

Metapadrão - Descrição e Integração de Padrões de Metadados

Alcione Benacchio

alcione@inf.ufpr.br

Departamento de Informática, Universidade Federal do Paraná (UFPR)

Caixa Postal 19.081 81.531-990 Curitiba PR Brazil

Maria Salete Marcon Gomes Vaz

salete@uepg.br

Universidade Estadual de Ponta Grossa (UEPG)
Ponta Grossa PR Brazil (também vinculada à

Universidade Federal do Paraná (UFPR)

Resumo

Iniciar o processo de gestão de dados e metadados de uma organização, localizar e compreender os dados disponíveis e o seu contexto é uma tarefa difícil de ser realizada. Metadados são elementos fundamentais para a integração, interpretação, organização e localização de informações. Por isso, organizações procuram adotar padrões de metadados para tratar cada tipo de informação que ela gerencia, resolvendo parte do problema, pois utilizando padrões distintos surgem os problemas de integração dos dados entre os padrões. O objetivo deste artigo é apresentar um padrão genérico para descrição de padrões de metadados que possibilita a integração dos dados comuns entre padrões distintos, proporcionando um repositório onde um único sistema possa gerenciar os metadados de vários padrões.

Abstract

Starting the managing process of metadata in a organization, locate and understanding the available data and its context is a difficult task to be performed. Metadata are basic elements for integration, interpretation, organization, and localization of information. Thus, organizations are looking for apply metadata standards to storage every single type of information that it manage. Solving part of the problem, therefore using different standards problems with data integration can appear. O goal of this paper is to present a generic standard for metadata description that provides integration of common data between different standards, providing a repository where a single system could manage metadata of several standards.

1. Introdução

A atividade de documentar dados facilita a reutilização, minimiza problemas de inconsistência e simplifica eventuais manutenções [13]. O Metapadrão é um modelo criado com o objetivo de tornar possível a integração das funcionalidades de gestão, descrição, reuso e padronização dos dados dentro de uma organização, provendo um repositório onde os metadados registrados estejam centralizados.

Um repositório de metadados que integra todos os padrões utilizados pela organização, facilita a indexação e a recuperação de dados comuns entre padrões distintos. Por isso, o objetivo deste trabalho é especificar a arquitetura de um repositório de metadados genérico que torna possível a integração das funcionalidades de gestão, descrição, reuso e padronização dos dados dentro de uma organização.

Este ambiente utilizará alguns conceitos de modelagem e de registro de metadados baseados na norma ISO/IEC 11179 [4] e especifica que os metadados registrados estejam centralizados com o intuito de integrar todos os padrões utilizados pela organização, facilitando a indexação e a recuperação de dados comuns entre padrões distintos.

Como contribuição, este trabalho propõe uma divisão entre a arquitetura de armazenamento dos metadados e a modelagem dos padrões, ou seja, em um local serão armazenados os dados dos metadados, e em outro serão armazenados os padrões que serão utilizados como *templates* (modelos) para o armazenamento dos dados, conforme a especificação de cada padrão descrito. Outra contribuição da modelagem dos padrões é a flexibilidade para constantes extensões dos padrões utilizados pela organização.

Não faz parte do escopo deste trabalho determinar forma ou tecnologia utilizada para o armazenamento para os padrões e os dados dos metadados.

Na Seção 2 é apresentada a definição de metadados, abordando suas aplicações e uma breve explanação sobre o seu processo de gestão. Na Seção 3 são introduzidos os conceitos sobre padrões de metadados e alguns dos padrões mais conhecidos.

A Seção 4 é dedicada a apresentar a arquitetura do Metapadrão, descrevendo de forma abstrata os atributos que o compõe e como deve ser o processo de gestão de dados ao utilizar o modelo Metapadrão. E finalmente, a última seção apresenta as conclusões e perspectivas de trabalhos futuros.

2. Metadados

Metadados são definidos como “dados que descrevem dados”. Podem ser utilizados para descrever objetos ou tornar pública sua existência. Eles disponibilizam, descrevem, localizam e auxiliam na compreensão dos dados, transformando-os em conhecimento [13]. Ao ter conhecimento de quais dados estão disponíveis, entender o seu contexto e onde esses estão localizados, informações precisas são obtidas e melhores decisões podem ser tomadas [12].

Os metadados podem ser Esquemas de Classificação, Domínios Conceituais, Contextos, Elementos de Dados, Conceitos de Elemento de dados, Classes de Objeto, Propriedades, Representações de classes e Valores de Domínio.

Ao serem identificados e registrados os metadados são gerenciados como elementos dentro do repositório, são nomeados e devem possuir pelo menos um Contexto. O Contexto possui um nome e definições podem ser especificadas em uma ou mais linguagens e define o escopo e o significado de cada tipo de metadados, podendo conter informações sobre o domínio de negócio, informações sobre áreas, informações sobre sistemas, banco de dados, modelagem ou sobre qualquer ambiente determinado pelo proprietário do registro.

Um Esquema de Metadados é um conjunto de atributos definidos para atender uma determinada finalidade. Através da identificação de problemas no armazenamento e recuperação de informações por falta de padronização, vários esquemas foram criados para atender diferentes propósitos, são chamados de padrões de metadados [9].

Os metadados possuem uma vasta aplicação e são utilizados para descrever uma infinidade de informações como imagens, vídeos, atributos, tabelas, bancos de dados, mapas espaciais, documentos, entre outros. As ferramentas que manipulam *Data Warehouse* [14], por exemplo, são baseadas em metadados, que descrevem as informações nelas contidas e por elas manipuladas, desde dados técnicos até regras de negócio.

Por isso estão sendo desenvolvidos sistemas especializados para a leitura e descrição desses metadados específicos, como sistemas de armazenamento e indexação de mapas digitais, gestão eletrônica de documentos arquivísticos [2].

3. Padrões de Metadados

A utilização de padrões já foi vista como forma de limitação entre a comunidade de desenvolvedores, hoje com o grande crescimento dos dados armazenados, é visto como um grande aliado.

Investir em padrões é um esforço correto para a solução de problemas conhecidos e trazem diversos benefícios a seus usuários. Facilita a atividade de análise, pois geralmente são amplamente documentados, trazendo a tona soluções para problemas que em alguns casos ainda não tinham sido previstos pelos analistas, facilitando a comunicação entre os usuários e proporcionando uniformidade e integração entre soluções.

Os padrões estão cada vez mais presentes dentro das organizações, por isso diversos padrões foram desenvolvidos para atender a diferentes áreas como Bibliotecas Digitais, Multimídia, Documentos Arquivísticos entre outros.

Dentre os vários padrões existentes pode-se destacar o Padrão *Dublin Core* [7]. Este padrão define um grupo de atributos que pode ser utilizado por autores para descrever seus próprios recursos na web, onde de acordo com DCMI o padrão Dublin Core se destaca pela Simplicidade, Interoperabilidade Semântica, Consenso Internacional, Extensibilidade e Modularidade de Metadados na Web. Possui 15 atributos, (Title, Creator, Subject, Description,

Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights), por serem atributos extremamente genéricos, este padrão pode ser utilizado nas mais diversas áreas de conhecimento [7].

O esquema METS é um padrão para a codificação de metadados descritivos, administrativos e estruturais, utilizados para a gestão e a troca de objetos de repositórios de bibliotecas digitais. Essas bibliotecas de objetos digitais requerem a manutenção de vários tipos de metadados estruturados. Para manter metadados descritivos sobre um livro e garantir que os dados possam ser utilizados sem dúvidas sobre sua precisão, é importante a utilização de metadados técnicos e metadados estruturais.

Codificado em XML, este padrão possui sete seções, em cada seção um grupo de atributos, sendo Cabeçalho METS, Metadados Descritivos, Metadados Administrativos, Seção de Arquivos, Mapa Estrutural, Ligações Estruturais, Comportamento. O padrão METS é um schema mantido pelo Network Development and MARC Standards Office da Biblioteca do Congresso dos Estados Unidos. [10]

O MPEG (*Movie Picture Experts Group*) é o grupo da ISO/IEC encarregado do desenvolvimento de padrões para a representação de codificação de dados digitais de áudio e vídeo. Este grupo produziu vários padrões como o MPEG-1 que resultaram em produtos como o Video CD e o MP3, MPEG-2 utilizado na Televisão Digital e o no qual o formato de vídeo do DVD é baseado, MPEG-4 que está sendo aplicado para vídeos de alta compressão como os utilizados por dispositivos móveis e vídeos pela web, MPEG-7 utilizado para descrever dados de áudio e vídeo.

O padrão MPEG-7 que é interface de descrição de conteúdo multimídia foi desenvolvido no intuito de prover um *template* para ser utilizado em repositórios de dados para que possam ser utilizados por pessoas ou para que possa ser recuperado de forma automatizada por aplicações [11].

4. Metapadrão

Com a existência e o crescente desenvolvimento de novos padrões, vários sistemas de registro de metadados foram e serão criados para atender os requisitos de cada padrão, fazendo os usuários de padrões utilizarem vários sistemas diferentes. Existe a necessidade de integração entre esses metadados, pois muitas vezes esses dados registrados são relacionados. Essa integração minimiza o esforço para sincronização e consistência dos dados [7].

A partir de um repositório baseado na arquitetura definida pelo Metapadrão, os metadados gerenciados podem ser localizados, atualizados e utilizados os dados que os metadados descrevem. A arquitetura proposta registra o significado dos dados e como esses dados são representados e descritos. A partir dessas descrições pode ser compreendido com precisão, como esses dados estão contextualizados e principalmente garantir que um tipo de metadados que já foi criado não seja duplicado.

Dentro do repositório são registrados os metadados e definidos os esquemas de metadados (padrões), o modelo prove um ambiente onde os padrões registrados sejam utilizados como modelos para o registro de valores dos metadados. Contribuindo com o trabalho de gestão de metadados e auxiliando o Administrador a decidir quais metadados devem ser coletados e mantidos e onde novos metadados devem ser criados, de forma que estarão sempre integrados e podendo ser reutilizados.

Os metadados registrados devem possuir as seguintes especificações:

- Identificador único para cada elemento
- Contexto dos metadados
- Definição de metadados obrigatórios, opcionais e obrigatórios baseados em condições
- Um metadado pode ocorrer muitas vezes
- Metadados relacionados hierarquicamente
- Tipos de valores de metadados

A padronização na descrição de dados, o entendimento claro dos dados por meio dos elementos organizacionais, o reuso dos dados através do tempo, espaço e aplicações, a padronização dos dados dentro de uma organização e a gestão e reuso dos componentes de dados são os principais benefícios desta arquitetura.

4.1 Aplicação do Metapadrão

Um modelo armazena todos os element sets(Atributos) que podem ser utilizados na descrição dos schemas de metadados(padrões), um element set pode ser a junção de vários element sets, cada element set também possui uma associação de contexto e se o dado relativo ao element set é obrigatório, opcional ou condicional.

Pode ser exemplificado da seguinte forma, o padrão Dublin Core e o MPEG-7 serão registrados no repositório da organização. O primeiro passo é registrar os atributos necessários para cada padrão. Pode ser observado que tanto o padrão Dublin Core quanto o MPEG-7 possuem o atributo Creator, esse atributo é relacionado ao contexto Autoria.

Neste caso o Administrador irá criar apenas um atributo Creator que será compartilhado entre os dois padrões, possibilitando o reuso de informações e também de localizar e recuperar informações relativas ao mesmo contexto entre padrões distintos (Crosswalking). O Administrador também poderá atribuir uma legenda ao atributo Creator para cada padrão onde ele é utilizado, tornando mais compreensível para os usuários do padrão.

Nesse caso específico seria simples encontrar todos os metadados criados por determinado autor, tanto no padrão Dublin Core quanto no MPEG-7 já que utilizam a mesma estrutura de armazenamento de autores como pode ser visto na Tabela 1.

	Dublin Core	MPEG-7
Author Context	Creator	Creator

Tabela 1. Atributo Creator utilizado pelos padrões Dublin Core e MPEG-7.

Ao termos todos os atributos registrados o Administrador identifica que o Atributo Creator, por si só não atende as suas necessidades. Após avaliar suas necessidades o Administrador chega a conclusão que o Atributo Creator ou qualquer outro atributo que esteja relacionado ao contexto autoria do metadado deve possuir separadamente os atributos name, last name, email como demonstrado na Tabela 2.

	Dublin Core	MPEG-7	
Author Context	Creator		
	name	last name	email
	name	last name	email

Tabela 2. Atributo Creator estendido com novos atributos.

O Administrador tem a possibilidade de estender todos os padrões que irá utilizar, de acordo com o seu contexto, tornando o Metapadrão mais flexível e abrangente.

4.2 Modelo Conceitual do Metapadrão

O modelo descrito na figura 1 é uma abstração do modelo proposto para armazenar padrões de metadados para que possa ser utilizado como *templates* para armazenamento de conteúdo de metadados.

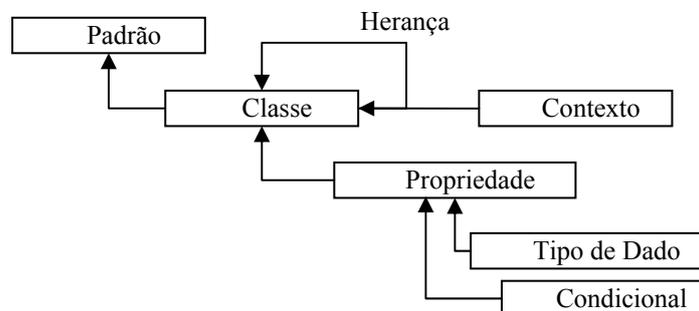


Figura 1. Modelo Metapadrão

Na entidade Padrão, serão registrados os padrões que serão armazenados no modelo, cada atributo será registrado como uma classe, possuindo propriedades que nela irão ser registrados as informações, conforme especificado nas propriedades por obrigatório, opcional ou condicional. Caso seja necessário utilizar alguma regra para preenchimento do atributo de propriedade, deverá ser registrada a condição na entidade condicional.

É obrigatória a especificação de um contexto para cada classe. Sendo que as classes podem ser utilizadas por qualquer padrão, desde que de acordo com as normas especificadas para o padrão seja necessário o uso desta classe com este mesmo contexto.

A Norma ISO/IEC 11179, que descreve um sistema de registro de metadados que pode ser utilizado por grandes organizações para o registro de seus metadados. Esta norma é dividida em seis partes (Framework, Classificação, Metamodelo e atributos básicos, Formulação e definição de dados, Nomeação e Identificação de princípios, Registro). A terceira parte da norma Metamodelo e Atributos básicos possui o mesmo objetivo deste trabalho, mas ela possui um foco mais amplo no que diz respeito aos gestores, envolvendo várias pessoas no processo. Neste trabalho foi definido como apenas um gestor o Administrador, também pode ser destacado que este modelo propõe que cada metadado seja um objeto e que cada padrão seja a composição dos objetos criados e armazenados no modelo [4] [5].

Baseadas nesta norma aplicações licenciadas foram desenvolvidas pela iniciativa privada e já estão disponíveis no mercado, como o *Data Foundations OneData Registry* e o *Oracle Enterprise Metadata Manager (EMM)*. Também é possível encontrar implementações desenvolvidas por órgãos governamentais como o Instituto Australiano de Saúde e Bem-estar que desenvolveu o *Metadata Online Registry (METeOR)* [1] e o Instituto Canadense para a Informação da Saúde que desenvolveu o *Canadian Institute for Health Information (CIHI) Data Dictionary* [3]. Ambos os casos são repositórios para padrões de dados relacionados à saúde, estatísticas de serviços para comunidade e informações.

Apesar de ter identificado várias implementações, até o mês de maio de 2007 o site da JTC1 SC32 WG2 informava que o órgão ainda não tinha validado nenhuma delas como aplicações que realmente atendessem os requerimentos descritos na Norma ISO/IEC 11179 [8].

5. Conclusões e Trabalhos Futuros

Através da utilização de metadados e padrões de metadados é possível promover a integração, interpretação, localização e reutilização dos dados, assegurando sua existência através do tempo. O modelo Metapadrão proposto provê um repositório onde as funcionalidades para o gerenciamento e padronização dos dados dentro de uma organização possam ser integrados, proporcionando ao Administrador um ambiente de gestão de metadados completo, permitindo que os metadados registrados possam ser acessados e utilizados por pessoas e aplicações.

Pode-se considerar que a principal vantagem do modelo proposto é a reusabilidade da própria aplicação que torna o ambiente muito simples de ser utilizado por seus usuários, pois o sistema poderá ser utilizado para registrar dados sobre todos os padrões e podendo haver a integração de informações entre padrões distintos.

A arquitetura proposta pela Norma ISO/IEC 11179 é mais indicada para organizações que controlam um grupo de organizações e que precisam promover a integração dos dados entre as organizações, demandando um maior número de pessoas para a gerência dos processos de registro. Por isso o Metapadrão pode ser mais apropriado para organizações que precisam padronizar seus dados sem precisar integrar seus dados com outras organizações.

6. Referências Bibliográficas

- [1] Australian Institute of Health and Welfare (2005) "Metadata Online Registry - METeOR". Última consulta em 28 de abril de 2007 <<http://meteor.aihw.gov.au/>>
- [2] Baldan, R. (2004) "EDMS: Gerenciamento Eletrônico de Documentos Técnicos". Érica.
- [3] Canadian Institute for Health Information "CIHI Data Dictionary". Última consulta em 28 de abril de 2007 <<https://eservices.cihi.ca/ddexternal/welcome.do>>
- [4] ISO/IEC 11179-1 (2004) "Information Technology – Metadata registries (MDR) – Part 1: Framework". ISO/IEC.
- [5] ISO/IEC 11179-3 (2003) "Information Technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes". ISO/IEC.
- [6] ISO/IEC 11179-6 (2005) "Information Technology – Metadata registries (MDR) – Part 6: Registration". ISO/IEC.

- [7] Duval, E., Hodgins, W., Sutton, S., Weibel S. (2002) “Metadata Principles and Practicalities”. D-Lib Magazine 8(4).
- [8] JTC1 SC32 (2006) “MDR Implementations”. Metadata Standards ISO/IEC JTC1 SC32 WG2. Última consulta em 28 de abril de 2007 <<http://www.metadata-standards.org/>>
- [9] NISO Press, National Information Standards Organization (2004) “Understanding Metadata”. NISO Press.
- [10] Metadata Encoding and Transmission Standard (METS) – Última Consulta em 20 de Junho de 2007 <<http://www.loc.gov/standards/mets/>>
- [11] Moving Picture Experts Group (MPEG) ISO/IEC JTC1 SC29 WG11 – Última Consulta em 20 de Junho de 2007 <<http://www.chiariglione.org/mpeg/>>
- [12] Tannenbaum, A. (2002) “Metadata Solutions: Using Metamodels, Repositories, XML and Enterprise Portals to Generate Information on Demand”. Addison Wesley.
- [13] Vaz, M. S. M. G. (2000) “MetaMídia – Um Modelo de Metadados na Indexação e Recuperação de Objeto Multimídia”, Tese de Doutorado, UFPE.
- [14] Vetterli, T., Vaduva, A., Staudt, M. (2000) “Metadata Standards for Data Warehousing: Open Information Model vs. Common Warehouse Metamodel”. ACM SIGMOD Record. ACM Press.



ALCIONE BENACCHIO

Graduado em Processamento de Dados pelo Centro de Ensino Superior de Foz do Iguaçu (2001), especialização em Redes de Computadores e Sistemas Distribuídos pela Universidade Federal de Santa Catarina (2005) e Mestrando em Informática pela Universidade Federal do Paraná. Atualmente é Professor Pesquisador da Faculdades Integradas do Brasil. Tem experiência na área de Ciência da Computação, com ênfase em Sistemas de Computação e Bancos de Dados e Metadados.

MARIA SALETE MARCON GOMES VAZ

Bacharelado em Processamento de Dados pela Universidade Estadual de Ponta Grossa (1988), Mestrado (1994) e Doutorado (2000) em Ciências da Computação pelo Centro de Informática, Universidade Federal de Pernambuco. Atualmente, Professora Adjunta da Universidade Estadual de Ponta Grossa. Diretora Adjunta do Setor de Ciências Agrárias e de Tecnologia/UEPG. Representante Docente no Conselho de Ensino, Pesquisa e Extensão - CEPE/UEPG. Presidente da Câmara de Pesquisa e Pós-Graduação do CEPE/UEPG. Editora Associada da Revista Publicatio - Ciências Exatas e da Terra, Ciências Agrárias e Engenharias. Professora Colaboradora e orientadora no Mestrado em Informática da Universidade Federal do Paraná. Avaliadora de Curso e Institucional do Ministério da Educação. Experiência na área de Ciência da Computação, com ênfase em Banco de Dados, atuando principalmente nos seguintes temas: banco de dados não convencionais, metadados, internet e sistema de informação.