

BIBLIOTECAS DIGITAIS E METADADOS: UMA ABORDAGEM INTEGRADORA

GALINDO, Marcos¹, PEREIRA, Marcos S.², LIMA, Cleiton M. V.³

O artigo discorre sobre a importância do uso de metadados em Bibliotecas Digitais e Sistemas de Arquivos e propõe uma solução de desenvolvimento que possibilite disseminar dados descritos em múltiplos padrões de metadados. O estudo baseia-se na representatividade de cada um dos padrões internacionalmente aceitos, e como eles suprem, ou não, a completude informacional necessária a cada tipologia documental. O estudo se dividiu nas seguintes fases: a) escolha do tema considerando-se sua relevância e importância para bibliotecas digitais; b) seleção de alunos interessados e capacitados para estudo do tema proposto, visando fazer uso da interdisciplinaridade informação-computação; c) levantamento de materiais com abrangência nas áreas de Ciências da Informação e Computação; d) criação de uma proposta de solução para o problema pesquisado, estudo da viabilidade das possíveis soluções, escolha e melhoria da mais adequada e, por fim, e) a escrita do artigo.

O Liber é um laboratório de pesquisa em Ciência da Informação e Tecnologia que desenvolveu o sistema de gerenciamento de informação acadêmica. Tal sistema foi desenhado para gerir o acervo de teses e dissertações produzida nos programas de pós-graduação da Universidade Federal de Pernambuco. Ele baseia-se no princípio universal do livre acesso, na Legislação do Depósito Legal e da Propriedade Intelectual vigente no Brasil e respeitada universalmente por força de protocolos de reciprocidade internacionais. O Liber busca instrumentalizar a produção intelectual e viabilizar a transferência de informação da academia para a sociedade utilizando-se de ferramentas hipermídicas e as redes de alcance mundial.

Alem das pesquisas sobre disseminação de dados, o Liber realiza, ainda, estudos na area de definição e uso de padrões de metadados, disseminação de dados via web, produção de softwares voltados ao gerenciamento e integração de arquivos e bibliotecas digitais e coleta de dados usando o protocolo Open Archives. A proposta maior do laboratorio é, através da interdisciplinaridade entre alunos de Ciências da Informação, Ciências da Computação e Design, prover um ambiente no qual seja possível a realização de estudos voltados a metadados e acervos digitais.

1 Liber - Universidade Federal de Pernambuco www.liber.ufpe.br galindo@ufpe.br
2 Liber - Universidade Federal de Pernambuco www.liber.ufpe.br mosp@cin.ufpe.br
3 Liber - Universidade Federal de Pernambuco www.liber.ufpe.br cmvl@cin.ufpe.br

Hoje, disseminação de informação abrange uma vasta área de instituições e aplicações que vão desde Sistemas de Arquivos Digitais, Catálogos on-line até Bibliotecas Digitais. A Internet como meio de publicação de dados revolucionou o desenvolvimento dos Sistemas de Informação. Vistos isoladamente, os dados não têm significado real, pois o verdadeiro interesse está na informação. Aqui, claro, convém conceituar dados, informações, bem como os próprios sistemas de informação. Zwass define dados como “fatos ‘crus’ que podem ser processados para se obter informação”. Informação, por sua vez, são dados inseridos num contexto, o que lhes garante significado real. Ainda segundo Zwass, “um sistema de informação é um conjunto organizado de componentes para coletar, transmitir, armazenar e processar dados em ordem para prover informação”.

Todavia, a que se considerar os problemas apresentados pela Internet, tais como grande número de repositórios, alto grau de autonomia dos mesmos, a variedade de representações para os dados e a grande irregularidade nas estruturas dos dados. Dessa maneira, repositórios interessados em compartilhar seus dados carecem, ou de uma representação universal, ou de um mecanismo que possibilite o tratamento dos dados nos diversos formatos de metadados existentes. Em tempo, metadados são, pela definição de Tronchin, “a descrição do dado, do ambiente onde ele reside, como ele é manipulado e para onde ele é distribuído”. Ou seja, metadado é uma abstração do dado. Criar uma representação universal é questionável quanto à sua completude informacional, visto que cada tipologia documental tem descrições específicas que obedecem às suas próprias características documentais. Além disso, usuários têm sempre objetivos muito próprios para realizar uma pesquisa e uma grande quantidade de informação poderia não ser relevante às especificidades do usuário. Padrões nem sempre são suficientemente representativos para um determinado domínio de aplicação, de forma que há um grande número de possíveis padrões a se escolher. Entretanto, cada um deles tem limitações quanto ao poder de representação, diferenças na linguagem de representação do próprio metadado e há, ainda, a possibilidade de alguns padrões serem complementares.

As assertivas acima levantadas demonstram a importância de se propor e construir uma solução que possibilite a comunicação entre os vários repositórios sem que haja a necessidade de que tais repositórios sigam o mesmo padrão de representação de metadados. A solução proposta é a criação de um mecanismo (software) que possibilite a manipulação de vários padrões de maneira que: a) não seja necessário adequar as estruturas já existentes em cada repositório a um dado padrão, b) permitir a melhor troca de dados entre repositórios, c) bem como prover maior qualidade de resposta às pesquisas.

Além de redimensionar a linguagem utilizada e revolucionar o acesso aos sistemas de informação, a Internet gerou a necessidade de modelos de dados com maior complexidade por causa da quantidade de informações. Grandes quantidades de dados, quando espalhados tornam-se um problema e exigem esforço para integrá-los, disseminá-los e gerar informações sobre dados. O paradigma atual de representações descritivas em metadados foi criado fora do ambiente de Biblioteconomia por tecnólogos que buscavam descritores mais eficientes e dotados de uma pluralidade tecnológica que permite o manuseio e gerenciamento de informação por máquinas com um mínimo de mediação de

bibliotecários, desta forma, diferentes padrões de metadados foram criados para resolver o problema de representação, dentre aceitos internacionalmente estão o Marc (MACHINE-Readable Cataloguing), mantido pela Library of Congress e a National Library of Canada, o Dublin Core, mantido pela Dublin Core Metadata Initiative e também o uso de XML em protocolos para sistemas de arquivos abertos, como o Open Archives.

Cada um desses padrões tem formas diferentes de especificação dos metadados, e podem utilizar diferentes linguagens de representação. Como cada um dos padrões possui comunidades e instituições que os utilizam para distribuição, uma abordagem ideal do problema, que é o princípio básico de Integração de Dados, seria dispor de mecanismos que permitam abstrair o tipo de representação documental, integrando-os numa mesma aplicação. A aplicação deve conter dois módulos específicos: a) Módulo específico que codifica os dados de acordo com o padrão: é o módulo responsável por formatar os dados de acordo com o formato requisitado. Para cada um dos padrões usados deve haver um módulo como esse que entenda exatamente como deve ser dada a resposta. Tal componente poderia ainda ser usado ou adaptado para diferentes repositórios que apresentassem semelhanças nas suas estruturas de dados; b) Módulo de recuperação dos dados no repositório: é responsável apenas por recuperar os dados no repositório específico. Esse módulo proverá serviço a todos os módulos responsáveis pela codificação requisitada. É ele quem fornece os dados que serão posteriormente formatados num padrão específico. A figura a seguir representa a disposição dos módulos:

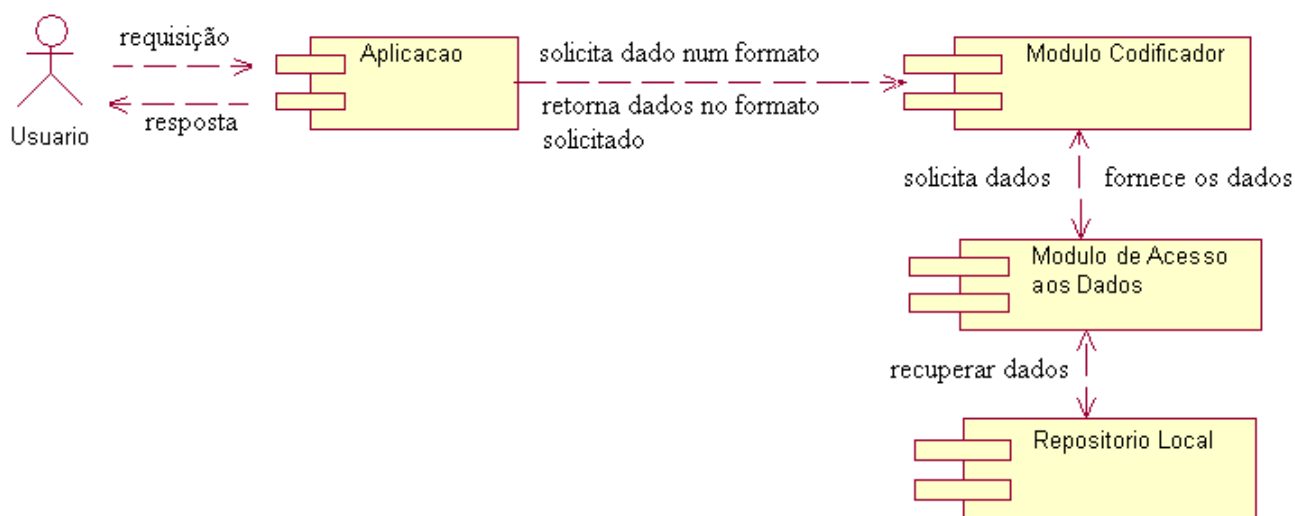


Figura 1: Arquitetura Proposta

Essa é uma proposta de solução simples e eficiente para o problema do tratamento de múltiplos padrões de metadados, a qual objetiva primordialmente disseminar informação através do suporte a múltiplos formatos buscando maior completude informacional dos acervos, oferecer maior qualidade de resposta, prover uma interface uniforme para os usuários do repositório abstraindo os detalhes da representação do mesmo. Há ainda a possibilidade de uso integrado de soluções existentes já prontas, ou parcialmente prontas. O foco maior de tal solução são os sistemas de arquivos abertos. Vale lembrar que há

iniciativas da Biblioteca Digital Brasileira mantida e desenvolvida pelo IBICT, Instituto Brasileiro de Informação em Ciência e Tecnologia.

A partir do estudo de propostas de implementação para distribuição de dados, tais como iniciativas relacionadas ao Open Archives usada pelo IBICT, padrões internacionalmente aceitos para definição de metadados, seu poder de representatividade para as mais variadas tipologias documentais, facilidade de extensão e limitações para representação de um dado acervo. Além do um estudo de viabilidade de outras possíveis soluções como um Sistema Agregador, o qual toma para si uma copia dos dados e define um esquema de representação global, um Sistema de Mediação para os repositórios existentes, propôs-se então uma solução, a qual foi explicada acima, que fosse o mais simples possível de implementar, ou seja, facilitando o uso incentiva-se a adesão dos repositórios interessados em disseminar seus dados. A simplicidade é fundamental para que não haja um esforço demasiadamente dispendioso com a construção do módulo que proverá os dados segundo vários padrões de metadados. A proposta se preocupa ainda com o impacto sobre as estruturas já existentes nos repositórios. Visto que, o sistema trabalhará quase que à parte do repositório, usando este apenas para a coleta de dados, não seria necessário alterar o repositório. A ideia é que o módulo se adeque ao repositório e não o contrário. Como o o sistema para prover informações em padrões diferentes de metadados funciona à parte do repositório, fica possível ir adicionando módulos sob demanda para cada um desses padrões. Assim, o repositório possibilitará a disseminação de informação segundo vários padrões de metadados, e não apenas um em especial. Tal solução oferece ainda boas possibilidades de reuso, construção conjunta por repositórios semelhantes e conseqüente diminuição no esforço de integração e manipulação dos dados. Assim, fica caracterizada a viabilidade de uso dessa abordagem no âmbito de Bibliotecas Digitais e Sistemas de Arquivos.

Referências

DUBLIN Core Metadata Initiative. Disponível em: <<http://dublincore.org>>.

MARC Standards. Disponível em: <<http://www.loc.gov/marc>>.

OPEN Archives Initiative. Disponível em: <<http://www.openarchives.org>>.

ZWASS, Vladimir: Introduction to Information Systems & Their Capabilities. In: ZWASS, Vladimir: Foundations of information system. Boston: Irwin McGraw-Hill, c1998. p.5.

TRONCHIN, Valsoir, Análise, Modelagem e Implementação de Data Warehouses – São Paulo: Fenasoft/98 em 20/07/98.