

USO DO *DUBLIN CORE* NA DESCRIÇÃO DE OBRAS RARAS NA *WEB*: A COLEÇÃO DA BIBLIOTECA BRASILIANA DIGITAL

Daniela Pires

Biblioteca Brasileira Guita e José Mindlin. Universidade de São Paulo, São Paulo, SP. Bibliotecária.
danipires@usp.br; tel. (11) 3091-1154

RESUMO

Acessível ao público desde junho de 2009, a Biblioteca Brasileira Digital, da Universidade de São Paulo tem por objetivo facultar para a pesquisa, a maior Brasileira custodiada por uma universidade. Sua intenção é disponibilizar virtualmente parte do acervo da Universidade oferecendo-se como um instrumento útil e funcional para a pesquisa e o estudo dos temas e cultura brasileiros, além de oferecer um modelo tecnológico de gestão que possa ser difundido a outras coleções, acervos e instituições. Este trabalho apresenta os resultados da implantação de um esquema de metadados baseado no formato *Dublin Core*, para a descrição de obras raras e especiais na *web*. Especificamente, apresenta os procedimentos e processos de descrição de conteúdos das diversas tipologias documentais (livros, periódicos, gravuras etc.) e formatos digitais (pdf, jpeg entre outros).

Palavras-Chave: Bibliotecas digitais; Metadados; Dublin Core.

ABSTRACT

Available for the public access since June 2009, the University of Sao Paulo Brasileira Digital Library, aims to provide to the research the largest Brasileira guarded by an University. The purpose is make available online part of the University Collection, as an useful and functional instrument to the research and study of Brazilian themes and culture, and also offer a technological management model able to be disseminated to other collections and institutions. This work shows the results of the implantation of a metadata schema based on Dublin Core format, for the description of rare and special works on the web. Specifically, it shows the content description procedures and processes of several documents typologies (books, journals, images etc) and digital formats (pdf, jpeg and others).

Keywords: Digital Libraries, Metadata, Dublin Core.

1 Introdução

Acessível ao público desde junho de 2009, a Biblioteca Brasileira Digital, da Universidade de São Paulo tem por objetivo facultar para a pesquisa, a maior Brasileira custodiada por uma universidade, por meio da consolidação de um modelo de biblioteca digital, fundamentada em padrões e normas internacionais. Sua intenção é disponibilizar virtualmente parte do acervo da Universidade oferecendo-se como um instrumento útil e funcional para a pesquisa e o estudo dos temas e cultura brasileiros, além de oferecer um modelo tecnológico de gestão que possa ser difundido a outras coleções, acervos e instituições. Numa primeira etapa do projeto, foram digitalizados parte do acervo da Biblioteca Brasileira Guita e José Mindlin, doada à USP pelo bibliófilo José Mindlin, que colecionou ao longo de sua vida um expressivo conjunto de livros e manuscritos, sendo sua coleção considerada a mais importante coleção do gênero formada por um particular. São cerca de 17.000 títulos, ou 40.000 volumes, compreendidos entre o Século XVI até o atual, e conta com obras de literatura brasileira (e portuguesa), relatos de viajantes, manuscritos históricos e literários (originais e provas tipográficas), periódicos, livros científicos e didáticos, iconografia (estampas e álbuns ilustrados) e livros de artistas (gravuras).

Com o desenvolvimento de novas tecnologias de gestão da informação, a gestão e a descrição dos conjuntos documentais têm passado por intensa transformação. Apresenta-se aqui, mais detalhadamente o modelo baseado em padrões atualizados e compartilhados de descrição e a consolidação de um esquema de metadados descritivos e gerenciais dos dados e informações produzidos pelo Projeto Brasileira Digital.

Este trabalho é resultado da implantação de um esquema de metadados baseado no formato *Dublin Core*, para a descrição de obras raras e especiais na *web*, a partir da coleção da Biblioteca Brasileira Guita e José Mindlin. Especificamente, apresentaremos os procedimentos e processos de descrição de conteúdos das diversas tipologias documentais (livros, periódicos, gravuras etc.) e formatos digitais (pdf, jpeg entre outros) além do estudo realizado para a definição de quais campos e qualificadores utilizados e como estes atendem às necessidades da coleção e de suas diferentes tipologias documentais e seus níveis especializados de descrição, além de atender as orientações do próprio repositório digital (DSpace) quanto ao uso de cada campo (e qualificadores) de informação. Apresentaremos inclusive as ferramentas de apoio utilizadas (catálogos de autoridades, vocabulários controlados, obras de referências entre outros). Além disso citaremos os esforços para a consolidação de um padrão de metadados baseado na simplicidade de descrição para acervos de várias naturezas documentais, tendo como base o esquema *Dublin Core* e a sua disponibilização na Plataforma Corisco (um sistema integrado de aplicativos e recomendações para sustentar a implantação e gerenciamento de bibliotecas (repositórios) digitais, que foi desenvolvida pelo Laboratório da Brasileira Digital). Por último trataremos da integração e a consolidação de um esquema de metadados gerenciais e administrativos produzidos pelo Projeto da Brasileira Digital.

2 Revisão de Literatura

As bibliotecas digitais são hoje um instrumento de multiplicação, universalização de acesso e de democratização de conteúdos documentais. São consideradas fundamentais para uma política de difusão cultural, pesquisa e formação. Essas novas bibliotecas são sistemas de informação que oferecem todos os tipos de conteúdo digital, incluindo gêneros documentais e publicações eletrônicas.

De acordo com a *Digital Library Federation (DLF)*:

bibliotecas digitais são organizações que fornecem recursos, incluindo pessoal especializado, para selecionar, estruturar, oferecer acesso intelectual,

interpretar, distribuir, preservar a integridade e assegurar persistência através do tempo de coleções de objetos digitais, para que sejam facilmente e economicamente disponíveis para uso de uma comunidade alvo definida ou um conjunto de comunidades. (DLF, 1998)

Com a evolução do conhecimento científico e advento de novas tecnologias de informação houve necessidade de serem criados mecanismos para facilitar a recuperação de documentos e agilizar a comunicação. Isso gerou também para as bibliotecas a necessidade de acompanharem este desenvolvimento e oferecem aos seus usuários outras formas de acesso ao conteúdo de seus acervos.

Hoje um dos grandes serviços de disseminação da informação oferecidos pelas bibliotecas aos seus usuários é a consolidação de um sistema de digitalização e disponibilização de seus acervos no formato digital, contribuindo para tornarem-se um instrumento de trabalho, investigação e difusão cultural e oferecendo aos usuários conteúdos em diversas tipologias documentais (livros, periódicos, mapas, gravuras etc) e em diversos formatos eletrônicos (pdf, mp3, jpeg entre outros).

Sendo assim podemos afirmar que as bibliotecas digitais não vêm para substituir o conceito tradicional da biblioteca como centro de transmissão da cultura e do saber, elas ampliam e renovam este conceito, como também abrem novos horizontes para as atividades das bibliotecas, uma vez que o acesso à informação não ocorrerá em determinado espaço físico, em um determinado horário, ele ocorre no momento em que o usuário sente a necessidade de obter a informação.

Desde a formação das primeiras coleções de documentos houve a preocupação de possibilitar o acesso à produção intelectual dos povos. As primeiras bibliotecas surgiram do instinto de preservar a memória coletiva dos grupos sociais e da paixão de colecionar. Isto gerou a necessidade de ordenamento para seu posterior uso. Esta organização/ordenamento permite que os documentos sejam recuperados com o intuito de oferecer informações/conhecimentos necessários ao desenvolvimento científico, ao exercício de atividades profissionais ou simplesmente pelo prazer da leitura. (LEMOS, 1998).

As atividades das bibliotecas compreendem, em termos simples, o atendimento aos usuários, a organização, o tratamento e a disseminação de informações para o público tendo como missão atender às cinco leis da biblioteconomia, de S. R. Ranganathan: - Os livros são para usar; - A cada leitor seu livro; - A cada livro seu leitor; - Poupe o tempo do leitor; - A Biblioteca é um organismo em crescimento.

A representação do conhecimento registrado, ou a catalogação de itens documentais, tem por objetivo central propiciar aos usuários que estes encontrem aquilo que procuram. Esta representação com sintaxe e semântica própria, consiste no levantamento das características dos registros e na cognição das características do usuário. Estes registros podem pertencer a vários acervos e estarem disponíveis em diversos suportes e formatos. Contudo, é esta caracterização que permitirá a individualização dos documentos, bem como reuni-los em diversas outras coleções, de acordo com suas semelhanças. (MEY; SILVEIRA, 2009)

As atividades realizadas pelas bibliotecas e centros de informação/documentação são hoje gerenciadas por bancos de dados e sistemas de informação que permitem integrar todos os processos envolvidos no tratamento de documentos, inclusive com a integração de ferramentas de apoio, como catálogos de autoridades, vocabulários controlados, tesouros, obras de referências entre outros.

Um dos principais conceitos difundido hoje entre os profissionais que atuam na descrição de recursos eletrônicos na Internet é o de metadados. De acordo com a *National Information Standards Organization* - NISO (2004, p. 1) metadados podem ser definidos como

informação estruturada que descreve, explica, localiza, ou ainda permite que um recurso informacional possa ser recuperado, usado e gerenciado. O termo metadados freqüentemente designa dados sobre dados, ou informação sobre informação.

As bibliotecas sempre produziram metadados, por meio de seus catálogos, índices entre outros instrumentos. São estes metadados, devidamente organizados e estruturados que garantem que a informação seja descoberta, localizada e possa ser utilizada. O desafio hoje está na descrição da informação que se encontra em diversos formatos digitais e que precisam atender a diferentes públicos e usos.

Os metadados podem ser descritos separadamente em um sistema de armazenamento de informações (bancos de dados), como podem estar embebidos nos arquivos digitais.

A utilização de padrões e normas para a descrição de registros informacionais é necessária para garantir a interoperabilidade entre os diversos sistemas de informações. Existem diferentes padrões e normas para a descrição de documentos, cabe a cada instituição definir o que melhor se adapta às necessidades de seu acervo e tipos documentais. Para o gerenciamento de acervos digitais um dos formatos mais conhecidos e utilizados por grandes projetos de bibliotecas digitais para a descrição de seus recursos na *web* é o formato *Dublin Core*.

O formato *Dublin Core* surgiu em 1995 no Workshop patrocinado pela *Online Computer Library Center* (OCLC) em Dublin, Ohio (Estados Unidos), com o objetivo de definir um conjunto de elementos que poderiam ser utilizados pelos autores e não-catalogadores para descrever seus próprios recursos na *Web* e facilitar a descoberta destes recursos na Internet. Desde então, este tem sido o formato utilizado por. Seu desenvolvimento é gerenciado pelo *Dublin Core Metadata Initiative* (DCMI). A partir de 2003 tornou-se o padrão ISO 15836.

O formato *Dublin Core* é composto por 15 elementos de descrição: *title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage* e *rights*. Este formato simples e conciso permite descrever vários tipos de recursos na Internet. Contudo, com a evolução do formato, estes 15 elementos de descrição se tornaram mais completos com a utilização de Qualificadores. Estes qualificadores têm a função de refinar ou tornar mais específico o recurso descrito. Portanto, é decisão do sistema sobre qual formato utilizar (simples ou qualificado). A opção do Projeto Brasileira Digital foi por utilizar o formato de descrição com qualificadores, contudo o repositório digital DSpace possui algumas orientações específicas quanto ao uso de alguns campos de metadados que são diferentes da Iniciativa *Dublin Core*, como pode ser observado no Quadro 1. No quadro 2 apresentamos os metadados utilizados pelo Brasileira Digital na plataforma Corisco.

Tabela 1 - Diferenças entre os elementos Dublin Core e seus qualificadores no repositório DSpace (as diferenças estão destacadas em negrito)

Nome do elemento DC	Qualificador	Dublin Core – DSPACE Elemento	Dublin Core – DSPACE Qualificador
Title	Alternative	Title	Alternative
Creator		Não recomenda a utilização	
Subject		Subject	Classification DDC LCC LCSH MESH Other
Description	Table of contents		Tableofcontents

	Abstract		Abstract Provenance Sponsorship Statementofresponsability URI Version
Publisher		Publisher	
Contributor		Contributor	Advisor Author Editor Illustrator Other
Date	Created Valid Available Issued Modified Copyrighted Submitted	Date	Created Accessioned Available Issued Copyrighted Submitted Updated
Type		Type	
Format	Extend Medium		Extend Medium Mimetype
Identifier	URI Bibliographic citation		URI Bibliographic citation Govdoc ISBN ISMN ISSN Other Sici Slug
Source		Não recomenda a utilização	
Language			ISO RFC3066
Relation	Is version of Has version Is replaced by Replaces Is required by Requires Is part of Has part Is referenced by References Is format of Has format Conforms to		Is version of Has version Is replaced by Replaces Requires Is part of Has part Is referenced by Is format of Isbasedon Ispartofseries URI
Coverage	Spatial Temporal	Coverage	Spatial Temporal
Rights	Access Rights License		Holder URI

Tabela 2 - Metadados utilizados pela Brasileira Digital na Plataforma Corisco

Elemento Dublin Core	Qualificador	Etiqueta / legenda para o site	Notas
DC.Contributor	author	Autor	Entidade responsável pelo conteúdo do item
DC.Contributor	other	Colaborador	Pessoas ou entidades que contribuíram para o conteúdo do item. Ex.: tradutores, ilustradores, gravadores, prefaciadores, impressor, etc
DC.Title	none	Título	Nome dado ao documento. Forma pelo qual o documento é formalmente conhecido
DC.Title	alternative	Título alternativo	Forma alternativa ao título. Neste campo incluem-se as traduções ou abreviaturas do título
DC.Date	none	Data de publicação	Data de publicação do item
DC.Description	none	Descrição/Notas	Descrição de particularidades do item
DC.Description	isversionof	Versão	Referencia a edição
DC.Description	provenance	Procedência	Descrição da custódia do item. Utilizar este campo para descrever a que biblioteca pertence o item
DC.Description	abstract	Resumo da obra	Sinopse do conteúdo do recurso
DC.Description	tableofcontents	Conteúdo	Descrição do conteúdo do item: sumário, lista de ilustrações
DC.Publisher	none	Local de Publicação	Entidade responsável por tornar o documento acessível em sua forma. Incluem pessoas, organizações ou serviços. (Colocar Local de publicação e o nome da editora)
DC.Format	medium	Descrição física	Descrição dos aspectos físicos do livro (n. de páginas, volume, ilustrações etc)
DC.Language	none	Idioma	Utilizar este campo para introduzir informação sobre a língua diferente do formato ISO
DC.Relation	none	Relação	Relação do item. (Nestes campos de metadados devem ser atribuídos as relações de uma obra, deve-se utilizar para se estabelecer os <i>links</i> no acervo ou fora dele)
DC.Relation	ispartofseries	Série	Nome e número da série/coleção
DC.Relation	ispartof	Parte de	Utilizar quando o item for parte de uma obra (Ex. capítulos de livros, para imagens, mapas que fazem parte de um item que já faz parte do acervo, inserir o handle)
DC.Relation	haspart	Possui parte	Recurso possui parte física ou lógica de outro item
DC.Relation	hasversion	Tem versão	Recurso está disponível em diferentes versões
DC.Relation	isbasedon	Baseado em	Recurso está baseado em outro
DC.Relation	isreferencedby	Referenciado por	Recurso é referenciado por outro recurso
DC.Relation	isformatof	Formato adicional	Recurso está disponível em mais de um formato

DC.Relation	requires	Vinculado com	Recurso requer algum suporte para seu funcionamento, distribuição ou coerência
DC.Relation	replaces	Continua em	Recurso continua em outro título ou forma
DC.Relation	isreplacedby	É continuação de	Recurso é continuação de outro
DC.Relation	uri	Recurso relacionado	Inserir neste campo o link para o catálogo eletrônico (quando item pertencer à biblioteca parceira)
DC.Rights	none	Direitos	Especificar sobre os direitos autorais (domínio público; proibido uso para fins comerciais etc)
DC.Subject	none	Assunto	Assunto. Consultar vocabulário controlado do SiBi-USP. (descritores em português)
DC.Type	none	Tipo	Tipologia documental

3 Materiais e Métodos

A digitalização do acervo da Biblioteca Brasileira Guita e José Mindlin é realizada a partir de um sistema integrado de digitalização robotizada de livros encadernados (APT 2400 RA *BookScan da Kirtas Technologies*, que possui 2 câmeras Canon de 25 megapixel) na sua mais moderna versão, que permitiu desdobrar espacialmente as atividades de digitalização, processamento das imagens (através do software integrado ao scanner, *BookScan Editor*), preparação dos objetos digitais, metadados (administrativos, gerenciais e descritivos) e gestão do repositório digital.

A gestão das informações bibliográficas da Biblioteca Digital é realizada pelo repositório DSpace (www.dspace.org), que é um *software* de código aberto que permite a construção e gerenciamento de repositórios institucionais e bibliotecas digitais. O DSpace provê mecanismos para submissão, aprovação e publicação de documentos, além de um motor de busca baseado no Apache Lucene (lucene.apache.org) e a possibilidade de interligação com outros repositórios digitais via protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH - www.openarchives.org/pmh/).

Contudo, diante do grande volume de dados e informações gerados em cada um dos processos envolvidos na disponibilização *online* de documentos (digitalização, processamento das imagens, OCR, compactação de arquivos, metadados bibliográficos, interface *web* e disponibilização no site), foi necessária a organização dos metadados de acordo com as seguintes características:

- ✓ Metadados descritivos: descrição e identificação dos recursos informacionais; trata-se da descrição bibliográfica, ou a catalogação dos itens;
- ✓ Metadados administrativos e gerenciais: administração e gerenciamento das informações relacionadas aos documentos, processos e pessoas;
- ✓ Metadados de preservação: gestão das informações de preservação digital dos documentos;
- ✓ Metadados técnicos: informações técnicas dos *softwares*, *hardwares* e processos utilizados;
- ✓ Metadados de uso: informações sobre usos e usuários dos recursos de informação.

Depois de definido o esquema de metadados (*Dublin Core* qualificado) para a descrição bibliográfica, houve a necessidade de adaptá-lo às necessidades da coleção, definindo quais campos e qualificadores seriam utilizados e como estes atenderiam às necessidades da coleção e as diferentes tipologias documentais, assim como atender as orientações do repositório

(DSpace), quanto ao uso dos campos. Abaixo descrevemos sucintamente como foi o planejamento das atividades:

- **Importação dos dados bibliográficos:** foi elaborada uma tabela de importação de dados, que definiram quais dados e campos bibliográficos seriam importados do banco de dados da Biblioteca Guita e José Mindlin, que está na plataforma ISIS (*software* de armazenagem e recuperação da informação) para um arquivo XML no formato *Dublin Core*, que é o formato de importação e exportação aceito pelo repositório; a partir desta etapa os dados são importados para o repositório conforme os documentos vão sendo digitalizados. Abaixo apresentamos o conteúdo de arquivo de metadados para ser importado no Repositório Digital.

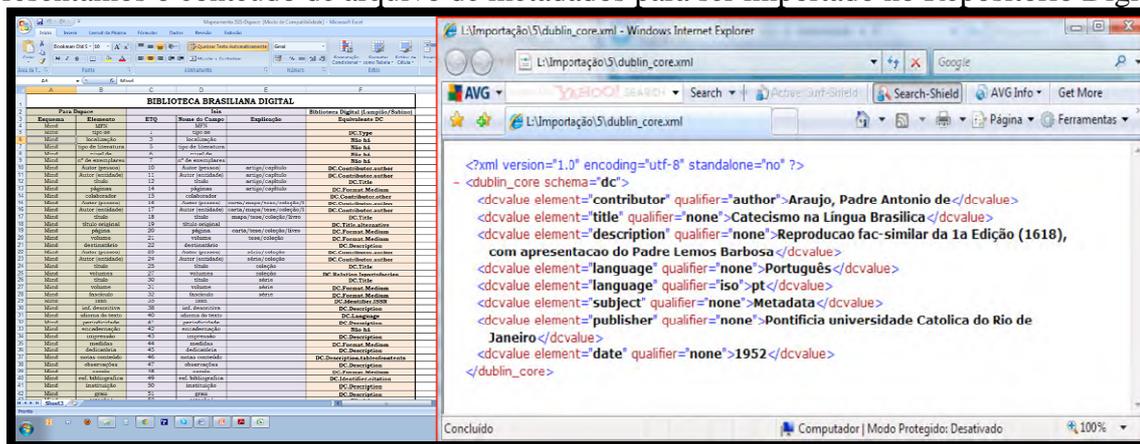


Imagem 1 – Tabela de importação do Banco de dados e seu resultado em XML

- **Definição do padrão de descrição dos metadados:** para garantir a uniformidade e a padronização da descrição dos metadados foram adotadas normas e padrões utilizados e consolidados pelas bibliotecas, o documento norteador utilizado é o AACR2. Após a importação dos metadados para o DSpace os metadados são conferidos e descritos conforme o padrão de descrição adotado pela Biblioteca Brasileira Guita e José Mindlin para recursos na *web*. (As orientações resumidas podem ser verificadas no Quadro 2).

- **Definição das tabelas auxiliares, vocabulários controlados e tesouros utilizados:** esta etapa consistiu em analisar as ferramentas auxiliares para a descrição bibliográfica, com o objetivo de uniformizar e padronizar a entrada de dados, como por exemplo, vocabulários controlados para a padronização das entradas de assuntos, tabelas de autoridades para a padronização dos nomes dos autores e colaboradores dos documentos, tesouros geográficos para a cobertura geográfica dos documentos. O formato Dublin Core também possui algumas sugestões de vocabulários controlados para auxiliarem no preenchimento dos campos. No Projeto os campos e as tabelas auxiliares utilizadas são:

Tabela 3 - Tabelas auxiliares utilizadas no Projeto Brasileira Digital

Campos DC	Tabelas auxiliares
<i>Subject</i>	- Vocabulário controlado da USP (Disponível para acesso no endereço: http://143.107.73.99/Vocab/SIBIX652.dll/Index)
<i>Contributor</i>	- Catálogo de Autoridades da Biblioteca Nacional (catálogos.bn.br) - Catálogo de Autoridades da LC (<i>Library of Congress Authorities</i> - http://authorities.loc.gov/)
<i>Type</i>	-Tabela de Tipologia Documental definida pelo Projeto Brasileira Digital
<i>Coverage (qualifier: spatial)</i>	- Tesouros de nomes geográficos da <i>Getty (TGN - Thesaurus of Geographic Names)</i>

- **Definição da estrutura de apresentação dos campos de metadados para o usuário:** depois de instanciados os metadados (descritivos, administrativos, preservação, técnicos e uso), estes foram selecionados para serem apresentados aos usuários da Biblioteca Digital em nosso *site*, contudo apenas os metadados descritivos dos itens digitalizados são apresentados aos usuários. Também foram definidos que os campos *dc.contributor* e *dc.subject* seriam formatados para funcionar como *links*, permitindo a busca a todos os itens que utilizem a mesma entrada de dados.



Imagem 2 – Apresentação dos metadados no site

- **Definição da tabela de tipologia documental:** esta etapa envolveu a definição e descrição dos tipos documentais que compõem o acervo digital e a definição das coleções no repositório que seriam criadas para abrigar o acervo. De forma simples apresentamos a tipologia documental definida que orientou a arquitetura das coleções no site: Livros (que também inclui os folhetos), Imagens, Mapas, Periódicos, Manuscritos e Obras de Referência (que possui também uma interface de busca fora do DSpace).



Imagem 3 – Arquitetura das coleções

- **Definição da rotina de cadastramento e conferência dos metadados:** esta etapa consistiu em cadastrar a equipe de catalogadores e sistematizar a entrada de dados no repositório, pois há diferenças para cada banco de dados. Depois de importadas as informações bibliográficas

são feitas alterações, correções, atualizações e a padronização destas informações. Como última etapa é realizada uma conferência no arquivo e nos dados para que estes possam ir para o servidor *web*.

- **Enriquecimento dos metadados:** uma das exigências para a descrição dos metadados é permitir uma contextualização do livro, oferecendo um pequeno resumo/sinopse do seu conteúdo. O campo *Dublin Core* utilizado para este tipo de informação é *dc.description.abstract*. Para esta etapa a Biblioteca Brasileira conta com um grupo de colaboradores especializados em determinadas coleções e temas, que são responsáveis pela elaboração de textos de apresentação que são publicadas na *homepage* e pela produção de sinopses para algumas obras. Além disso, é necessário também o uso de ferramentas de apoio como dicionários biobibliográficos, enciclopédias, dicionários e documentos especializados em brasileira. Além deste tipo de descrição, também procuramos descrever as subdivisões de um documento, sempre que estas forem importantes para o entendimento da obra, como por exemplo, o sumário, a relação de gravuras, ilustrações, estampas, mapas etc. O campo *Dublin Core* utilizado para descrever este tipo de informação é *dc.description.tableofcontents*.

- **Estabelecimento de relações entre os itens:** um dos pontos positivos para o uso do Dublin Core para a descrição de recursos na web se baseia no conceito de estabelecer relações entre itens do mesmo acervo ou mesmo com itens pertencentes a outros acervos ou catálogos. O campo utilizado para esta finalidade é o *dc.relation* com seus qualificadores. Como exemplo, podemos citar a descrição de todas as gravuras de álbuns de iconografias e estabelecer a relação entre a obra completa, as relações entre um item e suas partes é feita através dos campos *dc.relation.ispartof* (é parte de) e *dc.relation.haspart* (tem parte). Também estabelecemos as relações entre os vários volumes de uma mesma obra (uma vez que cada volume é descrito separadamente), através do campo *dc.requires* (vinculado com) e de todas as suas edições por meio do campo *dc.relation.hasversion* (tem versão/edição) e *dc.relation.isversionof* (é versão/edição).

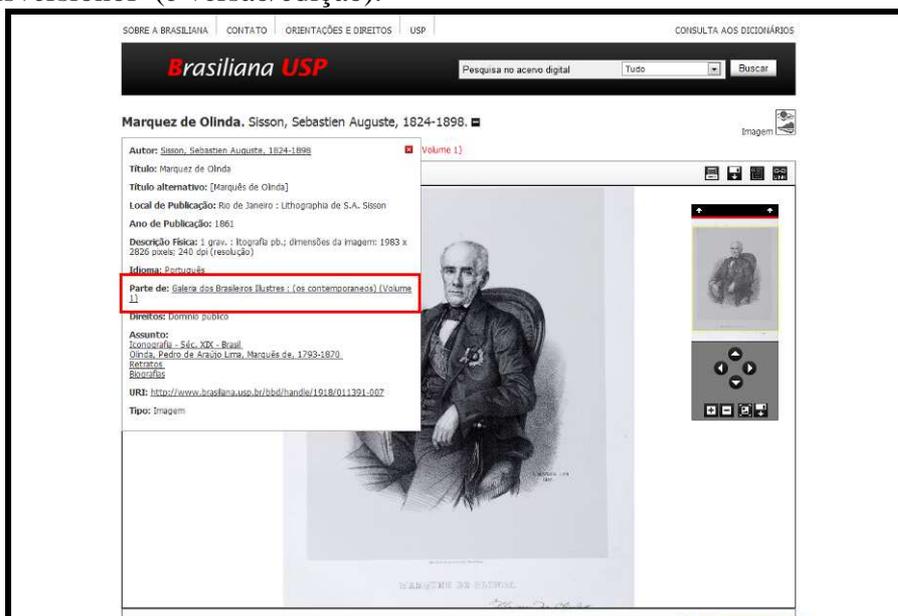


Imagem 4 – Exemplo de relação entre itens

- **Elaboração de manual de tratamento da informação para os diferentes tipos documentais:** para uniformizar a entrada de dados na Brasileira Digital está sendo produzido

um manual de tratamento da informação para os diferentes tipos documentais (livros, periódicos, imagens etc.).

Apresentamos de uma forma geral e sucinta as principais atividades realizadas pela equipe de metadados quanto à descrição bibliográfica dos documentos digitais da Brasileira Digital. Os metadados que são produzidos em cada um dos processos e etapas da disponibilização *online* de documentos passam por constantes reformulações e adaptações tendo como objetivo a adequação do repositório aos diferentes tipos e necessidades de descrição documental que são impostas à equipe de desenvolvimento da plataforma Corisco. Apresentamos a seguir alguns dos resultados finais e parciais de projetos de metadados em desenvolvimento.

4 Resultados Parciais

4.1 Trabalhos em desenvolvimento

Muitos são os desafios enfrentados para a consolidação de um modelo de gestão de acervos digitais, que contemple o gerenciamento de todos os processos, informações, pessoas e a comunicação das equipes e a consolidação da Plataforma Corisco, baseada no DSpace, como um sistema integrado de aplicativos e recomendações para sustentar a implantação e gerenciamento de bibliotecas digitais. Apresentamos agora os resultados obtidos, além de novos desenvolvimentos necessários:

- ✓ Consolidação do formato Dublin Core como o esquema de metadados da Biblioteca Brasileira Digital. Neste momento, está em andamento um estudo para a simplificação do esquema DC utilizado, cuja intenção é diminuir o uso de qualificadores para utilizar o esquema de forma mais concisa e objetiva;
- ✓ Aplicação na plataforma Corisco de vários esquemas de metadados, de forma que ela possa atender a acervos de várias naturezas e tipos documentais, como documentos de arquivos, vídeos, música etc. com o objetivo de se tornar uma ferramenta de fácil customização para diferentes necessidades de descrição;
- ✓ Consolidação de um banco de dados que reúna novos esquemas de metadados e que concentre as informações gerenciais; bibliográficas e de preservação digital, esta solução consiste em um único sistema de informações, com vários esquemas de metadados que permitam a unificação num mesmo ambiente dos dados e arquivos, utilizando o repositório DSPACE como o ambiente centralizador destas informações;
- ✓ Testes para a utilização de metadados embutidos utilizando o esquema *IPTC metadata*: padrão desenvolvido em 1970 pela *International Press Telecommunications Council* para descrever, inserir e editar dados dos arquivos de imagem digital. Este esquema de metadados tem por objetivo embutir dados nos arquivos digitais para facilitar a recuperação de arquivos por mecanismos de buscas.
- ✓ *Workflow* de produção: inicialmente, o projeto pretendia que o sistema de controle para o fluxo de processos da produção dos arquivos digitais da Brasileira Digital fosse feito conjuntamente com banco de dados gerencial. Contudo, hoje percebemos que estes devem ser bancos de dados diferentes, com controles próprios e interligados. Neste momento estamos desenvolvendo um sistema de controle para padronizar o sistema de monitoramento dos arquivos digitais, avaliar a eficiência do modelo adotado, acompanhar os processos e as pessoas envolvidas em cada etapa da produção de objetos digitais com o propósito de obter dados e resultados para relatórios além da produção de estatísticas sobre o fluxo de produção. Este sistema de controle utilizará uma ferramenta de trabalho colaborativo, uma vez que as equipes trabalham em ritmos

diferentes, e necessitam que a informação acessada esteja sempre atualizada e sincronizada.

4.2 – Projetos em desenvolvimento

A Biblioteca Brasileira Guita e José Mindlin sedia um Grupo de Pesquisa em Humanidades Digitais cujos projetos desenvolvidos no âmbito dos metadados são:

- ✓ **Iconografia em língua estrangeira:** O projeto tem como meta específica produzir descrições dos documentos iconográficos das obras em língua estrangeira do acervo da Biblioteca *Brasileana Guita e José Mindlin*, com a correta tradução dos títulos, legendas e demais informações relativas às imagens, cujo objetivo geral, a partir disso, é colaborar para a uniformização do catálogo de iconografia do acervo *Brasileana-USP* e para a ampliação do acesso a esse catálogo por parte de mecanismos de busca. Estudo para sistematizar a descrição de topônimos em catálogos para proporcionar pesquisas interativas e inteligentes. Projeto desenvolvido em parceria com a Faculdade de Filosofia, Letras e Ciências Humanas, Departamento de Letras Modernas, nas áreas de relacionadas às línguas: alemão, latim e francês.
- ✓ **Formação de bancos de dados a partir de obras de referência na Biblioteca Brasileira Digital:** O objetivo deste projeto é trabalhar com obras de referência pertencentes ao acervo da Brasileira, transformando-as em documentos digitais com funcionamento de bancos de dados. As obras, atualmente em formato de imagens digitais, serão convertidas para arquivos XML, anotados de forma a poderem servir de base para buscas cruzadas programadas em X-Query. Como produto deste projeto, ofereceremos ao usuário da Biblioteca Brasileira Digital da USP um índice remissivo de autores e personagens históricos, muitos deles presentes nos demais texto do acervo. Com isso, usuários de diversos perfis encontrarão, na Brasileira, um banco de informações importante para pesquisas – acadêmicas e escolares. O resultado, desta forma, será um fator de extroversão do conteúdo da Brasileira, e um fator de ampliação e democratização do seu acesso.

5 Considerações Finais

A tarefa de agregar valor aos documentos que estão sendo digitalizados e permitir que estes sejam recuperados é muito gratificante para todos os envolvidos na construção de acervos digitais. Pois além da tarefa de padronização e cadastramento das informações úteis e necessárias para a recuperação informação pelos mecanismos de buscas, há o processo de aprendizado que abrange tanto a área de modelagem e criação de uma Biblioteca Digital quanto à descoberta e o uso das fontes de informação relacionadas à história do país e a preservação de nossa memória documental.

No entanto, o processo de construção de um projeto desta dimensão é lento e complexo, pois estão envolvidos conceitos interdisciplinares, relacionados à gestão de projetos e pessoas, biblioteconomia e sistemas de bibliotecas, tecnologias da informação, *design*, história, literatura, linguística e um conjunto de profissionais e contextos que permitiram que a Biblioteca Brasileira Digital entrasse no ar, em pleno funcionamento em junho de 2009 e pudesse ser construída desde então com as sugestões dos usuários e diversos projetos de pesquisas.

Muitos processos são constantemente revistos e passam por um processo de adequação e atualização, inclusive neste momento uma das etapas inclui a implementação de um

protocolo de coleta de metadados em repositórios digitais com o objetivo de proporcionar a interoperabilidade, dar visibilidade e integrar metadados e informações de diferentes repositórios digitais, utilizando, por exemplo, o protocolo OAI-PMH da *Open Archives Initiative Protocol for Metadata Harvesting*.

Quanto à manutenção de bancos de dados em bibliotecas é importante salientar que os repositórios digitais não substituem nossos catálogos gerenciais do acervo físico, mantidos durante todos esses anos em formato MARC 21 ou em qualquer outro formato de descrição. Para o ambiente da web o esquema de metadados *Dublin Core* é hoje uma alternativa segura, interoperável e consolidada para a descrição de documentos, inclusive para a catalogação de obras raras, cujo nível de descrição deve ser mais específico e detalhista. Desta forma, podemos afirmar que o *Dublin Core* proporciona uma descrição mínima e objetiva dos itens do acervo, como também permite que os itens sejam encontrados no repositório digital. Para a Biblioteca Brasileira Digital este esquema se mostra muito eficiente e atende as expectativas de descrição, permitindo inclusive estabelecer conexões entre vários itens de um mesmo acervo ou de acervos diferentes.

Diante disso, o que nos motiva e constitui a nossa missão é poder “inocular o vírus da leitura” como dizia o Sr. José Mindlin, por meio da disponibilização destes documentos, contribuindo para a construção de uma sociedade de leitores. Contudo, devemos sempre nos lembrar que para que o leitor chegue até o livro é necessário que ele primeiro o encontre e esta é nossa principal tarefa enquanto catalogadores.

6 Referências

DIGITAL LIBRARY FEDERATION. *A Working Definition of Digital Library [1998]*. Disponível em <<http://www.diglib.org/about/dldefinition.htm>>. Acesso em: 04 maio 2007.

DUBLIN CORE METADATA INITIATIVE. **Dublin Core element set, version 1.0**. Disponível em: <<http://dublincore.org/documents/dces/>>. Acesso em: 25 maio 2008.

DUBLIN CORE METADATA INITIATIVE. **Using Dublin Core**. Disponível em: <<http://dublincore.org/documents/2005/11/07/usageguide/>>. Acesso em: 15 ago. 2008.

DUBLIN CORE METADATA INITIATIVE. **Using Dublin Core: Dublin Core Qualifiers**. Disponível em: <<http://dublincore.org/documents/2005/11/07/usageguide/qualifiers.shtml>>. Acesso em: 15 ago. 2008.

LEMOS, Antonio Briquet de. Bibliotecas. *In*: CAMPELLO, Bernadete (Org). **Formas e expressões do conhecimento**: introdução às fontes de informação. Belo Horizonte: Escola de Biblioteconomia da UFMG, 1998. p. 347-366.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofolletti. **Catalogação no plural**. Brasília: Briquet de Lemos/Livros, 2009.

NISO (National Information Standards Organization). (2004). *Understanding metadata*. Bethesda, MD: NISO Press. Available: <<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>>. Acesso em: 15 set. 2009.