhadas entre os sistemas transacionais e utilizadas de forma consistente pelos sistemas de apoio à decisão. Com isso, espera-se proporcionar movimentos rápidos da informação entre unidades ou entre níveis diferentes da mesma unidade funcional de uma organização.

Na bibliografia disponível é possível encontrar várias classificações para sistemas de informação. Laundon & Laundon (1996) apresentam uma classificação, baseada nos níveis de decisão de uma organização, composta por Sistemas de Informação Transacionais (Operacionais ou Operativos), Sistemas de Informação Especialistas, Sistemas de Informação Gerenciais, Sistemas de Apoio à Decisão, Sistemas Táticos e Sistemas Estratégicos.

Os sistemas de informação transacionais armazenam as informações necessárias para as operações diárias da organização, são utilizados por todos os funcionários para registrar e executar operações pré-definidas e normalmente possuem seus dados organizados de maneira estruturada, como, por exemplo, em tabelas de banco de dados relacionais. Estes sistemas são responsáveis por grande parte das informações requeridas pelos demais sistemas, os quais, por

*Doutorando em CI – UNB - mevictorino@unb.br

**Doutora em CI – UNB - marisab@unb.br

sua vez, produzem as informações para os sistemas no nível seguinte de complexidade. Por exemplo, os sistemas transacionais geram informações para os sistemas especialistas e gerenciais, que geram informações para os sistemas de apoio à decisão. Os sistemas de apoio à decisão geram informações para os sistemas táticos e estes geram informações para os sistemas estratégicos.

Por outro lado, os sistemas de apoio à decisão armazenam dados analíticos, destinados às necessidades da gerência no processo de tomada de decisões em uma organização. Integram dados oriundos de diversas fontes, como, por exemplo, sistemas transacionais, planilhas, arquivos textos e fontes de dados externas à organização, em repositórios multidimensionais denominados Data Warehouse (DW). Para compor um DW, os dados passam por um processo de extração, transformação e carga.

A extração de dados é o primeiro passo para a construção de um DW e inicia-se com a análise dos sistemas transacionais pré-existentes, através da leitura da documentação desses sistemas ou até mesmo através da engenharia reversa, caso a documentação seja precária. Outro passo inicial na extração de dados é a análise de dados externos que podem ser obtidos através de convênios ou comercialização.

A sugestão deste trabalho é que a organização da informação seja planejada no momento da modelagem dos sistemas de informação computadorizados através da criação de metadados para a representação da informação e de sistemas de organização do conhecimento para a representação do conhecimento. O objetivo é documentar a informação de forma consistente durante da concepção dos sistemas que a manipularão, diferentemente da engenharia de software que se restringe à criação de glossários e modelos de dados do tipo entidade-relacionamento para a documentação da informação.

2 SISTEMAS DE INFORMAÇÕES COMPUTADORIZADOS

Segundo Robredo (2004), a preocupação com a preservação de documentos iniciou-se há mais de 4.0000 anos, na Mesopotâmia. No entanto, a mudança de paradigma do documento como foco para seu conteúdo, ou seja, para a informação, aconteceu com o surgimento da Ciência da Informação.

Miranda (2002) afirma que a Ciência da Informação tem origem no fenômeno da “explosão da informação” e no esforço subsequente de “controle bibliográfico” e de tratamento da documentação implícita no processo.

Em 1962, um grupo de pesquisadores reunidos no Georgia Institute of Technology afirma que Ciência da Informação é a que investiga as propriedades e comportamento da informação, as forças que regem o fluxo da informação e os meios de processamento da informação para um máximo de acessibilidade e uso. O processo inclui a origem, disseminação, coleta, organização, armazenamento, recuperação, interpretação e uso da informação. O campo deriva ou relaciona-se com a matemática, a lógica, a lingüística, a psicologia, a tecnologia computacional, as operações de pesquisa, as artes gráficas, as comunicações, a biblioteconomia, a gestão e alguns outros campos (Robredo, 2003).

Na mesma década, Borko (1968) define Ciência da Informação como a disciplina que investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo, e os meios de processá-la para otimizar sua acessibilidade e uso. A CI está ligada ao corpo de conhecimentos relativos à origem, coleta, organização, estocagem, recuperação, interpretação, transmissão, transformação e uso de informação. Ela tem tanto um componente de ciência pura, através da pesquisa dos fundamentos, sem atentar para sua aplicação, quanto um componente de ciência aplicada, ao desenvolver produtos e serviços.

Wersig e Neveling (1975) abordam a multidisciplinalidade da Ciência da Informação e apresentam quatro enfoques para a mesma: a visão orientada para o fenômeno (fenômeno informação), a visão orientada para os meios (relacionada à prática), a visão orientada para a tecnologia (relacionada à Ciência da Computação) e a visão orientada para os fins (relacionada às necessidades sociais).

Na década de 1990, Saracevic (1992) redefine Ciência da Informação como um campo dedicado às questões científicas e às práticas profissionais voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais.

Para Bottle (1997) a Ciência da Informação é uma disciplina que investiga as características da informação e a natureza dos processos de sua transferência, que envolvem a

coleta, combinação e avaliação da informação e a organização de sua disseminação através de aparatos intelectuais e tecnologias apropriadas.

Saracevic (1999) divide a Ciência da Informação em duas grandes áreas: uma com foco na recuperação da informação; e outra com o enfoque na comunicação e no uso da informação.

Robredo (2003) define Ciência da Informação como “o estudo, com critérios, princípios e métodos científicos, da informação”. O autor ainda ressalta que a informação é indissociável de algum tipo de sistema e define sistema de informação como “uma entidade complexa e organizada que capta, armazena, processa, fornece, usa e distribui a informação”.

O início da aplicação dos computadores no processamento da informação organizacional, na década de 1950, de acordo com Davis e Olson (1985), *apud* Lyytinen (1987), causou um profundo impacto na maneira de tratar-se a informação, originando os sistemas de informações computadorizados.

Inicialmente esses sistemas eram desenvolvidos de forma artesanal, até o final da década de 1960, quando surgiu a “Engenharia de Software”, que, segundo Sommerville (2003), é uma disciplina que se ocupa de todos os aspectos da produção de software, desde os estágios iniciais de especificação até a manutenção do sistema. Técnicas e métodos de desenvolvimento de softwares foram sistematizados, dando origem às metodologias de desenvolvimento de sistemas de informação computadorizados, também denominadas metodologias de engenharia de software.

As metodologias de engenharia de software vêm evoluindo consideravelmente nos últimos tempos, partindo do modelo em cascata proposto por Royce (1970), caracterizado por ser um modelo linear, até o Processo Unificado (LARMAN, 2004), processo de desenvolvimento de software mais utilizado na década de 2000, caracterizado por ser um modelo iterativo incremental.

Ambas as estratégias de desenvolvimento apresentam sérias deficiências no que diz respeito à representação da informação. De maneira geral, as atividades da engenharia de software que possuem tarefas relacionadas à modelagem de dados ou da informação são as atividades de levantamento e análise de requisitos. Normalmente, os resultados dessas atividades são glossários, modelos de dados relativos às entidades pertencentes ao domínio modelado e uma base de dados para cada sistema de informação desenvolvido.

Nas metodologias da engenharia de software não há a preocupação com o uso de linguagens documentárias, definições consensuais, metadados ou qualquer outro recurso capaz de

abordar os aspectos semânticos da informação nas organizações. O foco da engenharia de software é a automação dos processos organizacionais por meio do desenvolvimento intensivo de software, sem que sejam contemplados aspectos relativos à modelagem da informação. Como consequência, os sistemas de informação podem se apresentar dispersos e desconectados, fazendo com que as informações, em uma organização, sejam manipuladas por sistemas de informação isolados incapazes de trocarem informações entre si.

No entanto, as organizações precisam cada vez mais acelerar o processo de tomada de decisão, devendo, para isso, reagir rapidamente às modificações do ambiente, normalmente através da análise integrada dos seus dados. Segundo Kimball *et al.* (2008), a tecnologia mais utilizada para a integração dos dados dos sistemas em uma organização tem sido o Data Warehouse, que consiste na coleção de informação corporativa, derivada diretamente dos sistemas transacionais e das fontes externas.

2.1 Data Warehouse

Segundo Inmon (1997), tido como pioneiro do conceito, DW é uma coleção de dados orientada por assuntos, integrada, não volátil e variável em relação ao tempo, que tem por objetivo dar suporte aos processos de tomada de decisão. Kimball *et al.* (2008) afirmam que DW é todo processo que provê informação para o suporte à tomada de decisão.

O objetivo de um ambiente de DW é integrar e transformar os dados gerados pelos sistemas transacionais em dados analíticos. A figura 1 apresenta a Arquitetura Técnica de um DW proposta por Kimball (1998), na qual podem ser observadas duas partes principais: a Área Interna (*back room*) e a Área Externa (*front room*).

A área interna é a parte da arquitetura de dados em que ocorre o processo de organização de dados. A principal preocupação dos responsáveis por essa área é a transferência dos dados do ponto “A” para o “B”, utilizando os serviços de extração, transformação, carga e controle do trabalho no tempo apropriado. A área interna é composta pelos seguintes componentes: Sistemas Fonte, Área de Organização de Dados e Servidor de Apresentação.

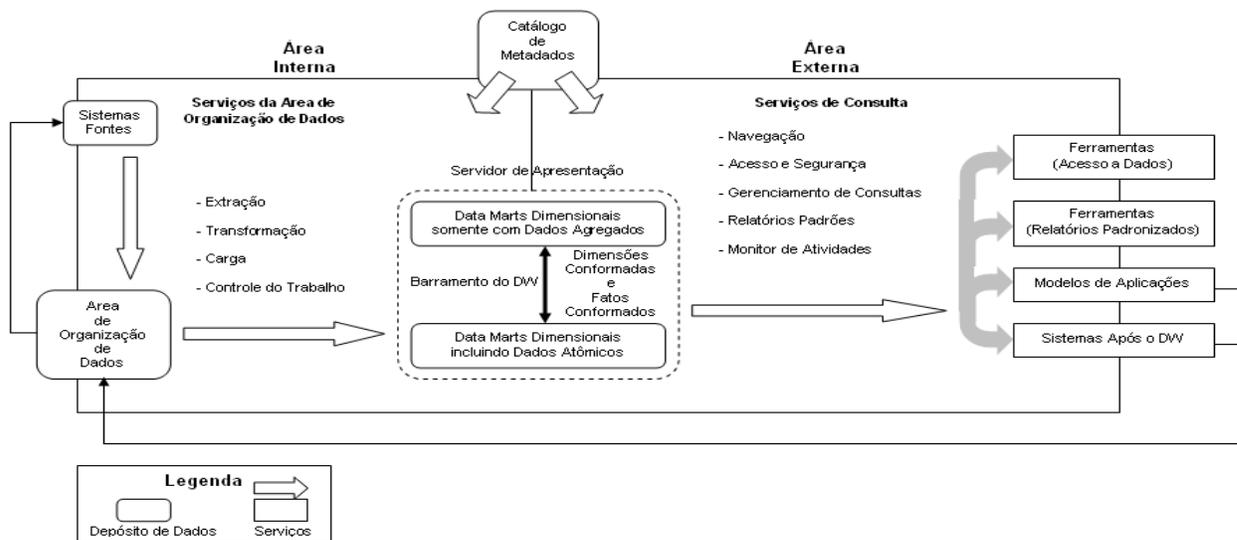


Figura 1 - Arquitetura Técnica de um DW (KIMBALL, 1998)

Os sistemas fonte correspondem aos sistemas transacionais de uma organização e/ou às fontes externas que serão integradas para compor o DW. Na grande maioria dos casos, o ambiente operacional apresenta-se não integrado devido ao fato de as aplicações terem sido construídas uma a uma, separadamente das outras aplicações.

A área de organização de dados basicamente é o local de construção do DW. A organização de dados é um processo complexo que inclui, dentre outros, os seguintes subprocessos: extração, transformação, carga e indexação. É o local onde os valores a serem adicionados ao DW são criados. Os principais tipos de repositórios de dados utilizados nessa área incluem arquivos texto, tabelas relacionais e estruturas proprietárias.

O Servidor de Apresentação corresponde às plataformas de destino da área interna, na qual os dados do DW são organizados e armazenados para consulta direta pelos usuários finais, por geradores de relatórios e por outras aplicações da área externa. Pode ser baseado em bancos de dados relacionais ou multidimensionais. O servidor de apresentação é compartilhado tanto pela área interna como pela área externa.

Esse servidor normalmente é constituído dos seguintes componentes: Data Marts (DM) somente com dados agregados, Data Marts (DM) com dados atômicos, barramento do Data Warehouse e catálogo de metadados.

Os Data Marts atômicos são aqueles que armazenam dados no mais baixo nível de detalhes necessário, possibilitando identificar a maioria das exigências do negócio, através de

consultas à base de dados analítica. Já os Data Marts agregados só armazenam dados sumarizados, a partir dos DM atômicos.

O catálogo de metadados é uma descrição geral para todo o conjunto de metadados usados no DW. Os metadados englobam todas as etapas de construção de um DW, mantendo informações sobre o que e onde está no DW.

Uma das tarefas mais complexas na construção de um DW é a extração de dados dos sistemas fonte para carga na área de organização de dados. Consiste basicamente em ler e entender as fontes de dados e copiar as partes necessárias para uma Área de Organização de Dados, a fim de serem trabalhadas posteriormente (BECKER, 1999). Na grande maioria dos DW, os dados provêm de fontes heterogêneas (planilhas, arquivos textos, banco de dados, etc). Após esse passo, os dados passam pelos processos de limpeza, transformação e integração para serem armazenadas em um ou mais repositórios que não sejam limitados por tabelas e linhas estritamente relacionais.

Data Warehouse é, portanto, um repositório de dados corporativo, no qual os dados obtidos de sistemas fontes são devidamente tratados e posteriormente depositados em bancos de dados informacionais, que oferecem um enfoque histórico para permitir um suporte efetivo à decisão.

3 ORGANIZAÇÃO DA INFORMAÇÃO

Segundo Guimarães (2006), a Organização da Informação elabora repositórios estruturados de informação e desenvolve técnicas que fornecem subsídios para evitar a criação de redes de conceitos confusas, em que os usuários gastam muito tempo navegando sem encontrar o que precisam.

Em um ambiente organizacional, a informação pode estar registrada em vários suportes, como, por exemplo, livros, relatórios, mapas, fotografias, planilhas eletrônicas e tabelas relacionais. No intuito de usar uma expressão mais ampla que “documento” para referenciar estes suportes e respectivos conteúdos, emprega-se a denominação “objetos informacionais”, conforme sugerido por Robredo (2005).

Segundo Dittrich e Domenig (1999), os objetos informacionais em mídia podem ser categorizados em três tipos: estruturados, semi-estruturados e não estruturados. Os objetos informacionais estruturados possuem uma estrutura rígida de armazenamento (planilhas

eletrônicas, tabelas relacionais, etc.), os semi-estruturados possuem estrutura de armazenamento, porém esta não é rígida (e-mail, páginas HTML, etc.) e os não estruturados não possuem estrutura de armazenamento definida além de seqüências de bytes ou caracteres (imagens, textos, etc.).

Resumidamente, pode-se afirmar que o objetivo da organização da informação é dar suporte ao fluxo de tratamento e recuperação dos objetos informacionais estruturados, semi-estruturados e não-estruturados nas organizações.

Segundo Svenonius (2000, p. 6), o ato de organizar a informação pode ser visto como um tipo particular de uso da linguagem e afirma que:

A vantagem a ser obtida por considerar o ato de organizar a informação como a aplicação de uma linguagem específica é que os *constructos* da lingüística, tais como vocabulário, semântica e sintaxe, podem ser utilizados para generalizar o entendimento e avaliar diferentes métodos de organização da informação. Outra vantagem é que esses *constructos* possibilitam a conceitualização que pode unificar métodos, antes díspares, de organização da informação – catalogação, classificação e indexação.

Taylor (2004, p. 1) lembra que organizar é uma característica básica dos seres humanos e afirma que “o aprendizado humano baseia-se na habilidade de analisar e organizar dado, informação e conhecimento”.

Bräscher e Café (2008) analisam o emprego dos termos “organização da informação” (OI) e “organização do conhecimento” (OC) em diferentes contextos e observam a falta de clareza quanto à delimitação desses conceitos. No contexto deste trabalho adota-se a proposta conceitual das autoras, na qual a organização da informação “é um processo que envolve a descrição física e de conteúdo dos objetos informacionais. O produto deste processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico”, enquanto que a organização do conhecimento “visa à construção de modelos de mundo que se constituem em abstrações da realidade”.

As autoras concluem:

Esses dois processos (OI e OC) produzem, conseqüentemente, dois tipos distintos de representação: a representação da informação, compreendida como o conjunto de atributos que representa determinado objeto informacional e que é obtido pelos processos de descrição física e de conteúdo, e a representação do conhecimento, que se constitui numa estrutura conceitual que representa modelos de mundo.

Segundo a proposta deste trabalho, o reuso da informação nas organizações é viabilizado pela representação da informação (RI) e representação do conhecimento (RC). A RI é realizada por meio de metadados relacionados aos objetos informacionais e a RC por meio de diferentes tipos de sistemas de organização do conhecimento (SOC), que são “sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles” (Bräscher; Café, 2008). Os SOC são utilizados para a modelagem do domínio em que os objetos informacionais estão inseridos.

Sendo assim, apresenta-se aqui uma abordagem que integra princípios de organização da informação e do conhecimento às metodologias de desenvolvimento de sistemas de informações computadorizados. Os princípios de OI e OC constituirão o processo denominado de “modelagem da informação”.

4 OBJETIVOS DA PESQUISA

Esta pesquisa tem por objetivo geral propor um processo de modelagem da informação para ser empregado em conjunto com as metodologias de engenharia de software utilizadas para a modelagem e desenvolvimento de sistemas de informação computadorizados. Esse processo proporcionará subsídios tanto ao compartilhamento das informações entre os sistemas transacionais, quanto à utilização das informações geradas por esses sistemas em sistemas de apoio à decisão e também à documentação das informações externas de interesse das organizações.

São objetivos específicos desta pesquisa:

- apresentar a relevância do fluxo informacional nas organizações;
- analisar as principais metodologias da Engenharia de Software utilizadas para a modelagem e o desenvolvimento de sistemas de informação transacionais;
- apresentar uma arquitetura de sistema de apoio à decisão;
- analisar os conceitos, métodos e técnicas de Organização da Informação preconizados pela Ciência da Informação, a fim de identificar aqueles aplicáveis à modelagem da informação no contexto da modelagem de processos e da modelagem de sistemas de informação;

- propor um processo de modelagem da informação a ser utilizada em conjunto com as metodologias de engenharia de software utilizadas para a modelagem e o desenvolvimento de sistemas de informação transacionais;
- validar a proposta por meio de um estudo de caso a ser implementado em uma organização pública brasileira.

5 PROCEDIMENTOS METODOLÓGICOS DA PESQUISA

Este trabalho está organizado em três etapas distintas. A primeira refere-se à pesquisa bibliográfica em busca de um aprofundamento em aspectos teóricos nas áreas de Engenharia de Software e Organização da Informação, a segunda etapa diz respeito à proposição de um processo de modelagem da informação embasado nos conhecimentos proporcionados pela etapa anterior e, finalmente, a última etapa consistirá na construção de um estudo de caso para a validação do processo proposto.

A pesquisa bibliográfica teve por objetivo buscar esclarecimentos dos principais conceitos que envolvem o tema da pesquisa.

Para o desenvolvimento do processo de modelagem da informação proposto, foi escolhido o metamodelo Unified Method Architecture (UMA) (SHUJA; KREBS, 2007).

Para avaliar a eficiência deste processo, houve a necessidade de planejar a implementação de um estudo de caso que terá por objetivo verificar a contribuição do processo de modelagem para o tratamento da informação nas organizações.

Este tipo de estudo de caso denomina-se instrumental, pois será desenvolvido com o propósito de auxiliar no conhecimento ou redefinição de determinado problema (STAKE, 2000). O autor da pesquisa não tem interesse específico no caso, tendo em vista que o processo de modelagem proposto poderá ser utilizado em qualquer tipo de organização, mas reconhece que pode ser útil para alcançar o objetivo de validar este processo.

O caso de uso validará a seguinte hipótese: **“As organizações quando utilizam um processo formal de modelagem da informação em conjunto com suas metodologias de desenvolvimento de sistemas de informação computadorizados, tendem a compartilhar suas informações entre tais sistemas e reutilizá-las em sistemas de apoio à decisão”**.

Gil (2002) afirma que não há consenso por parte dos pesquisadores quanto às etapas a serem seguidas em um desenvolvimento de um estudo de caso. No entanto, Yin (2001) e Stake

(200), *apud* Gil (2002), sugerem um conjunto de etapas a serem seguidas em uma pesquisa deste tipo, dentre elas:

- formulação do problema;
- definição da unidade-caso;
- determinação do número de casos;
- coleta de dados;
- análise dos dados; e
- preparação do relatório.

5.1 Problema da Pesquisa

O armazenamento e o acesso à informação nos ambientes organizacionais foi significativamente facilitado com a evolução da Tecnologia da Informação e Comunicação (TIC) ocorrida nas últimas décadas. No entanto, conseguir reutilizar a informação existente nesses ambientes é um problema que se agrava, principalmente nas organizações que dependem de movimentos rápidos e eficientes da informação.

Apesar de as informações geradas por um sistema muitas vezes servirem de insumo para outros, esses diferentes sistemas estão dispersos e não apresentam nenhum tipo de conexão entre si nessas organizações. Na maior parte das vezes, os sistemas de informação são desenvolvidos isoladamente, como se trabalhassem de maneira estanque, o que normalmente dificulta ou impede o reúso de informações pelos demais sistemas da organização.

Outro problema existente nas organizações consiste no fato de as metodologias de engenharia de software, utilizadas para a modelagem e desenvolvimento de sistemas de informação computadorizados, não possuírem uma estratégia para documentar os aspectos semânticos da informação manipulada. Essas metodologias vêm evoluindo consideravelmente nos últimos tempos, porém ainda apresentam deficiências no que diz respeito ao tratamento da informação, se comparadas com os métodos e as técnicas utilizados pela Ciência da Informação.

Finalmente, ainda existe a necessidade de as organizações usarem, em seus sistemas de apoio à decisão, informações geradas por diversos sistemas internos mal documentados e fontes externas de dados difíceis de serem compreendidas.

5.2 Definição da Unidade-Caso e Número de Casos

O universo onde será desenvolvida a pesquisa é o das Organizações Militares do Exército Brasileiro. A unidade-caso será apenas um órgão departamental do Exército Brasileiro.

5.3 Coleta de Dados

A coleta de dados ocorrerá em dois momentos, o primeiro em 2009, antes do processo de modelagem da informação proposto ser utilizado pelo órgão escolhido. O segundo momento ocorrerá após a utilização deste processo. Os dados estão sendo obtidos mediante análise de documentos, entrevistas, observação espontânea, observação participante, questionários e estudo de publicações e documentos internos.

5.4 Análise dos Dados e Preparação do Relatório

Pode-se afirmar que a análise de dados será de natureza qualitativa, pretende-se comparar os dados obtidos nos dois momentos citados anteriormente.

O objetivo da análise será verificar se, após a adoção do processo de modelagem proposto os tomadores de decisão conseguirão acessar a informação de seu interesse de maneira ágil, se as informações geradas pelos sistemas de informações transacionais serão compartilhadas entre tais sistemas e se essas informações serão reutilizadas em sistemas de apoio à decisão.

Após a análise será gerado um relatório com as observações e uma conclusão sobre a pesquisa.

6 O PROCESSO DE MODELAGEM DA INFORMAÇÃO E SUA DOCUMENTAÇÃO

Modelagem da informação consiste em um conjunto de procedimentos, técnicas, ferramentas e documentos auxiliares que ajudam os profissionais de informação em seus esforços para representar o domínio observado e os objetos informacionais pertencentes a este domínio. A modelagem da informação contempla tanto a descrição física – características físicas do meio e do formato em que a informação está registrada – quanto à descrição do conteúdo informacional.

Para a construção do processo de modelagem da informação proposto neste trabalho foi escolhido o metamodelo *Unified Method Architecture* (UMA) (SHUJA; KREBS, 2007), que é composto por disciplinas e fases. As disciplinas caracterizam o conteúdo do método, ou seja, documentam o que deve ser feito; as fases caracterizam a execução propriamente dita das

atividades que compõem o método.

Os principais elementos das disciplinas são: produto de trabalho, papel e tarefa. O produto de trabalho consiste em uma informação que é produzida, modificada ou usada por um processo, e que define uma área de responsabilidade e está sujeita a controle de versão. Ele pode ser um modelo, um elemento de modelo ou um documento. O papel é uma definição do comportamento de um indivíduo ou conjunto de indivíduos trabalhando em equipe. Ele especifica um conjunto de habilidades, competências e responsabilidades relacionadas e é utilizado por tarefas para especificar quem as desempenha, bem como definir um conjunto de produtos de trabalho, pelos quais são responsáveis. A tarefa é uma unidade de trabalho que um indivíduo, desempenhando um papel, pode ser chamado a realizar.

Os principais elementos das fases são: marco e iteração. O primeiro consiste em um ponto no qual termina formalmente uma iteração; corresponde a um momento em que foram alcançados objetivos significativos para a modelagem da informação. Já o segundo iteração é uma seqüência distinta de atividades com um plano criado e critérios de avaliação que resultam em um produto de trabalho. As atividades são compostas por um conjunto específico de tarefas.

As disciplinas do processo de modelagem proposto são:

- **requisitos:** descreve as atividades relacionadas ao levantamento das necessidades informacionais dos usuários;
- **análise da informação:** tem por finalidade transformar os requisitos em um conjunto de especificações de esquemas de representação da informação e do conhecimento;
- **implementação:** define as estratégias para a construção dos esquemas de representação da informação e do conhecimento;
- **validação:** identifica os passos a serem seguidos para que os esquemas de representação da informação e do conhecimento sejam validados;
- **disponibilização:** descreve as atividades que garantem que o produto da modelagem da informação será disponibilizado a seus usuários finais.

A documentação das disciplinas é materializada por meio da construção de diagramas de atividades da UML (BOOCH *et al.*, 2005), sendo que as atividades são detalhadas em tarefas, papéis e produtos de trabalho. Uma das principais vantagens dessa abordagem é permitir ao usuário do processo de modelagem da informação configurar as disciplinas de acordo com as

necessidades da organização alvo. Assim, o processo torna-se flexível tratando de maneira diferente organizações com objetivos ou porte diferentes.

As fases do processo proposto são: iniciação, elaboração, construção e transição. A meta da fase de iniciação é identificar o escopo da modelagem da informação. Na fase de elaboração, a meta é validar a Arquitetura da Informação proposta por Victorino e Bräscher (2009). Os esquemas de representação da informação serão criados na fase de construção. Finalmente, na fase de transição, os produtos gerados nas fases anteriores serão disponibilizados integralmente para os usuários que passarão por treinamento quando for necessário.

Os produtos de trabalho mais importantes resultantes do processo de modelagem da informação são o repositório informacional corporativo, metadados e os SOC, mais especificamente, tesouros, taxonomias e ontologias. Como já detalhado por Victorino e Bräscher (2009), todos estes artefatos devem ser harmoniosamente conectados.

Os tipos de metadados a serem registrados são administrativos, estruturais, descritivos, técnicos e de negócio. Os administrativos proporcionam o gerenciamento e a manutenção do registro dos objetos informacionais, enquanto os estruturais descrevem o suporte físico do objeto informacional que está sendo descrito. Já os metadados descritivos descrevem as características intelectuais do conteúdo de um objeto informacional e o relacionamento do objeto com os conceitos do domínio modelado e os metadados técnicos completam a descrição das estruturas dos objetos informacionais estruturados, como, por exemplo, tabelas, campos, tipos de dados. Por fim, os metadados de negócio descrevem o conteúdo dos objetos informacionais que contenham informações sobre o negócio de maneira que o usuário final possa entender.

Os tesouros são utilizados para permitir ao usuário encontrar o termo que represente um determinado significado para o que procura. Enquanto as taxonomias navegacionais, à semelhança dos diretórios *web*, são utilizadas para permitir que os usuários leigos naveguem pelo conteúdo do repositório e, por esse motivo, são criadas levando em conta o comportamento do usuário. Por outro lado, as taxonomias descritivas auxiliam os especialistas em suas buscas por informações. Finalmente, as ontologias permitem o aprimoramento das buscas realizadas pelos usuários com a delimitação do contexto.

O processo de modelagem da informação não deve ser entendido como algo isolado e estanque. Na verdade, o processo proposto objetiva gerar uma infra-estrutura composta por

artefatos que sofrerão freqüentes atualizações para acompanhar a evolução do ambiente informacional das organizações.

A escolha do momento mais apropriado para a execução das tarefas relativas à modelagem da informação dependerá do ambiente organizacional. No entanto, a sugestão da abordagem proposta é que a modelagem da informação seja realizada em paralelo à modelagem dos sistemas transacionais e decisórios.

7 O COMPARTILHAMENTO E O REÚSO DA INFORMAÇÃO

O processo de modelagem da informação proposto proporcionará subsídios ao compartilhamento das informações entre os sistemas de informação transacionais e ao reúso das informações geradas por esses sistemas em sistemas de apoio à decisão.

Como já foi citado anteriormente, este trabalho propõe que a modelagem da informação seja feita em conjunto com as metodologias de modelagem de sistemas. Isso se deve ao fato de estas metodologias não darem a atenção devida às informações utilizadas pelos sistemas de informações transacionais.

As metodologias de engenharia de software utilizadas para o desenvolvimento de sistemas transacionais adotam a notação Entidade/Relacionamento (ER) (CHEN, 1976) e UML (BOOCH *et al.*, 2005) e documentam os objetos informacionais, normalmente tabelas relacionais, por meio de glossários, diagramas ER e modelos de classes.

Este trabalho sugere que os esquemas de representação da informação e do conhecimento, representados por metadados e SOC, sejam criados, consultados e atualizados durante a modelagem dos sistemas de informação, proporcionando, assim, a localização de um objeto informacional a partir de suas características físicas e/ou semânticas. Esse recurso proporcionará o compartilhamento de um objeto informacional existente e evitará a criação de objetos informacionais com nomes diferentes, mas com a mesma semântica ou objetos informacionais com o mesmo nome, porém com semânticas diferentes.

A consulta e atualização dos esquemas de representação da informação e do conhecimento, antes da criação de novos objetos informacionais, têm por objetivo evitar os problemas citados que dificultam o compartilhamento e o reúso da informação nos ambientes organizacionais.

Por exemplo, poderíamos ter em uma organização duas tabelas referentes a funcionários, uma denominada “Pessoa” e outra “Funcionário”, que, apesar de possuírem nomes diferentes, representam a mesma semântica. Por outro lado, poderíamos ter em uma organização duas tabelas denominadas “Processo”, das quais uma armazena dados dos processos organizacionais e outra armazena dados dos processos judiciais da empresa, ou seja, possuem semânticas totalmente diferentes, porém com o mesmo nome.

A figura 2 apresenta uma simplificação de um ambiente organizacional convencional com sistemas de informações transacionais e sistemas de apoio à decisão. Nesse ambiente, o desenvolvimento de sistemas transacionais inicia-se com o levantamento das necessidades dos usuários finais, em seguida são confeccionados diagramas ER e UML e finalmente são desenvolvidos os programas que processarão os dados, bem como também serão criadas as respectivas bases de dados, normalmente uma para cada sistema.

O desenvolvimento de sistemas de apoio à decisão inicia-se com o levantamento das necessidades dos usuários gerenciais, logo após são confeccionados modelos dimensionais, em seguida as bases transacionais internas à organização e fontes externas de dados são analisadas e documentadas por meio de metadados e, por fim, são criadas e executadas as rotinas de extração, transformação e carga (ETL) sobre os dados que carregarão o DW – o repositório dimensional criado para este fim.

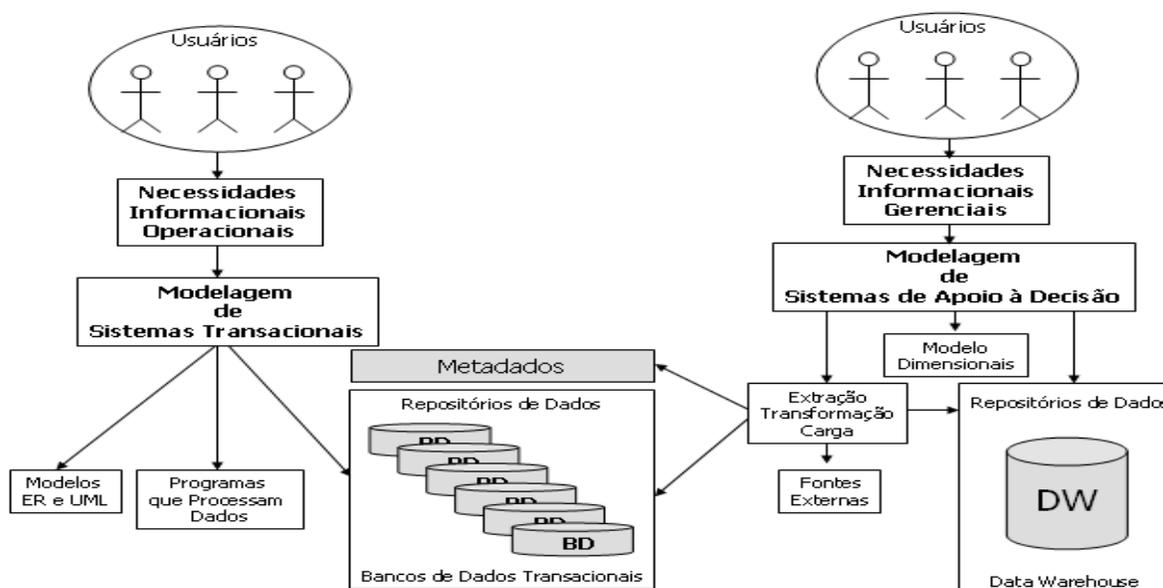


Figura 2: Ambiente Organizacional Convencional com Sistemas de Informações Transacionais e Sistemas de Apoio à Decisão

Pode-se perceber que o único recurso disponível nesse ambiente (figura 2) para o entendimento da semântica de um objeto informacional e sua localização são os metadados gerados durante a análise das bases que compõem o DW.

A figura 3 apresenta uma simplificação do ambiente organizacional com sistemas de informações transacionais, sistemas de apoio à decisão e os artefatos previstos na metodologia de modelagem da informação proposta neste trabalho.

Neste caso, além de todos os passos previstos no ambiente convencional, os metadados, os tesouros, as taxonomias e as ontologias são desenvolvidos em conjunto com a modelagem dos sistemas de informação transacionais. Ao término do desenvolvimento de um sistema, o domínio e os objetos informacionais criados estarão documentados. Quando a organização necessitar desenvolver novos sistemas, os técnicos responsáveis consultarão esses artefatos.

Essa estratégia proporciona o uso compartilhado da base de dados gerada para um sistema de informação por outros sistemas que venham a ser desenvolvidos posteriormente, gerando uma base corporativa única e integrada para todos os sistemas transacionais da organização.

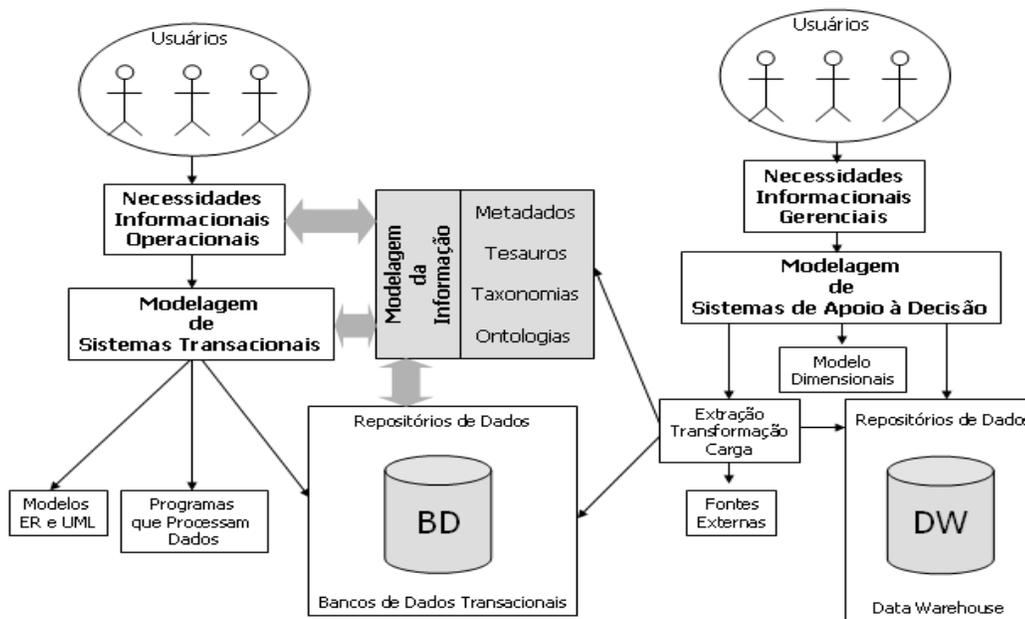


Figura 3: Ambiente Organizacional Proposto com Sistemas de Informações Transacionais e Sistemas de Apoio à Decisão

Outro aspecto importante a ser observado na figura 3 é que os produtos de trabalho gerados pela modelagem da informação, durante o desenvolvimento de sistemas transacionais, também serão consultados e atualizados durante o desenvolvimento de sistemas de apoio à

decisão, auxiliando a execução das tarefas de ETL, consideradas as mais complexas em um projeto de DW. Essa estratégia proporcionará o entendimento e, conseqüentemente, o reuso eficiente e eficaz, em um DW, dos objetos informacionais gerados pelos sistemas transacionais.

8 CONSIDERAÇÕES FINAIS

Este trabalho apresentou um processo de modelagem da informação fundamentado nos conceitos, métodos e técnicas de Organização da Informação preconizados pela Ciência da Informação.

Tendo em vista o dinamismo das informações, esse processo de modelagem foi concebido para ser realizado em conjunto com as metodologias de desenvolvimento de sistemas de informação computadorizados para não se tornar uma coisa estanque e isolada incapaz de acompanhar as transformações nos ambientes organizacionais.

Essa proposta visa aproximar esforços metodológicos das áreas de ciência da informação e engenharia de software para que o tratamento da informação seja o mais amplo possível.

Os principais produtos de trabalho gerados são os esquemas de representação da informação e do conhecimento e o Repositório Informacional Corporativo. Os esquemas de representação do conhecimento contribuem para a descrição semântica de um domínio, enquanto que os esquemas de representação da informação são utilizados para a descrição do suporte e do conteúdo dos objetos informacionais armazenados no Repositório Informacional Corporativo.

MODELING OF INFORMATION IN COMPUTER INFORMATION SYSTEMS FOR REUSE OF INFORMATION IN ORGANIZATIONS

ABSTRACT

This work presents an inter-disciplinary doctorate study in progress that proposes the use of Information Science concepts, methods and techniques to model information generated in external and internal organizations environment. The purpose is contribute so that those information are shared among different transactional systems and used in a consistent way for the decision support systems, providing fast information movements among units or between different levels of the same functional unit within organizations. The proposed information modeling is based on concepts, methods and techniques of Information Organization – all of them based on the principles of Information Science.

Keywords: Information Systems; Information Organization; Software Engineering.

REFERÊNCIAS

- BECKER, K. e, PEREIRA, W. A. L. Data Warehouse: arquitetura funcional e ferramentas de apoio ao projeto e implementação. In: XIV SIMPÓSIO BRASILEIRO DE BANCO DE DADOS (SBBD99), Florianópolis, Brasil, outubro 1999.
- BOOCH, G.; RUMBAUGH, J.; JACOBSON, I. **UML: Guia do Usuário**. 2. ed. Rio de Janeiro: Campus. 2005. 474 p.
- BORKO, H. Information science: what is it? *American Documentation*, Jan. 1968.
- BOTTLE, R.T. **Information science**. In: John Feather and Paul Sturges, editors. *International encyclopedia of information and library science*. London: Routledge.
- BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo, **Anais...** São Paulo: ANCIB, 2008. Disponível em: < <http://www.enancib2008.com.br> >.
- CHEN, P., P. “The Entity-Relationship Model - Toward a Unified View of Data”. **ACM Transactions on Database Systems** 1 (1), pp. 9-36, 1976.
- DITTRICH, K. e, DOMENIG, R. Towards Exploitation of the Data Universe. In: 3RD INTERNATIONAL CONFERENCE ON BUSINESS INFORMATION SYSTEM, abril 1999.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas. 2002. 171 p.
- GUIMARÃES, J. A. C. Interfaces Hipertextuais para a Representação da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 7, 2006, Marília, **Anais...** Marília: ANCIB, 2006. Disponível em: <<http://www.portalppgci.marilia.unesp.br/enancib/viewpaper.php?id=230>>.
- INMON, W. H. **Como Construir o Data Warehouse**. Rio de Janeiro: Campus, 1997. 388 p.
- LARMAN, C. **Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development**. 3.ed. New York: Prentice-Hall, 2004.
- LAUNDON, K C.; LAUNDON, J. P. **Management Information Systems. Organization and Technology**. 4. ed. São Paulo: Prentice-Hall, 1996.
- LYYTINEN, K. Different Perspectives on Information Systems: Problems and Solutions. **ACM Computing Surveys**. Vol.19(1):5-45, 1987.
- KIMBALL, R., REEVES, L., ROSS M. e, THORNTHWAITE, W. **The Date Warehouse Lifecycle Toolkit**. New York: John Wiley & Sons, 1998. 771 p.
- KIMBALL, R.; ROSS M.; THORNTHWAITE, W.; MUNDY J.; BECKER, B. **The Date Warehouse Lifecycle Toolkit**. 2. Ed. New York: John Wiley & Sons, 2008. 636 p.

MIRANDA, A. A ciência da informação e a teoria do conhecimento objetivo: um mal necessário. In: AQUINO, Mirian de Albuquerque, (org). **O Campo da ciência da informação: gênese, conexões e especialidades**. João Pessoa, Editora Universitária/UFPb, 2002. P.9-24.

ROBREDO, J. **Da Ciência da Informação Revisitada aos Sistemas Humanos de Informação**. Brasília: Thesaurus, 2003. 245 p.

ROBREDO, J. Organização dos documentos ou organização da informação: uma questão de escolha. **DataGramaZero** - Revista de Ciência da Informação - v.5 n.1 fev/04. Disponível em: <http://www.dgz.org.br/fev04/F_I_art.htm>. Acesso em: 10 abr. 2009.

ROBREDO, J. **Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas**. 4. ed. Brasília: Reproart, 2005. 409 p.

ROYCE, W. W. **Managing the Development of Large Software Systems: Concepts and Techniques**. Proc. WESCON, ago. 1970.

SARACEVIC, T. Information science: origin, evolution and relations. In: Conference on Concepts of Library and Information Science, historical, empirical and theoretical perspectives, ed. P. Vakkari and B. Cronin:5-27. Tampere, Finland, 1991. London: Taylor Graham, 1992.

SARACEVIC, T. Information science. *Journal of the American Society for Information Science*, p. 1051-1063, 1999.

SHUJA, A. K.; KREBS J. **IBM Rational Unified Process Reference and Certification Guide**. Boston: IBM Press Pearson plc, 2007. 307p.

SOMMERVILLE, I. **Engenharia de Software**. 6. ed. São Paulo Addison Wesley, 2003. 592 p.

STAKE, R. E. Case studies. In: DENZIN, N; LINCOLN, Y. (Ed.). **Handbook of qualitative research**. 2 ed. Thousand Oaks: Sage, 2000.

SVENONIUS, E. **The Intellectual Foundation of Information Organization**. Boston: MIT Press. 2000. 255 p.

TAYLOR, A. G. **The Organization of Information**. London: Libraries Unlimited. 2004. 417 p.

VICTORINO, M. C.; BRÄSHER, M. Organização da Informação e do Conhecimento, Engenharia de Software e Arquitetura Orientada a Serviços: uma Abordagem Holística para o Desenvolvimento de Sistemas de Informação Computadorizados. **DataGramaZero** - Revista de Ciência da Informação - v.10 n.3 jun/09. Disponível em: <http://www.dgz.org.br/jun09/F_I_art.htm>. Acesso em: 8 jun. 2009.

WERSIG, G. ; NEVELING, U. **The phenomena of interest to information science**. *Information Scientist*, v. 9, n. 4, p. 127-140, Dec. 1975.