

**METADADOS DIGITAIS: revisão bibliográfica da evolução e
tendências por meio de categorias funcionais**
*DIGITAL METADATA: bibliographical review on evolution and
trends using functional categories*

Luiz Fernando de Barros Campos - lfbcampos@gmail.com
Especialista em Gestão Estratégica da Informação, Analista de Sistemas e Engenheiro
Eletricista. Mestrando em Ciência da Informação pela ECI-UFMG

Resumo

Revisão bibliográfica sobre metadados digitais no campo da ciência da informação, objetivando precisar as várias funcionalidades dos metadados. Empregando-se uma técnica de análise de conteúdo temática, constatou-se a recorrência de tecnologias como XML, RDFS, ontologias, *data warehouses*, Web Semântica, serviços Web, entre outras, e de dez categorias relativas às funções dos metadados que fundamentavam os textos examinados. Com suporte na análise temática, adotando-se uma perspectiva técnica e histórica, relacionaram-se as categorias funcionais às tecnologias e mostrou-se como as categorias são tratadas em dezenove dos artigos revisados, destacando-se as ênfases empregadas e a convergência dos temas. Notou-se que trabalhos e tecnologias que versam sobre padrões e modelos para preservação digital adotam perspectivas mais abrangentes e integradas, incorporando todas ou quase todas as categorias. Com base nos resultados, foram comentadas tendências e questões latentes percebidas na revisão bibliográfica, e sugeridas abordagens e metodologias para a análise de metadados e das tecnologias relacionadas.

Palavras-chave: Metadados digitais. Categorias funcionais de metadados. Tecnologias de metadados.

1 INTRODUÇÃO

Metadados são habitualmente definidos simplesmente como dados descrevendo outros dados. No entanto, cada vez mais, especialmente no meio digital, o conceito tem sido empregado em variados contextos que envolvem diversos propósitos e tecnologias. Com o objetivo primário de identificar as funções de metadados nesses diferentes contextos empreendeu-se uma revisão bibliográfica sobre metadados digitais abrangendo trabalhos do campo da ciência da informação. Na revisão, utilizou-se a técnica de análise de conteúdo temática. Esse procedimento, habitualmente utilizado para análise de conteúdo de comunicações (STRAUSS, 1987; VALENTIM, 2005), foi empregado para discriminar os temas relativos às funcionalidades de metadados nos

trabalhos revisados e as relações entre eles. Como resultado, discerniram-se dez temas, denominados categorias funcionais dos metadados, que ressaltam diferentes ênfases em sua utilização, assim nomeadas: Descritividade, Padronização, Interoperabilidade, Modularidade, Reflexibilidade, Visibilidade, Flexibilidade, Automatização, Administração e Preservação. Ao mesmo tempo, constatou-se a recorrência de procedimentos, padrões, modelos, linguagens e projetos (que foram denominados tecnologias) como catalogação digital, XML, RDF, XMLS, RDFS, ontologias, *data warehouses*, Web Semântica, Serviços Web e outros.

Este texto tem por objetivo mostrar como as categorias funcionais estão relacionadas às tecnologias citadas anteriormente e como elas são abordadas em alguns artigos do campo da ciência da informação, destacando as ênfases empregadas e a convergência dos temas e possibilitando, assim, o discernimento de algumas tendências no tratamento do assunto. O artigo estrutura-se da seguinte forma: na seção 2 são comparadas algumas definições de metadados. Em seguida, na seção 3, descrevem-se atributos, funções e uma classificação dos metadados. As dez categorias são então apresentadas em blocos que contêm suas definições e descrições das tecnologias que melhor as representam (seções 4 a 9). As categorias são então utilizadas para codificar 19 textos selecionados da revisão bibliográfica e os resultados são analisados (seção 10). Essa seção retrata o próprio processo de discriminação das categorias funcionais e visa possibilitar uma compreensão de como a literatura do campo da ciência da informação tem ressaltado, sob diversos prismas, o tópico dos metadados e tecnologias afins. Na seção final, com base nos resultados obtidos, são comentadas tendências e questões latentes percebidas na revisão, e sugeridas abordagens e metodologias para a análise de metadados e das tecnologias relacionadas.

Cabe aqui esclarecer alguns pressupostos que fundamentam a exposição. O primeiro é que descritividade e representação estão intrinsecamente relacionadas nas aplicações de tecnologias digitais. Linguagens como HTML e XML representam um documento descrevendo-o. Uma página HTML é o documento e sua descrição. Como detalhado a seguir, na seção 4.1, as *tags* descritivas HTML delimitam trechos do conteúdo do documento, funcionando como metadados internos, na nomenclatura de Gilliland-

Swetland (2000). Em relação a XML, Furgeri (2006, p.238) expressou o raciocínio do seguinte modo: “...a XML passou a permitir que a própria estrutura do documento se transformasse num metadado”. O segundo ponto é que as categorias não são estanques. As tecnologias podem facilmente referir-se a mais de uma categoria, justamente por causa de suas inúmeras aplicações. Em última instância, todo metadado descreve algum objeto. No entanto, descrevem esse objeto para fins variados. Por exemplo, RDF descreve recursos, ou melhor, descreve semântica de dados que seja processável por máquinas (FENSEL, 2004, p. 21). Tendo em vista essa funcionalidade inerente de representar metadados, ela foi classificada na categoria modularidade (em contraste, por exemplo, com HTML, que basicamente descreve estrutura de documentos). Do mesmo modo, Dublin Core é claramente um conjunto de simples metadados descritivos. No entanto, enfatiza-se, no caso, o considerado mais fundamental: a emergência e consolidação do padrão (categorias de padronização e interoperabilidade). Tenha-se em mente, então, que as categorias funcionais aplicam-se a virtualmente todas as estruturas de metadados, em maior ou menor grau, especialmente, as categorias Descritividade e Padronização (isso se evidencia na discussão do METS, uma estrutura de metadados de preservação digital, na seção 9.2, ou na seção 8.3, que se refere à Web Semântica). O terceiro ponto advém diretamente dos argumentos anteriores. As descrições efetuadas por metadados estão diretamente relacionadas ao processamento por máquinas. De fato, a *representação de conhecimento*, especialmente na Internet, é usualmente a construção de metaestruturas descritivas que possibilitem o intercâmbio e processamento automáticos de objetos digitais para fins humanos (lembre-se que, na literatura, conhecimento é constantemente definido em termos de ação (SVEIBY, 1998; POLANYI, 1967). Procurou-se ilustrar o argumento na seção referente às categorias de flexibilidade e automatização, especialmente ao se cotejar as diferentes definições de ontologias, efetuadas em níveis taxonômico e lógico. Por fim, ressalte-se que na exposição buscou-se distinguir princípios e tendências gerais, procurando-se adotar uma abordagem histórica (cuja importância é destacada ao se comentar a evolução dos sistemas transacionais na seção 7.1) e observadora das práticas em domínios diversos que, em última instância, definem o conhecimento e os diferentes objetivos.

2 UMA DEFINIÇÃO DE METADADOS

O conceito de metadados como “*dados sobre dados*” tem sua origem na ciência da computação. Gill (2000) afirmou que essa definição tem levado à utilização do termo metadados onde o termo dados teria sido anteriormente considerado suficiente (como em descrições de pessoas ou objetos), acarretando confusões e mal-entendidos. Vellucci (1998) destacou que o suporte nessa definição pode levar à idéia de que metadados são simplesmente dados catalográficos ou bibliográficos com outro nome. Assim, aqui destacam-se definições mais específicas e próximas das tendências de conceituar dado e informação valorizando o contexto e o usuário.

O *World Wide Web Consortium* – W3C define metadados como informação sobre objetos da Web compreensível por máquinas. A ênfase é no processamento automático, mas falta melhor especificação do tipo e finalidade das informações sobre os objetos da Web. Dempsey e Heery (1997, p. 5) conceituaram metadados como “*dados associados com objetos que desoneram os usuários potenciais de ter conhecimento completo antecipado da existência e características desses objetos*”. Agora o foco concentra-se na finalidade: o metadado é, de certa forma, uma “economia” informacional. Gilliland-Swetland (2000, p. 1) adotou uma definição abrangente: metadado é “*a soma total do que pode ser dito sobre algum objeto informacional em algum nível de agregação*”. A autora entende objeto informacional como um item ou grupo de itens que pode ser manipulado ou endereçado, independente de tipo ou formato, como um objeto singular ou discreto, por um humano ou computador. Consolidando essas definições, Vellucci (1998, p. 192) empregou a conceituação de metadados digitais aqui adotada por privilegiar o uso, o ambiente, a estrutura e as finalidades, e ainda sugerir sua categorização:

(...) dado que descreve atributos de um recurso, caracteriza suas relações, apóia sua descoberta e uso efetivo, e existe em um ambiente eletrônico. Usualmente consiste em um conjunto de elementos, cada qual descrevendo um atributo do recurso, seu gerenciamento, ou uso.

3 CLASSIFICAÇÃO, ATRIBUTOS E FUNÇÕES DOS METADADOS

Gilliland-Swetland (2000) idealizou uma classificação de metadados digitais baseada em seus atributos, descrita a seguir. A fonte dos metadados pode ser interna (normalmente quando há geração pelo criador do objeto informacional) ou externa (geração após a criação do objeto, por outro que não o criador). Os metadados podem ser criados automaticamente por computadores ou manualmente por humanos. Podem ter uma natureza leiga ou especializada, dependendo de terem sido criados ou não por especialistas. Podem ter diferentes estados: (a) estáticos; (b) dinâmicos, sendo alterados com a manipulação; (c) de longo prazo, para garantir seu acesso e uso; ou (d) de curto prazo, para transações. Metadados podem ter uma estrutura estabelecida (como o padrão MARC) ou não ter nenhuma estrutura (simples anotações em certos campos). Sua semântica pode ser controlada, obedecendo a um vocabulário standardizado (um tesouro, por exemplo), ou não controlada (como *tags* HTML). E, finalmente, podem se referir a uma coleção ou a itens individuais.

Esses atributos são majoritariamente características estruturais intrínsecas aos metadados. Diferenciam-se das categorias funcionais empregadas, pois estas enfatizam as aplicabilidades dos metadados e sinalizam tendências de operacionalização e desenvolvimentos futuros. Vellucci (1998) apontou inúmeras funções inerentes aos metadados, especialmente no meio digital, relacionadas às categorias utilizadas, que transcendem as funções de identificação, descrição e recuperação. Metadados passam a englobar dados de diferentes tipos, como administrativos, de avaliação ou medição do conteúdo, de proveniência, de relacionamento e de estrutura, e também os termos e condições dos próprios dados. Eles certificam a autenticidade e completude do conteúdo, procuram estabelecer e especificar o contexto do conteúdo, identificam e exploram as relações estruturais internas e externas ao objeto e provêm diversos pontos de acesso para uma ampla gama de usuários.

4 A CATEGORIA DA DESCRITIVIDADE

A função principal de metadados é descrever o recurso ou objeto informacional de modo a permitir sua identificação, localização, recuperação, manipulação e uso. Pode-se considerar que cada objeto informacional apresenta quatro características principais: conteúdo, contexto, estrutura e apresentação (GILLILAND-SWETLAND, 2000; BAX, 2001). O entendimento da função descritiva do metadado pode ser refinado a partir da identificação da característica que está sendo retratada. Descrições de contexto, estrutura e apresentação possibilitam a implementação de várias funcionalidades dos metadados. No entanto, a categoria em foco enfatiza a descrição do conteúdo do objeto, usualmente conferindo-lhe vários atributos. Desse modo, está fortemente vinculada aos parâmetros da comunidade de catalogação. Considere-se que Wynar (1985) conceitua catálogo basicamente como uma lista dos itens de uma coleção. Essa lista registra, descreve e indexa os conteúdos da coleção. Um registro bibliográfico descreve intelectual e fisicamente um item e o localiza na coleção.

4.1. Linguagens utilizadas para descrição: HTML e XML

Um documento é usualmente descrito e estruturado, no meio digital, por meio de metalinguagens de marcação. Pode haver marcação procedimental ou descritiva. Na marcação procedimental, códigos são inseridos no documento. Esses códigos indicam como o texto deve ser formatado por um editor ou processador de textos. Normalmente, os padrões são proprietários. Na marcação descritiva, são utilizadas marcas ou *tags* que indicam a função ou propósito da informação no documento. Cada unidade de informação, delimitada pelas marcas, transforma-se em um objeto. Há grande estruturação da informação e as marcas são claramente separadas do conteúdo, ao contrário da marcação procedimental.

A SGML (*Standard Generalized Markup Language*), aprovada pelo padrão ISO 8879, é uma metalinguagem ou linguagem descritiva usada para a definição de um conjunto de marcas. Cada documento SGML carrega consigo uma especificação formal que define sua estrutura, um arquivo DTD (*Data Type Document*). Já a HTML (*HyperText Markup*

Language) é uma linguagem de marcação originada da SGML, constituindo um padrão predominante na WWW (*World Wide Web*) para inúmeras aplicações, incluindo os navegadores (*browsers*), editores, *softwares* de *e-mail*, servidores de bases de dados, dentre outros (BAX, 2001). Seu conjunto de marcas é pré-definido e não há uma separação entre a estrutura, a semântica e a apresentação da informação. Suas *tags* delimitam o início e o fim de título, parágrafo ou tabela de um documento, havendo *tags* próprias para cada uma dessas marcações. O W3C padroniza as versões da linguagem (em <http://www.w3.org/MarkUp/>). Por sua vez, a XML (*Extensible Markup Language*) foi concebida pelo W3C como uma simplificação da SGML (em <http://www.w3.org/XML/;http://www.w3.org/TR/REC-xml/>). Trata-se de uma metalinguagem para definir marcas e a relação estrutural entre elas, que permite descrever estruturas em árvores por meio de uma sintaxe linear. A Figura 1 ilustra um trecho de um documento XML (um registro catalográfico). Nesse código XML, *metsHdr* e *agent* são marcas (elementos). Há duas ocorrências do elemento *agent* aninhadas dentro do elemento *metsHdr*. *CREATEDATE*, *RECORDSTATUS*, *ROLE* e *TYPE* são atributos, seguidos de seu conteúdo.

Um ponto a destacar é que as marcas XML, ao contrário das HTML, não são fixas, ou seja, são convencionalmente criadas pelas comunidades de usuários. Isso aproxima a linguagem da semântica dos dados e tem ocasionado uma especialização e surgimento de padrões diversos adaptados a finalidades de cada domínio, como a linguagem *Chemical Markup Language* (CML) ou a *Mathematical Markup Language* (MathML) (BAX, 2001). A XML representa um compromisso entre as limitações da HTML e a complexidade da SGML.

```
<metsHdr CREATEDATE="2006-07-04T15:00:00"
RECORDSTATUS="Completo">
    <agent ROLE="CREATOR" TYPE="INDIVIDUAL">
        <name>Uri Geller</name>
    </agent>
    <agent ROLE="ARCHIVIST" TYPE="INDIVIDUAL">
        <name>Ana Flávia</name>
    </agent>
</metsHdr>
```

Figura 1 – Trecho de um documento XML. Adaptado de METS (2004).

5 AS CATEGORIAS DE PADRONIZAÇÃO E INTEROPERABILIDADE

A padronização é condição necessária para o entendimento das representações (estruturas, convenções, sintaxe e significado consensual dos metadados). É um compromisso ou promessa realizados de maneira cooperativa e repetitiva, resultante da aceitação de convenções. Dela resulta comunicação efetiva, reutilização e melhor compartilhamento de componentes, diminuição de custos e menos retrabalho. A interoperabilidade é a habilidade para transferir e utilizar informações entre sistemas com eficiência e uniformidade, exigindo padronização e flexibilidade em certo nível. Relaciona-se com a integração, cooperação, intercâmbio, interação e atuação em conjunto.

5.1. Padrões de metadados utilizados para a catalogação – MARC e Dublin Core

Os padrões de catalogação digital são um exemplo claro das necessidades de padronização e interoperabilidade. O MARC (*Machine Readable Catalogue*) é uma estrutura de registro eletrônico desenhada pela *Library of Congress* e a *British Library* para conter dados bibliográficos, possibilitando a comunicação em formato legível por computador e a reformatação dos registros para qualquer fim. Desenvolvida a partir de 1966, é uma das mais antigas estruturas de metadados. Um dos primeiros projetos de catalogação de recursos da Internet a utilizar o MARC foi o projeto InterCAT da OCLC (*Online Computer Library Center's*) (WOODWARD, 1996). O formato provê meios de integrar metadados em sistemas existentes. No entanto, o padrão é projetado primariamente para descrições de objetos simples e de acesso linear, adaptando-se com dificuldade a objetos em hipertextos de muitos níveis (VELLUCCI, 1997).

O padrão Dublin Core (*Dublin Metadata Core Element Set*, <http://dublincore.org/>) surgiu devido à necessidade de equilibrar a quantidade de informações constantes nos registros bibliográficos para permitir uma descrição mais rápida e barata. Mantido pela OCLC e desenvolvido por meio de cooperação internacional, é constituído de 15 elementos e diversos qualificadores. O esquema foi concebido almejando algumas

características básicas: simplicidade, interoperabilidade semântica, consenso internacional, flexibilidade e modularidade na Web. Não foi projetado com a intenção de substituir modelos mais complexos como a catalogação AACR2/MARC, mas apenas para prover um conjunto básico de elementos que pudessem ser usados por profissionais ou leigos para a descrição de recursos simples (WEIBEL, 1997).

6 A CATEGORIA DA MODULARIDADE

Ao se considerar as diversas características de um objeto digital (conteúdo, contexto, estrutura e apresentação), ele deixa de ser visto e representado como um bloco, passando a ser concebido em dimensões ou módulos. O ganho em praticidade é evidente e está vinculado à utilização de metadados direcionados para cada um desses módulos. O caso exemplar é quando uma descrição (usualmente estática) efetuada por metadados de qualquer uma das dimensões de um objeto pode ser utilizada em uma dinâmica ou ação, como a alteração da estrutura ou apresentação do objeto de acordo com contextos - este é o princípio, por exemplo, da aplicação de folhas de estilo (CSS - *Cascading Style Sheet*) sobre páginas HTML, possibilitando diferentes visualizações do mesmo documento.

6.1. Um padrão modularizado para a construção de metadados: RDF e Namespaces

No meio digital, a construção e desenvolvimento de serviços e aplicativos seguem usualmente uma estrutura modular, com linguagens e padrões sendo criados para aprimorar e expandir outros recursos já existentes. É o caso do *Resource Description Framework* (RDF, <http://www.w3.org/RDF/>), uma recomendação do W3C que objetiva padronizar a criação e uso das descrições de recursos baseados na Web, possibilitando a criação, reutilização e intercâmbio de metadados. É interessante uma comparação com a XML, cuja informação semântica está entrelaçada com estrutura em árvore do documento, já que compreendida em um contexto caracterizado pela marcação e posicionamento dos elementos. Em contraste, o RDF é um meio de adicionar semântica a um documento de maneira independente de sua estrutura, possibilitando a codificação, troca e reutilização de metadados estruturados, e permitindo, assim, com bastante

facilidade, a reutilização de elementos de ontologias ou esquemas de metadados (JACOB, 2003), isso tudo sem incorporar semântica de nenhum domínio particular nem depender de um conjunto determinado de valores. O cerne do padrão é a representação de metadados por meio de uma declaração (sentença ou tripla) que atribui um valor (V) a uma propriedade (um predicado ou atributo P) de um recurso (um objeto R). Essa declaração é expressa na forma P (R,V). Por exemplo, autor_de (“www.thefinalcut.ar”, “Alfred”), significa que o autor do sítio *www.thefinalcut.ar* é Alfred. O RDF propicia alto nível de reificação: recursos e valores são intercambiáveis e uma declaração pode ser recurso ou valor de uma outra declaração, permitindo aninhamentos. Como a XML, o RDF não provê mecanismos para declarar os nomes das propriedades que são utilizadas. Ao contrário, constitui uma forma eficiente de acessar padrões de metadados existentes por meio de *namespaces* (vocabulários controlados que identificam conceitos de maneira unívoca, evitando ambigüidades). Um documento RDF pode iniciar referenciando o *namespace* do padrão de metadados Dublin Core, ou qualquer outra especificação consensual e compartilhada, e, por meio de um prefixo, especificar recursos, atributos ou valores das declarações, utilizando o vocabulário referenciado. O processo é exemplificado na Figura 2, que mostra um trecho de um documento RDF anotado em XML (na figura, *xmlns* é uma atributo que associa um prefixo ao identificador – URL – de um *namespace*. *Title*, *creator*, *date* e *subject*, referenciados pelo sufixo *dc*, pertencem ao vocabulário do Dublin Core). A possibilidade de relacionar documentos e referenciar *namespaces* provêem uma estrutura associativa mais flexível que a XML e próxima ao humano, abrindo um leque de aplicações potencias na ciência da informação, entre elas a constituição de padrões para a elaboração de tesouros (FURGERI, 2006).

```
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<Description about="http://www.w3.org/Press/99Folio.pdf">
<dc:title>The W3C Folio 1999</dc:title>
<dc:creator>W3C Communications Team</dc:creator>
<dc:date>1999-03-10</dc:date>
<dc:subject>Web development, World Wide Web
Consortium, Interoperability of the Web</dc:subject>
</Description>
</RDF>
```

Figura 2 –Trecho de documento RDF. Exemplo extraído de <http://www.w3.org/Metadata/Activity.html>.

6.2. A construção em camadas de metadados com vistas a uma semântica: DTD e Esquemas

Assim como tecnologias diversas são dispostas em camadas para a construção de projetos na Web, cada objeto informacional é concebido e abordado em seus próprios módulos. Essas abordagens reforçam-se, com tecnologias em camada superior tratando de certas características ou atributos de objetos em uma camada inferior. Por exemplo, considerando que as marcas de linguagens como XML são livremente criadas, surgiram metamódulos para descrevê-las e a seus usos, como a especificação DTD (*Document Type Definition*) ligada a uma instância de um documento XML. O DTD contém metainformações de definições sintáticas sobre os elementos e atributos do documento XML, definindo como as marcas devem ser interpretadas por meio de várias regras, como o encadeamento de elementos, ligação de atributos a elementos ou valores permitidos para atributos. O documento XML é bem formado se obedece à sintaxe XML, atendendo condições como: começar com uma declaração XML, possuir delimitadores de início e fim para todo elemento, referenciar sua raiz, e outras. Dizer que um documento XML é válido implica em admitir que ele é bem formado, usa um DTD e respeita esse DTD.

O Esquema XML (*XML Schema*) é outra maneira de especificar as restrições na estrutura e sintaxe de documentos XML. O próprio esquema é um documento XML que possui várias vantagens em relação aos DTD: prover um rico repertório de tipos de dados para definir os valores dos elementos (permitindo derivação de dados por restrição ou extensão); criar uma maneira mais rica de descrever o aninhamento de elementos; e utilizar um mecanismo de *namespaces* para combinar documentos XML com um vocabulário heterogêneo (permitindo a reutilização de definições feitas em outro esquema). Uma desvantagem é o processamento mais complexo (FENSEL, 2004).

O papel do Esquema RDF ou RDFS (*RDF Vocabulary Description Language Schema*, <http://www.w3.org/TR/rdf-schema/>) é diferente do Esquema XML. RDFS fornece

interpretações sobre as sentenças e não restrições sintáticas ou estruturais (BRAGANHOLLO e HEUSER, 2001), definindo um modelo de dados a ser referenciado por documentos RDF e provendo mecanismos para descrição dos recursos e propriedades. Mais especificamente, o esquema estende o RDF ao definir uma linguagem declarativa e passível de processamento por máquinas que pode (a) descrever formalmente um conjunto de classes (ou tipos de recursos) e suas propriedades, (b) estabelecer relações de classes com classes e propriedades, ou de propriedades com propriedades, e (c) especificar restrições às propriedades. Por meio desses mecanismos, RDF e RDFS estabelecem uma sintaxe padronizada, um conjunto padrão de primitivas modeladoras para especificar relações (FENSEL, 2004) e uma semântica formal (JACOB, 2003), possibilitando a construção de ontologias, uma aplicação fundamental para o campo da ciência da informação.

Um exemplo simples pode ajudar a compreensão intuitiva dos mecanismos do esquema RDF. Alguns de seus fundamentos são: (a) o vocabulário do RDFS está dividido em dois *namespaces*: *rdf* e *rdfs*; (b) recursos são organizados em classes; (c) há a construção de uma hierarquia de classes e pode haver herança múltipla; e (d) uma propriedade pode assumir diversas classes dentro de um determinado intervalo. Parcialmente com base nesses princípios, fica fácil entender que a Figura 3, que ilustra trechos de um esquema RDF, significa: toda classe é um tipo de classe, animal é uma classe que tem a sub-classe cachorro e Agamêmnon é um cachorro que tem o nome (uma propriedade) “Agamêmnon”.

```
Rdfs:Class rdf:type Rdfs: Class
:Animal rdf:type rdfs: Class
:Cachorro rdf:type :Animal
:Agamêmnon rdf:type :Cachorro
:nome rdf:type rdf:Property
:Agamêmnon rdf:name “Agamêmnon”
```

Figura 3 – Trechos de um Esquema RDF em Notation3, notação alternativa ao XML para RDFS criada por Tim Berners-Lee. O exemplo é construído a partir de Palmer (2001).

7 AS CATEGORIAS DE REFLEXIVIDADE E VISIBILIDADE

Na Web, os objetos informacionais devem revelar sua estrutura e padrões de utilização. O mero funcionamento ou apresentação do conteúdo não é suficiente; devem ser criadas as condições para a flexibilização e automatização dos usos presentes e potenciais. A reflexividade é um ato de auto-orientação e descrição que acarreta melhor visibilidade dos objetos informacionais e mais versatilidade em seu uso. Também relaciona-se com a documentação, confiabilidade, facilidade de manutenção, interoperabilidade e potencialidade de usos futuros.

Desde os registros bibliográficos convencionais e sistemas legados até os sistemas digitais atuais, como Web Semântica ou Serviços Web, tem havido uma necessidade (econômica, inclusive) de utilização mais complexa e intensiva de metadados, aumentando a reflexividade e transparência. Cada vez mais, metadados, além de referirem-se aos objetos informacionais primários, voltam-se para a caracterização, descrição e representação dos próprios sistemas, regras e processos que representam, organizam e tratam esses objetos.

Pode-se perceber a necessidade do uso de metadados e sua construção em camadas em um exemplo oriundo da catalogação. Wynar (1985) chamou a atenção para a utilização de várias fichas catalográficas para criar mais de um ponto de acesso. As fichas secundárias referenciam uma ficha principal que contém os metadados referentes ao item da coleção. Por sua vez, essa ficha principal referencia as fichas secundárias, para o caso em que o registro deva ser apagado do catálogo. Por meio de referências cruzadas, foi criado um metadado que possui uma nova função (administrativa) para além da função de referenciar o item da coleção. Esse metadado é metadado do sistema de metadados (metametadado).

7.1. Dos sistemas transacionais aos *data warehouses*

A categoria de reflexividade é sinalizadora de uma tendência e mais bem compreendida em uma perspectiva histórica. Significativas na evolução dos sistemas de informação computacionais são as transformações advindas nas estruturas de dados e o conseqüente aprimoramento do uso de metadados. Nos primeiros sistemas transacionais, cada

aplicação possuía arquivos de dados próprios, devendo preocupar-se com o formato dos dados e com a criação de rotinas para acessá-los e atualizá-los. Vários problemas foram gerados nesse ambiente tradicional de arquivos. Como as aplicações eram criadas separadamente, ocorria duplicação de funcionalidades comuns, com alta carga de retrabalho e desperdício de recursos, além de muita *redundância* de dados. Esta redundância gerava a possibilidade de *inconsistência* dos dados, pois eles apareciam em pontos diferentes e eram manuseados separadamente por cada aplicação. Já que os dados eram armazenados em formatos distintos, havia problemas de *integridade*. A organização diversa dos arquivos também ocasionava complicações no acesso aos dados, criando *isolamento*. Ainda, a *segurança* ficava comprometida.

A organização em bancos de dados surgiu como uma resposta natural ao caos e à perda de produtividade gerada pelo ambiente tradicional de arquivos. A idéia básica é simples: um Sistema de Gerenciamento de Banco de Dados (SGBD) agrupa os dados em uma única localização física. As aplicações não precisam se preocupar mais com as especificidades dos dados nem em criar rotinas para manuseá-los, que agora são providas pelo próprio SGBD. Para desempenhar suas funções, um SGBD possui usualmente um modelo de dados (estruturação conceitual dos dados), uma linguagem para a definição da estrutura dos dados (forçando uma especificação formal e concentrada), uma linguagem de manipulação dos dados (para permitir sua criação, alteração e destruição) e um dicionário de dados que contém a definição dos dados, sua fonte, proprietário, formato, funções empresariais relacionadas, informações de segurança e outras. Todas essas funcionalidades convergem para o conceito de metadados. Foi exatamente nesse ponto de evolução da computação que o termo metadados (e a definição usual de “*dados sobre dados*”) consolidou-se ao aparecer mais freqüentemente na literatura sobre SGBD na década de 80 (VELLUCCI, 1998).

A situação tornou-se mais emblemática com a criação dos Armazéns de Dados (*data warehouse* – DW) na década de 90. Os DW são classicamente definidos por Inmon (1997) como “*um conjunto de dados baseado em assuntos, integrado, não volátil, e variável em relação ao tempo, de apoio às decisões gerenciais*”. Eles são criados replicando dados do ambiente transacional para dar suporte a atividades de análise e

previsão. DW variam no tempo, devido às freqüentes cargas de dados e às alterações estruturais. Há processamento e armazenamento de dados de alta granularidade (agregados).

Segundo Han e Kamber (2001), em um DW os metadados mantêm informações sobre sua estrutura, a história da migração dos dados e das transformações realizadas, o monitoramento efetuado, os algoritmos usados para as agregações, o desempenho do sistema e os próprios parâmetros do negócio, como os termos utilizados e os proprietários dos dados. Inmon (1997) também ressalta a função dos metadados em estabelecer vínculos claros com o modelo de negócio. Vê-se que os metadados no DW assumem uma outra dimensão em relação ao ambiente transacional. Como os usuários de DW usualmente apresentam diferentes perfis, quase nunca técnicos, os metadados tornam-se especialmente úteis em facilitar a navegação e o planejamento das consultas. O mapeamento das transformações dos dados do ambiente operacional para o DW é registrado pelos metadados. Outro fato relevante que acentua a função dos metadados é o longo horizonte temporal do DW. No ambiente transacional, há uma estrutura correta em determinado momento. O DW deve manter controle, com o apoio de informações descritivas, das estruturas de dados modificadas ao longo de um período de tempo. Adquire importância uma nova dimensão dos dados: o contexto, pouco estruturado e em transformação constante e rápida. A utilização de metadados é também colocada em destaque dada à presença nos DW de dados externos, usualmente pouco estruturados e de formato incomum e imprevisível, tanto em relação às fontes quanto ao momento de aquisição.

8 AS CATEGORIAS DE FLEXIBILIDADE E AUTOMATIZAÇÃO

As categorias da descritividade, modularidade, reflexividade e visibilidade viabilizam a flexibilidade e a automatização. A flexibilidade é um modo de lidar com a diversidade (existência de inúmeras estruturas e padrões) e abrandá-la, a fim de, idealmente, produzir interoperabilidade e automatização dos processos (para serem realizados com o mínimo de intervenção humana). Subjaz à discussão a pretensão de interoperabilidade

semântica, o que coloca em evidência as ontologias e a Web Semântica. A flexibilidade mantém um intercâmbio com a padronização e a interoperabilidade. Uma padronização total não seria viável e tenderia a reduzir a criatividade e inovação. Os metadados são intensamente explorados, em parte, devido à necessidade de padronizar mantendo a flexibilidade.

8.1. Os esquemas e as transformações: em direção à flexibilidade e interoperabilidade

Uma maneira de gerar flexibilidade são as transformações dos objetos digitais. Um caso elucidativo acontece com as verificações e transformações de documentos marcado em XML. Um verificador sintático (um *parser*) pode checar se um documento XML é coerente com o DTD vinculado, ou seja, se é válido, desde que incorpore em seu próprio código as regras sintáticas e estruturais expressas pelo DTD. No entanto, seria um verificador apenas para aquele tipo de documento XML. Uma vantagem da utilização do metadocumento DTD é permitir a reusabilidade do verificador sintático. Esse é um ganho significativo, mas muitas vezes é necessário transformar um documento XML. Um esquema XML permite a translação de um documento XML em outro (XML ou mesmo HTML) que utiliza um diferente DTD por meio de uma transformação XSL (XSLT). Decker *et al* (2000) lembram que essa transformação é, no fundo, trabalhosa. Não se trata de mapear gramáticas, mas sim objetos e relações entre domínios diversos de interesse. Isso implica que, no caso, deve ser efetuada uma reengenharia nos modelos de domínio originais, definindo um mapeamento entre conceitos e relações. Esses modelos de domínio são usualmente descritos em termos de objetos e relações utilizando UML (*Unified Modeling Language*) ou modelamento de entidade-relacionamento. Ir além da reutilização de componentes de programas verificadores implica utilizar RDF e RDFS, que possibilitam a representação natural de objetos e suas relações e eliminam a necessidade dos vários passos das transformações XML.

8.2. Ontologias e metadados: convergências e divergências

As ontologias são úteis na construção de sistemas automáticos e flexíveis que atuam em nível semântico e estão intrinsecamente relacionadas aos metadados. Estabelecer essa relação implica rever conceitualmente o termo ontologia. Moreira (2003) e Moreira *et al* (2004), parcialmente com base em Guarino e Giaretta (1995), destacam quatro interpretações para o termo: (1) um sistema conceitual subjacente a uma base de conhecimento; (2) um tipo especial de base de conhecimento; (3) um vocabulário usado por uma teoria lógica; (4) uma especificação de uma conceitualização.

A primeira interpretação considera ontologia como um sistema conceitual subjacente a uma base de conhecimento, não como um artefato concreto. Como não há representação simbólica, a ontologia não pode ser armazenada e operada conceitualmente. Nas outras interpretações, a ontologia é considerada denotadora de uma entidade no nível simbólico, podendo ser compartilhada e transmitida. Na segunda interpretação, a ontologia é considerada uma base de conhecimento, ou um conjunto de sentenças descrevendo o estado de um domínio por meio de uma teoria lógica. Na linha da terceira interpretação, o sentido inclina-se para um artefato sintático, sem o rigor de uma teoria formal, podendo ser definido como um vocabulário adotado em um domínio específico. Assim, uma ontologia estaria associada a uma base de conhecimento (não seria a base). O entendimento encontra consonância em Jacob (2003), que entende que todo esquema de metadados é um exemplo de ontologia que especifica um conjunto de características físicas ou conceituais de recursos que foram consideradas relevantes para uma comunidade particular de usuários, como o padrão Dublin Core. Greenberg *et al* (2003, p.17) consideram ontologias como sistemas de metadados (ou vocabulários de metadados), constituindo a “*artéria central de metadados para a Web Semântica*”. Por fim, a quarta interpretação é a consagrada e vastamente adotada definição de Gruber: “*especificação formal e explícita de uma conceitualização compartilhada*”. Formal significa que a ontologia é expressa em linguagem formal. Explícita, que é um objeto de nível simbólico. Compartilhada, que é consensual e aceita por uma comunidade de usuários. A conceitualização é uma visão abstrata e simplificada do mundo que se quer representar por algum propósito. Jacob (2003, p.20) destaca o aspecto de conceitualização parcial para criar uma definição de ontologia no ambiente da Web Semântica:

(...) uma conceitualização parcial de um certo domínio de conhecimento, compartilhada por uma comunidade de usuários, que foi definida por meio de uma linguagem formal e processável por máquinas, com o propósito específico de compartilhar informação semântica entre sistemas automatizados.

Uma contraposição entre a terceira interpretação (ontologia como terminologia) e as demais é notada de diferentes maneiras e intensidades em vários autores. O princípio geral é que as ontologias comportam dois níveis: um taxonômico, definidor de um vocabulário, e um lógico, consistindo de regras de inferência que possibilitam responder a questões e escolher cursos de ação. Assim, Chandrasekaran *et al* (1999) consideram ontologias como vocabulários - seja enfatizando a representação de fatos (conceitos) no domínio ou a especificação de um conjunto de fatos a serem compartilhados - ou como teorias de conteúdo, com ênfase nos processos ou mecanismos. Fensel (2004) chama a atenção para DTD, XMLS e ontologias terem em comum a finalidade de prover vocabulário e descrever fontes de informações visando o intercâmbio.

8.3. Um modelo em camadas para a Web Semântica

A Web Semântica não é uma aplicação, um programa ou uma implementação, mas sim um conceito, uma abordagem e um projeto genérico. Os fundamentos da Web Semântica são apontados no seminal artigo de Berners-Lee *et al* (2001). A intenção é estabelecer uma extensão da Web atual com semântica compartilhável. Isso significa criar padrões e interoperabilidade entre agentes que interagiriam entre si e com pessoas, possibilitando efetiva estruturação e utilização automática da informação já existente e dispersa na Internet. Esses agentes são entidades programadas que empregam técnicas de inteligência artificial com o objetivo de ajudar pessoas a realizar tarefas, atuando de forma autônoma e utilizando a metáfora de um assistente pessoal (SOUZA e ALVARENGA, 2004). Ao contrário de alguns projetos de Inteligência Artificial, a “inteligência” adviria da descentralização na Web semântica. As ferramentas para realizar o projeto são padrões, linguagens e recursos como XML, RDF e ontologias. A Figura 4 mostra a arquitetura da Web Semântica, na visão de Berners-Lee, que tem como base a codificação (Unicode) e identificação unívoca (URI) dos objetos

informacionais e culmina com a confiança em informação semanticamente assimilada e com a prova de sua proveniência, envolvendo a utilização intensa de metadados e das tecnologias anteriormente comentadas, que são organizadas em camadas. Greenberg *et al* (2003) ressaltam o papel fundamental dos metadados na composição da Web Semântica. Souza e Alvarenga (2004) vêem a Web Semântica aproximando-se de um grande sistema de recuperação de informações com melhoras apreciáveis da revocação e precisão devido a marcações e semântica compartilhadas, que poderiam contextualizar informações e minorar os problemas de polissemia e sinonímia. A Web Semântica é encarada como uma evolução da Web atual e é enfatizada a adoção e o compartilhamento de padrões de metadados.

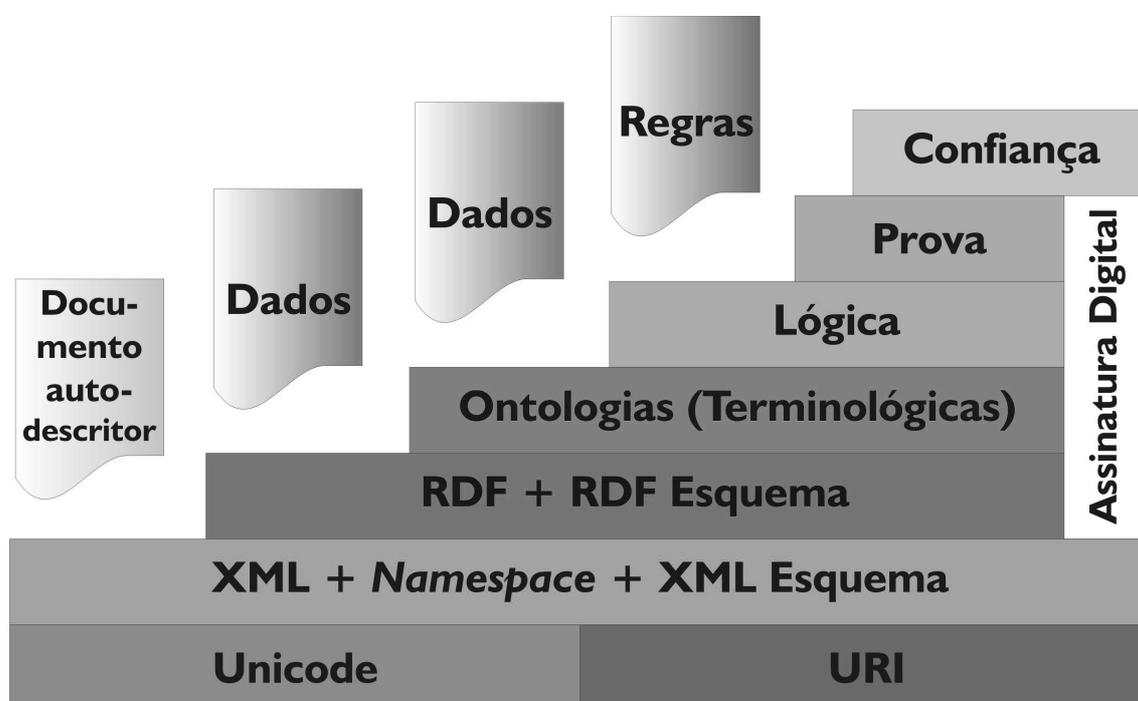


Figura 4 - Arquitetura modular da Web Semântica segundo Bernes-Lee.
Fonte: Adaptado de <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-1.html>>

9 AS CATEGORIAS DE ADMINISTRAÇÃO E PRESERVAÇÃO

Metadados referentes à administração, acesso, preservação e uso das coleções são constantemente empregados em repositórios. Isso é particularmente notável em

bibliotecas e arquivos digitais. Nesses ambientes, os metadados não apenas descrevem e identificam um objeto informacional, mas explicitam as condições corretas ou ideais de seu gerenciamento, as relações do objeto com outros na coleção, sua função, utilização, comportamento, contexto de criação e condições de preservação. À medida que a utilização da informação em ambientes digitais sofisticada-se, é esperado que haja uma ampliação no escopo das funcionalidades dos metadados. A direção que seu desenvolvimento tomará está fortemente relacionada às características da comunidade que os utiliza e às tarefas a serem desempenhadas. O certo é que a criação e a manutenção de metadados tendem a ser um componente fundamental para a maioria das estratégias de preservação digital (THOMAZ e SOARES, 2004).

A tipicidade de documentos digitais em relação aos documentos impressos revela-se na importância da preservação lógica e intelectual (já que a preservação física é fundamental para ambos). A preservação lógica está relacionada à garantia da conversão dos formatos originais dos documentos digitais enquanto a intelectual, à integridade e autenticidade das informações. A preservação intelectual é crítica devido aos materiais digitais serem passíveis de alterações em seu leiaute, apresentação ou interação no formato de publicação (ARELLANO, 2004). As necessidades de preservação intensificam-se quando se recorda a tendência da representação do objeto digital compartilhar-se nas características de conteúdo, contexto, estrutura e apresentação.

Os principais métodos de preservação dos objetos digitais são usualmente classificados em operacionais e estruturais. Os operacionais englobam a migração e conversão de formatos, a emulação (o dado é preservado em seu formato original e o ambiente original é reproduzido) e a conservação de *hardware* e *software* (normalmente bastante dispendiosa). Os estruturais envolvem os metadados para preservação, a adoção de padrões e a elaboração de normas. Os metadados de preservação são utilizados para garantir autenticidade, facilitar a identificação das fontes de informação, gerenciar e registrar direitos incidentes sobre os itens da coleção, gerenciar os fluxos dentro dos processos e representar as estruturas para possibilitar ou facilitar a busca e o acesso, entre outras funções.

9.1. OAIS: um modelo de referência internacional para metadados de preservação

Dentre os padrões para modelos de metadados de preservação destaca-se a iniciativa ISO (*International Organization for Standardization*) do modelo de referência *Open Archival Information System* (OAIS), publicado pelo *Consultive Committee for Space Data Systems* (CCSDS). Atuam quatro entidades no modelo: produtores, consumidores, administração e o arquivo propriamente dito. O modelo inclui um esquema de informação para inserção dos metadados almejando a preservação de informação digital por longo prazo. O objeto de dados, que pode existir digital ou fisicamente, é interpretado como objeto de informação por meio da combinação da base de conhecimento da comunidade-alvo com a informação de representação associada ao objeto. O objeto de informação pode ser de quatro tipos: Informação de Conteúdo, Informação de Descrição de Preservação, Informação de Pacote e Informação Descritiva. A Informação de Descrição de Preservação (IDP) serve para preservar a Informação de Conteúdo associada (junto com sua informação de representação) e consiste de: Referência (descritores), Contexto (relacionamentos com o ambiente), Proveniência (*copyright*, histórico) e Rigidez (informação para a comprovação de integridade e autenticidade). A Informação de Pacote reúne a Informação de Conteúdo e a Informação de Descrição de Preservação em um pacote identificável. Ela pode ser de submissão (produtor para arquivo), de arquivamento de informação (efetivamente armazenada) e de disseminação (arquivo para consumidor). Por fim, a Informação Descritiva facilita o acesso à Informação de Pacote por meio de ferramentas de pesquisa e recuperação. O modelo OAIS ainda inclui uma parte funcional com as entidades Recepção, Armazenamento, Gerenciamento de Dados, Administração do Sistema, Planejamento de Preservação e Acesso.

Thomaz e Santos (2003), com base no modelo de referência OAIS, sintetizam e criticam três projetos de metadados produzidos por entidades internacionais de manutenção de documentos eletrônicos de arquivos, para construir um modelo preliminar de metadados para preservação digital.

9.2. METS: estrutura de metadados para manutenção e troca de objetos digitais

METS (2004) (*Metadata Encoding and Transmission Standard*) é uma iniciativa da *Digital Library Federation* (<http://www.diglib.org/>) que procura especificar um modelo para codificar metadados descritivos, administrativos e estruturais que referenciam objetos em uma biblioteca digital utilizando o Esquema XML (XMLS). O objetivo é tanto a gestão dos objetos no repositório quanto a troca desses objetos entre repositórios, construindo um mecanismo flexível para a codificação dos metadados e para exprimir as ligações complexas entre várias formas de metadados. O modelo é mantido pelo *Network Development and MARC Standards Office* (<http://www.loc.gov/marc/ndmso.html>) da *Library of Congress*.

Um documento METS consiste de sete seções principais. O cabeçalho (seção 1) contém metadados que descrevem o próprio documento METS, abrangendo informações como o criador e o editor, data da última alteração, estado do documento, identificadores alternativos, e outros. Metadados descritivos constituem a seção 2. Eles podem ser externos ao documento (como um registro MARC) ou embutidos, e pode ocorrer em múltiplas instâncias. Metadados administrativos (seção 3) informam como os arquivos foram criados e armazenados, direitos de propriedade intelectual, referências sobre o objeto original do qual o objeto digital foi derivado e proveniência dos arquivos que compõem o objeto digital (como relações de arquivos originais e derivados, migração ou transformações ocorridas). Como os metadados descritivos, podem ser externos ou internos. A seção de arquivos (seção 4) lista todos os arquivos que contêm as versões eletrônicas do objeto digital, permitindo subdivisões por versão do objeto. O mapa estrutural, seção 5, a parte principal do documento METS, define uma estrutura hierárquica para o objeto (permitindo navegação) e liga os elementos dessa estrutura aos correspondentes arquivos de conteúdo e metadados. As ligações estruturais, sexta seção, permitem o registro das hiperligações entre nós na hierarquia esboçada no mapa estrutural. Essa seção é básica para arquivamento de sítios. A sétima e última seção do documento associa comportamentos executáveis com os conteúdos do objeto.

Nesse esquema de metadados distinguem-se todas as categorias: seu fim último é descrever objetos digitais para futura recuperação (descritividade), administrá-los e

preservá-los (administração e preservação). Outra finalidade importante é permitir o intercâmbio de objetos entre repositórios diferentes sem intervenção humana (flexibilidade, automatização, interoperabilidade). Os diferentes tipos de metadados aparecem separados (modularidade). O cabeçalho é metadado de metadados, informando sobre a própria estrutura que representa os itens da coleção (refletividade e visibilidade). E, enfim, o próprio modelo é, ou pretende ser, um padrão (padronização e interoperabilidade).

10 TRATAMENTO DAS CATEGORIAS POR ARTIGOS DO CAMPO DA CIÊNCIA DA INFORMAÇÃO

Foram selecionados 19 textos da revisão bibliográfica para indicar a maneira com que as categorias funcionais foram abordadas e salientadas, destacando a convergência das categorias com o assunto do artigo, sua frequência e agrupamento, e a abrangência do artigo. Privilegiaram-se artigos publicados em periódicos brasileiros, mas foram escolhidos 7 trabalhos internacionais devido ao interesse da abordagem e para instituir um marco de comparação. Empregou-se a técnica de análise de conteúdo temática (STRAUSS, 1987; VALENTIM, 2005) a fim de descobrir os núcleos de sentido cuja presença significasse algo para o objetivo visado, a distinção das funcionalidades de metadados. Houve uma pré-análise, a exploração e codificação dos artigos, e a interpretação dos resultados obtidos. As codificações são necessariamente subjetivas. No entanto, foram efetuadas em um processo recursivo envolvendo leituras atentas e ativeram-se aos conceitos das categorias que iam sendo determinadas, observando-se, secundariamente, a ocorrência de palavras-chave (tais como: integração, intercâmbio ou interação para interoperabilidade – outros exemplos constam da definição de cada categoria nas seções anteriores). Também é importante ressaltar que as codificações, apresentadas no Quadro 1, não pretendem avaliar a qualidade dos artigos, mas refletem apenas o tratamento dado ao tema.

Artigo	Assunto Tratado	Categorias																			
		D	P	I	M	R	V	F	A	AD	PR										
VELLUCCI (1998)	Metadados																				
WOODWARD (1996)	Catálogo na Internet																				
IKEMATU (2001)	Metadados																				
SOUZA&ALVARENGA (2004)	WEB Semântica																				
DZIEKANIAK&KIRINUS (2004)	WEB Semântica																				
DECKER et al (2000)	WEB Semântica, XML, RDF																				
BAX (2001)	Linguagens de marca																				
ALMEIDA (2002)	Linguagens de marca																				
FURGERI (2006)	Linguagens de marca																				
CHAVES et al (2001)	Ontologias																				
JACOB (2003)	Ontologias																				
BAX&LEAL (2001)	Serviços WEB																				
RYMAN (2003)	Serviços WEB																				
SOUZA et al (2000)	Dublin Core																				
WEIBEL (1997)	Dublin Core																				
ARELLANO (2004)	Preservação documentos digitais																				
THOMAZ&SANTOS (2003)	Metadados Arquivos Eletrônicos																				
THOMAZ&SOARES (2004)	OAIS																				
METS (2004)	Metadados Arquivos Eletrônicos																				
LEGENDA																					
D - Descritividade	R - Reflexibilidade	Ad - Administração																			
P - Padronização	V - Visibilidade	Pr - Preservação																			

Quadro 1 – Resultado da codificação de artigos selecionados da revisão bibliográfica.

Fonte: Desenvolvido pelo autor.

Analisam-se aqui apenas os principais resultados. Todos os artigos abordaram os metadados pelo prisma da descritividade. Isso era esperado, considerando que essa é a função básica e mais evidente de metadados. Praticamente o mesmo acontece com as categorias da padronização e interoperabilidade. Discutir metadados é em alguma medida abordar um padrão, com vista a obter interoperabilidade. Em geral, os artigos que discutem um padrão de metadados específico, como o Dublin Core, enfatizam a descritividade e a padronização, como Souza *et al* (2000), cujo foco restrito na descritividade é explicitado já no título.

Os textos mais abrangentes, incorporando quase todas as dimensões, foram os referentes a metadados para a preservação digital. METS (2004), um tutorial expositivo de um modelo de metadados para manutenção e troca de objetos entre bibliotecas digitais, englobou todas as categorias. Thomaz e Santos (2003) e Thomaz e Soares (2004) enfatizaram todas as categorias, menos a interoperabilidade, o que ocorre com frequência quando os metadados para preservação são estudados, exceto no caso de

haver menção explícita a troca de objetos entre repositórios (como é o caso do METS). Também os textos que versam sobre projetos abrangentes, como Web Semântica e Serviços Web, que utilizam metadados em camadas baseando-se em padrões e buscando incrementar a flexibilidade e automatização, tendem a gerar abordagens multifacetadas. Esse é o caso do artigo de Souza e Alvarenga (2004) que, não obstante, ressaltou menos a modularidade, visibilidade e automatização, por priorizar a descrição da Web Semântica sob a perspectiva de um sistema de recuperação de informação. Já Decker *et al* (2000) abarcou mais categorias, dentre os artigos que tratam da Web Semântica, por abordar em detalhes as combinações dos metadados e tecnologias para a construção da Web Semântica. Observe-se que os metadados só foram analisados sob o prisma da administração por artigos que tratavam especificamente do tema de preservação digital. Esses resultados sugerem que os projetos da Web, como a Web Semântica, pretendem aprimorar a comunicação e dar sentido à informação dispersa na rede, mas estão descuidando da administração e preservação dessa informação. Note-se que o mesmo ocorre com os dois artigos que versam sobre Serviços Web (RYMAN, 2003; BAX e LEAL, 2001), cujo foco na interoperabilidade não prescindiu da análise das demais categorias, com exceção da administração e preservação.

Dentre os artigos codificados, dois são revisões publicadas na ARIST (*Annual Review of Information Science and Technology*): Vellucci (1998) adotou uma abordagem bastante ampla, incorporando as novas tecnologias, como transformações entre modelos de metadados, de maneira significativa e destacando os esforços cooperativos, o que gerou uma análise rica e antecipadora das pressões das mudanças tecnológicas, ao contrário de Woodward (1996), escrito dois anos antes, que focou projetos tradicionais de catalogação e classificação transpostos para a Internet.

11 COMENTÁRIOS FINAIS

As categorias empregadas estão intrinsecamente relacionadas às funcionalidades em domínios específicos e apresentam-se interligadas. No entanto, algumas ênfases e tendências são perceptíveis. As tecnologias envolvidas na expansão da Web, ao requerer interoperabilidade semântica, ensejam uma utilização matizada de metadados, envolvendo aspectos como modularidade, reflexibilidade, visibilidade, flexibilidade e

automatização. Assim, essas categorias são valorizadas por artigos que versam sobre linguagens de marcas, Serviços Web, Web Semântica, ontologias, e outros. Entretanto, esses artigos e tecnologias usualmente negligenciam a administração e preservação da informação. Por outro lado, artigos sobre catalogação digital ou convenções de metadados enfatizam as categorias de descritividade e padronização. Nesses casos, a interoperabilidade resulta principalmente da aceitação do padrão. Abordagens que integram todas as categorias, inclusive pela extensão do tema, são mais raras. Esse é o caso dos trabalhos sobre preservação digital, que tendem a abordar o estudo de metadados multidimensionalmente, amalgamando os princípios da catalogação com as novas tecnologias, em parte por necessidades funcionais.

A vinculação entre a descrição e o processamento está presente em muitos textos revisados. Melhores descrições, abrangendo diversas características dos objetos digitais, entre elas o contexto, podem permitir maior economia processual, com reutilização de componentes e aumento da flexibilidade e automatização. De certa maneira, isso exige uma reorientação das abordagens e ampliação do escopo da área de tratamento e organização da informação, que passaria a enfatizar menos as representações de objetos a serem recuperados (o esforço original da catalogação), e mais a informação organizada necessária a uma economia dos processos de trabalhos e à implementação de funcionalidades, considerando a dinâmica da evolução das novas tecnologias. Para isso, o papel dos metadados é primordial.

Outra questão que se coloca é o escopo em que se deve atuar: em um nível geral ou em um domínio. A discussão não é nova, ocorre em ontologias quando se indaga sobre a conveniência de uma macroontologia utilizável globalmente (VICKERY, 1997) ou de uma ontologia definida em um domínio e o reconhecimento de seus compromissos (JACOB, 2004). A sugestão é que, para o estudo e implementação de metadados, dado a regionalização das taxonomias que ocorre na prática, uma análise do domínio seria mais adequada, com observância das dimensões históricas e epistemológicas (HJØRLAND, 2002). Princípios gerais (não absolutos) nasceriam do confronto entre saberes de diversos domínios, o que pressupõe um redirecionamento da forma como a ciência da

informação estuda as tecnologias de metadados de interesse comum com a ciência da computação.

REFERÊNCIAS

ALMEIDA, M. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da Informação**, v. 31, n. 2, p. 5-13, 2002.

ALMEIDA, M.; BAX, M. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, v. 32, n. 3, 2003.

ARELLANO, M. A. Preservação de documentos digitais. **Ciência da Informação**, v. 33, n. 2, 2004.

BAX, M. P. Introdução às linguagens de marcas. **Ciência da Informação**, v. 30, n.1, 2001.

BAX, M. P.; LEAL, G. J. Serviços web e a evolução dos serviços em TI. **DataGramZero**, v.2, n.2, 2001.

BERNERS-LEE, T. et al. **The semantic web**. Scientific American, maio, 2001.

BRAGANHOLO, V. P.; HEUSER, C. A. XML schema, RDF(S) e UML: uma comparação. **IDEAS** 2001, p. 78-90, 2001.

CHANDRASEKARAN, B.; JOSEPHSON J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent Systems**, jan./fev., 1999.

DECKER, S. et al. The semantic web: the roles of XML and RDF. **IEEE Internet Computing**, v. 4, n. 5, p. 63-74, 2000.

DEMPSEY, L.; HEERY, R. **A review of metadata**: a survey of current resource description formats. Work Package 3 of telematics for research Project DESIRE (RE1004), 1997.

DZIEKANIAK, G. V.; KIRINUS, J. B. Web semântica. **Encontros Bibli**, Florianópolis, n.18, 2004.

FAYYAD, U.M. et al. (Eds.). **Advances in knowledge discovery and data mining**. MIT Press, 1996.

FENSEL, D. **Ontologies: a silver bullet for knowledge management and electronic commerce**. 2 ed. Berlin: Springer-Verlag, 2004.

FURGERI, S. O papel das linguagens de marcação para a Ciência da Informação. **Transinformação**, v. 18, n. 3, p. 225-239, 2006.

GILL, Tony. **Metadata and the World Wide Web**. 2000. Disponível em: <http://www.getty.edu/research/conducting_research/standards/intrometadata/> Acesso em: 04 nov. 06.

GILLILAND-SWETLAND, A. J. **Introduction to metadata: Setting the stage**. 2000. Disponível em: <<http://www.slis.kent.edu/~mzeng/metadata/Gilland.pdf>> . Acesso em: 04 nov. 06.

GREENBERG, J.; SUTTON, S.; CAMPBELL, D. Metadata: a fundamental component of the semantic web. **Bulletin of the American Society for Information Science**, v. 29, n. 4, 2003.

GRUBER, T. **What is an ontology?** Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>> . Acesso em: 04 nov. 06.

GUARINO, N.; GIARETTA, P. Ontologies and knowledge bases: towards a terminological clarification. In: MARS, N. (Ed.) **Towards very large knowledge bases: knowledge building and Knowledge sharing**. Amsterdam: IOS Press, p. 25-32, 1995.

HAN, J; KAMBER, M. **Data mining: concepts and techniques**. Morgan Kaufmann Publishers, 2001.

HEERY, R. Review of metadata formats. **Program**, v. 30, n. 4, p. 345-373, 1996.

HJØRLAND B., Domain analysis in information science. Eleven approaches—Traditional as well as innovative. **Journal of Documentation**, v. 58, n. 4, p. 422–462, 2002.

IKEMATU, R. S. Gestão de metadados: sua evolução na tecnologia da informação. **DataGramZero**, v.2, n.6, 2001.

INMON, W. H. **Como construir o data warehouse**. Rio de Janeiro: Campus, 1997.

JACOB, E. K. Ontologies and the semantic web. **Bulletin of the American Society for Information Science**, v. 29, n. 4, 2003.

JUL, E. Cataloging Internet resources: survey and prospects. **Bulletin of the American Society for Information Science**, v. 24, n. 1, p.6-9, 1997.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos, 1993.

METS. **METS: An overview & tutorial**. Network Development and MARC Standards Office. Library of Congress. 2004. Disponível em: <<http://www.loc.gov/standards/mets/METSOverview.html>>. Acesso em: 04 nov. 06.

MILSTEAD, J; FELDMAN, S. Metadata projects and standards. **ONLINE**, v. 23, n. 1, p. 32-38, 1999.

_____. Metadata: Cataloging by any other name. **ONLINE**, v. 23, n. 1, p. 24-31,1999.

MOREIRA, A. **Tesauros e ontologias**: estudo de definições presentes na literatura das áreas das ciências da computação e da informação, utilizando-se o método analítico-sintético. Dissertação, Mestrado em Ciência da Informação. Belo Horizonte: ECI-UFMG, 2003.

MOREIRA, A; ALVARENGA, L.; OLIVEIRA, A. P. O nível do conhecimento e os instrumentos de representação: tesauro e ontologias. **DataGramaZero**, v.5, n.6, 2004.

PALMER, Sean B. **The Semantic Web: An Introduction**. 2001. Disponível em: <http://infomesh.net/2001/swintro/>. Acesso em: 27 jun. 2005.

POLANYI, M. **The Tacit Dimension**. London: Routledge and Kegan Paul, 1967.

RYMAN, A. **Understanding web services**. 2003. Disponível em: <http://www-128.ibm.com/developerworks/websphere/library/techarticles/0307_ryman/ryman.htm>. Acesso em: 04 nov. 06.

SOUZA, M. I. F. et al. Metadados para descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, v. 29, n. 1, 2000.

SOUZA, R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n. 1, 2004.

SOWA, John F. **Knowledge representation: logical, philosophical, and computational foundations.** Pacific Grove: Brooks/Cole, 2000.

STRAUSS, A. L. **Qualitative analysis for social scientists.** Cambridge, 1987.

SVEIBY, K. E. **A nova riqueza das organizações.** RJ: Campus, 1998.

THOMAZ, K. P.; SANTOS, V. M. Metadados para o gerenciamento eletrônico de documentos de caráter arquivístico - GED/A: estudo comparativo de modelos e formulação de uma proposta preliminar. **DataGramZero**, v. 4, n. 4, 2003.

THOMAZ, K. P.; SOARES, A. J. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramZero**, v. 5, n.1, 2004.

VALENTIM, M. L. P. Análise de conteúdo. In: _____. (Org.) **Métodos qualitativos de pesquisa em Ciência da Informação.** São Paulo: Polis, p. 119-134, 2005.

VELLUCCI, S. L. Metadata. **Annual Review of Information Science and Technology**, v. 33, 1998.

_____. Options for organizing electronic resources: The coexistence of metadata. **Bulletin of the American Society for Information Science**, v. 24, n. 1, p.14-17, 1997.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277-286, 1997.

WEIBEL, S. The Dublin Core: a simple content description model for electronic resources. **Bulletin of the American Society for Information Science**, v. 24, n. 1, 1997.

WOODWARD, J. Cataloging and classifying information resources on the internet. **Annual Review of Information Science and Technology**, v. 31, 1996.

WYNAR, B. S. Principles of cataloging. In: _____. **Introduction to cataloging and classification.** 7 ed. Littleton: 1985.

ABSTRACT

By means of a bibliographical review on digital metadata in the information science field, aiming to distinguish the various functionalities of metadata and employing thematic content analysis, it was noticed the recurrence of technologies like XML, RDFS, ontology, data warehouses, semantic Web, Web services, among others, and of certain categories connected to metadata functions that grounded the examined works. Based on these findings, adopting a technical and historical approach, the functional categories were related to the technologies and it was showed how the categories were expounded in 19 texts selected from the review. It was observed that studies and technologies that deal with models and standards for digital preservation adopt more comprehensive and integrated perspectives, encompassing all or nearly all categories. Based on the results, trends and issues noticed in the bibliographical review were commented and approaches and methodology for analysis of metadata and related technologies were suggested.

KEYWORDS: Digital metadata. Metadata functional categories. Metadata technologies.

Originais recebidos em 15/11/2006

Texto aprovado em 15/03/2007