

Isa M. Freire, Lilian M. A. R. Alvares,
Renata M. A. Baracho, Mauricio B. Almeida,
Beatriz V. Cendon, Benildes C. M. S. Maculan
(Org.)



ALÉM DAS NUvens:
EXPANDINDO AS FRONTEIRAS DA CIÊNCIA DA
INFORMAÇÃO

ISSN 2177-3688

BELO HORIZONTE

ECI/UFMG

2014

DIREITO AUTORAL E DE REPRODUÇÃO

Direitos de autor ©2014 para os artigos individuais dos autores. São permitidas cópias para fins privados e acadêmicos, desde que citada a fonte e autoria. E republicação desse material requer permissão dos detentores dos direitos autorais. Os editores deste volume são responsáveis pela publicação e detentores dos direitos autorais.

E56a
2014 Encontro Nacional de Pesquisa em Ciência da Informação : além das nuvens, expandindo as fronteiras da Ciência da Informação (15. : 2014 : Belo Horizonte, MG).

Anais [recurso eletrônico] / XV Encontro Nacional de Pesquisa em Ciência da Informação : além das nuvens, expandindo as fronteiras da Ciência da Informação, 27-31 de outubro em Belo Horizonte, MG. / Organizadores: Isa M. Freire, Lilian M. A. R. Álvares, Renata M. A. Baracho, Maurício B. Almeida, Beatriz V. Cendon, Benildes C. M. S. Maculan. – Belo Horizonte, ECI, UFMG, 2014.

ISSN 2177-3688

Evento realizado pela Associação Nacional de Pesquisa em Ciência da Informação (ANCIB) e organizado pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais (PPGCI-ECI/UFMG).

1. Evento – Ciência da Informação. 2. Evento – Pesquisa em Ciência da Informação. I. Título.

CDU: 02(063)(81)

CONVERSÃO DE REGISTROS EM XML PARA MARC 21: UM MODELO BASEADO EM XSLT¹⁵

RECORD CONVERSION FROM XML TO MARC 21: A MODEL BASED ON XSLT STYLESHEETS

Fabrcio Silva Assumpção
Plácida Leopoldina Ventura Amorim da Costa Santos

Resumo: Considerando a existência de sistemas de gerenciamento de bibliotecas que não utilizam os Formatos MARC 21 ou quaisquer outros padrões de metadados internacionalmente aceitos, este trabalho tem por objetivo elaborar um modelo para a conversão de registros de distintos padrões de metadados codificados com a *Extensible Markup Language* (XML) em registros nos Formatos MARC 21, tendo como um dos pontos centrais do modelo a utilização de folhas de estilo de transformação elaboradas com a linguagem *Extensible Stylesheet Language for Transformation* (XSLT). Após uma revisão de literatura, é apresentado o modelo e seus processos: mapeamento, elaboração de regras de conversão, elaboração da folha de estilo XSLT, verificação da folha de estilo, exportação e conversão dos registros. Como resultados alcançados, destaca-se que o modelo para a conversão de registros (1) busca apresentar uma generalização e uma formalização dos processos e dos componentes necessários à conversão de registros utilizando folhas de estilo XSLT e (2) apresenta novas possibilidades para o intercâmbio de dados que favorecem a interoperabilidade entre diferentes aplicações de informática que comunicam dados sobre recursos informacionais.

Palavras-chave: Conversão de registros. Padrões de metadados. Folhas de estilo XSLT. Formatos MARC 21. MARCXML.

Abstract: Considering the existence of integrated library systems that do not use neither MARC 21 Formats nor any other international metadata standard, this paper aims to formulate a model for conversion of records created using different metadata standards and encoded in Extensible Markup Language (XML) into MARC 21 records, by using stylesheets created with Extensible Stylesheet Language for Transformation (XSLT). After a literature review, we present the model and its process: mapping, formulation of the conversion rules, XSLT stylesheet formulation and test, and record exporting and conversion. As conclusions, we highlight that the model for record conversion (1) aims to present a generalization and formalization of the processes and components needed for record conversion using XSLT stylesheets and (2) presents new possibilities for the data interchange and supports interoperability between and among different computer applications that communicate data about information resources.

Keywords: Record conversion. Metadata standards. XSLT stylesheets. MARC 21 Formats. MARCXML.

1 INTRODUÇÃO

Uma vez que suas preocupações podem ser de naturezas diversas (BORKO, 1968, p. 3; LE COADIC, 2004, p. 25), a Ciência da Informação busca aporte em campos com

¹⁵ Pesquisa desenvolvida com financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

especificidades que auxiliam em suas discussões. Assim, a Ciência da Informação mostra-se como um campo interdisciplinar, relacionando-se com campos como a Biblioteconomia e a Ciência da Computação (SARACEVIC, 1996; LE COADIC, 2004). Como apontado por Alves (2010, p. 122), em razão de seu caráter interdisciplinar, a Ciência da Informação

utiliza-se, em parte, dos processos de tratamento descritivo da informação (TDI) desenvolvidos na disciplina de Catalogação na área de Biblioteconomia para solucionar os problemas inerentes a essa ciência: tornar acessível e disponível a variedade crescente de recursos informacionais, bem como promover uma melhora na recuperação desses recursos.

A catalogação, entendida como “um processo de representação documentária que desde a antiguidade atua como instrumento de acesso a informação e ao documento” (PEREIRA; SANTOS, 1998, p. 123), busca “individualizar um item documentário e ao mesmo tempo multidimensionar suas formas de acesso por meio de recursos tecnológicos” (PEREIRA; SANTOS, 1998, p. 122).

No domínio bibliográfico, a partir da década de 1960 surgiram padrões de metadados que, com as estruturas para a codificação, permitiram que os dados resultantes das atividades de catalogação fossem processados e intercambiados entre aplicações de informática. Atualmente, alguns dos principais padrões de metadados internacionalmente utilizados para o intercâmbio de registros no domínio bibliográfico são os Formatos MARC 21.

Os registros criados com esses padrões, de modo geral, ainda são codificados com a norma ISO 2709 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008), que define uma estrutura para a codificação semelhante àquela utilizada na década de 1960. Além da ISO 2709, os Formatos MARC 21 dispõem hoje de estruturas para codificação tecnologicamente mais vantajosas e mais condizentes com a atualidade, como é caso da *Extensible Markup Language* (XML).

No Brasil, o suporte a importação e a exportação de registros nos Formatos MARC 21 tem se tornado uma característica buscada nos sistemas de gerenciamento de bibliotecas (CAFÉ; SANTOS; MACEDO, 2001; CÔRTE et al., 1999; PARANHOS, 2004). No entanto, em uma análise exploratória, observa-se a existência de sistemas que não satisfazem esse requisito. Alguns desses sistemas, entre eles o *Personal Home Library* (PHL), utilizam padrões de metadados próprios.

A não utilização de padrões de metadados internacionalmente aceitos, como os Formatos MARC 21, reduz as possibilidades de intercâmbio de registros e pode ocasionar conflitos durante a migração entre sistemas de gerenciamento de bibliotecas. Diante dos

sistemas de gerenciamento de bibliotecas que não utilizam qualquer padrão de metadados internacionalmente aceito, tem-se como problema a questão: como converter os registros exportados por esses sistemas em registros nos Formatos MARC 21?

Observa-se que os sistemas de gerenciamento de bibliotecas que não utilizam padrões de metadados internacionalmente aceitos têm, às vezes, seus padrões de metadados próprios baseados nas tabelas de seus bancos de dados. Nesses casos, as diferenças entre os bancos de dados dos sistemas inviabilizam a criação de uma aplicação de informática que possa ser utilizada na conversão dos registros exportados a partir de diferentes bancos de dados. Assim, para cada banco de dados diferente é necessário criar uma nova aplicação ou modificar uma já existente, o que se mostra dispendioso (ZAFALON, 2012, p. 24, 27 e 28).

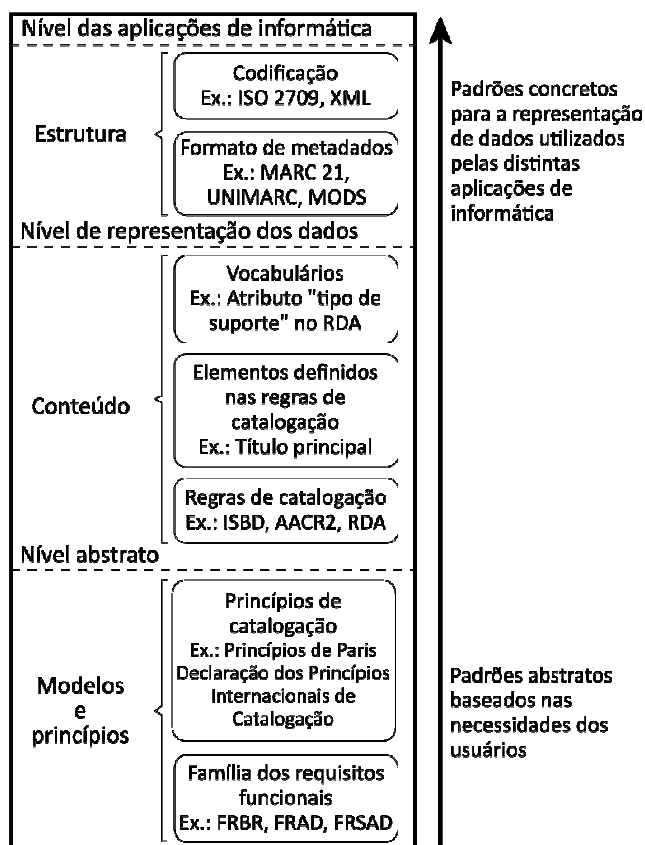
Por outro lado, as tecnologias atualmente disponíveis para a codificação trouxeram novas possibilidades para a conversão de registros entre padrões de metadados. Uma dessas tecnologias é a XML, que trouxe a possibilidade do uso de folhas de estilo elaboradas com a linguagem *Extensible Stylesheet Language for Transformation* (XSLT) na transformação de documentos XML de modo a atender distintos propósitos.

Considerando a provável efemeridade das aplicações de informática específicas para a conversão dos registros de cada sistema e a possibilidade do uso de folhas de estilo XSLT como instrumentos para a conversão, este trabalho tem por objetivo apresentar um modelo para a conversão de registros de distintos padrões de metadados codificados com a XML em registros nos Formatos MARC 21, tendo como um de seus pontos centrais a utilização de folhas de estilo de transformação elaboradas com a linguagem XSLT.

2 PADRÕES DE METADADOS E CODIFICAÇÕES

Os instrumentos de descrição utilizados no domínio bibliográfico podem ser categorizados de diversas formas. Entre as categorizações existentes, é utilizada aqui a categorização desenvolvida por Picco e Ortiz Repiso (2012, p. 149) apresentada na FIGURA 1.

FIGURA 1 – Instrumentos do domínio bibliográfico



Fonte: Picco e Ortiz Repiso (2012, p. 149, tradução nossa).

As autoras consideram três principais níveis: um nível abstrato abrangendo os modelos e os princípios do domínio bibliográfico; um nível de representação dos dados, com os instrumentos destinados ao conteúdo das representações; e um nível relacionado às aplicações de informática, compreendendo os formatos ou padrões de metadados e as codificações.

Uns dos principais padrões de metadados utilizados no domínio bibliográfico são os Formatos MARC 21. Um registro MARC 21, em qualquer um de seus formatos (para dados bibliográficos, de itens, de autoridade, de classificação ou para informação comunitária), envolve três componentes: a estrutura do registro, a designação do conteúdo e o conteúdo (LIBRARY OF CONGRESS, 1996). Seguindo a categorização de Picco e Ortiz Repiso (2012), a **estrutura do registro** é dada pela codificação, a **designação do conteúdo** é provida pelo padrão de metadados e o conteúdo está de acordo com regras de catalogação e vocabulários.

Enquanto padrão de metadados, cada Formato MARC 21 apresenta “um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados” (ALVES, 2010, p. 47-48). Em cada

Formato MARC 21, os metadados estão materializados nas designações de conteúdo, que podem ser de três tipos: etiqueta de campo (composta por três caracteres numéricos), indicador (composto por um caractere) ou código de subcampo (composto por um caractere alfanumérico).

A codificação, que permite que os metadados e os valores a eles atribuídos sejam processados por aplicações de informática, é entendida por Balby (1995, p. 31) como as regras para a estruturação dos dados em meio legível por computador. Para a autora, as codificações são definidas por normas, “conjuntos de regras, externos ao formato de intercâmbio, que orientam a distribuição dos dados no meio físico de armazenamento (fita, disquete etc.)”.

Os Formatos MARC 21 são derivados no Formato MARC, desenvolvido pela *Library of Congress* (LC) na década de 1960. Uma vez que, em sua origem, o MARC tinha entre seus objetivos possibilitar o intercâmbio de registros via fitas magnéticas, foi utilizada com esse padrão uma codificação condizente com o tipo de suporte em questão, a fita magnética.

A codificação de caráter sequencial utilizada com o MARC foi reconhecida primeiramente nos Estados Unidos como um padrão nacional, posteriormente, em 1981, foi reconhecida pela *International Organization for Standardization* (ISO) como a norma ISO 2709: *Documentation – Format for bibliographic information interchange on magnetic tape* (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008).

Apesar da norma ISO 2709 ter sido atualizada no decorrer dos anos, nota-se que não houve mudanças substanciais na codificação, de modo que os registros nos atuais Formatos MARC 21 são, em sua maior parte, codificados quase que da mesma forma com que eram codificados os registros na década de 1960, seja para propósitos de recuperação, por exemplo, via protocolo Z39.50, de importação e de exportação entre sistemas de gerenciamento de bibliotecas ou de armazenamento em bancos de dados.

O desenvolvimento das tecnologias de informática trouxe, além de outros suportes para o armazenamento e posterior intercâmbio de registros, outros ambientes digitais, como a Web. Com vistas ao uso das tecnologias de informática nos processos de catalogação, buscaram-se alternativas à codificação tradicionalmente utilizada com os registros nos Formatos MARC 21.

Algumas alternativas foram buscadas pela LC, que, na década de 1990, iniciou estudos sobre a utilização da *Standard Generalized Markup Language* (SGML) na codificação de registros. Com o surgimento da XML, a LC voltou-se para essa tecnologia e,

em 2002, lançou o *MARC 21 XML Schema*, mais conhecido como MARCXML: esquema XML que apresentava uma forma de codificação de registros MARC 21 utilizando a XML (LIBRARY OF CONGRESS, 2004, 2014).

A linguagem de marcação MARCXML foi pautada em uma série de considerações: ser um esquema simples e flexível para a codificação de registros MARC com a XML; não causar perdas no conteúdo quando os registros são convertidos a partir de registros ISO 2709; permitir a recriação de registros ISO 2709 a partir de registros MARCXML; facilitar a apresentação, a edição e a conversão dos dados por meio de folhas de estilo; permitir a validação dos registros; e facilitar a criação de ferramentas para a utilização, manipulação e conversão de registros MARC (LIBRARY OF CONGRESS, 2004).

3 CONVERSÃO RETROSPECTIVA E REUTILIZAÇÃO DE METADADOS

Em relação à conversão de registros no domínio bibliográfico, a literatura apresenta estudos e relatos que podem ser classificados em dois principais tipos: a **conversão retrospectiva** e a **reutilização de metadados** (*repurposing metadata*).

A conversão retrospectiva envolve a inserção dos dados de registros analógicos, de modo geral presentes em fichas catalográficas, em registros digitais (ASENSI ARTIGA; RODRÍGUEZ MUÑOZ, 2001; CASTRO, 2003; OLIVEIRA et al., 1998). Segundo Bowman (2007, p. 331), a conversão retrospectiva tem acompanhado a catalogação desde o início da automação de bibliotecas, o que ocorreu a partir da década de 1960 com a criação do Formato MARC e dos padrões dele derivados.

Devido à ampla utilização de sistemas digitais de gerenciamento de bibliotecas e de outras aplicações de informática destinadas ao gerenciamento, ao armazenamento e/ou à disseminação de recursos informacionais, por exemplo, os repositórios e as bibliotecas digitais, a literatura sobre conversão passou a preocupar-se não somente com a conversão retrospectiva (do analógico para o digital), mas também com a conversão dos registros já presentes no ambiente digital. Dada à existência de diversos padrões de metadados, tornou-se necessário converter os registros criados com um padrão em registros de acordo com outros padrões, por exemplo, para possibilitar o intercâmbio ou a migração desses registros entre sistemas que não adotam o mesmo padrão (ASENSI ARTIGA; RODRÍGUEZ MUÑOZ, 2001; RUDIĆ; SURLA, 2009; SCHMIDT; PATEL, 1999). Esse tipo de conversão tem sido chamado de reutilização de metadados e sua necessidade estende-se ainda mais devido a diversidade de padrões internacionais, nacionais e locais (WOODLEY, 2008).

Para Woodley (2008, p. 6), o processo de reutilização de metadados compreende um amplo conjunto de atividades: converter ou transformar registros de um padrão de metadados para outro; migrar de um padrão legado para outro; integrar registros criados de acordo com diferentes padrões; e coletar ou agregar registros criados utilizando um padrão compartilhado pela comunidade ou diversos padrões.

Na literatura são encontrados relatos e estudos sobre a reutilização de metadados tanto no contexto dos catálogos dos sistemas de gerenciamento de bibliotecas, quanto em outros ambientes informacionais, tais como os repositórios institucionais.

Averkamp e Lee (2009) apresentam um *workflow* para a reutilização dos metadados de teses e de dissertações, oriundos da base de dados *ProQuest UMI Dissertation Publishing*, na criação de registros para inserção em um repositório institucional e em um catálogo de biblioteca. Keenan (2010) relata a reutilização de registros no padrão Dublin Core da base de dados *U.S. Congressional Serial Set, 1817-1994* para a criação de registros bibliográficos no Formato MARC 21, visando à inserção desses em um catálogo de biblioteca. Essa base de dados oferece aos seus assinantes os registros em Dublin Core sem qualquer custo adicional, ao passo que vende seus registros no Formato MARC 21. A reutilização dos metadados nesse caso resultou em uma significativa redução dos custos, se comparada à compra dos registros já no Formato MARC 21.¹⁶

Na literatura brasileira são encontrados estudos e relatos sobre a conversão retrospectiva (CASTRO, 2003; DIAS, 1999; PEREZ; LIMA, 2002) e sobre a reutilização de metadados (BOICA; OLIVEIRA, 2008; MURAKAMI, 2012; RAPOSO; OLIVEIRA; SHINOTSUKA, 1985).¹⁷ Este estudo, partindo da necessidade de converter registros presentes em ambientes digitais, é considerado do tipo reutilização de metadados e seu principal resultado, um modelo para a conversão de registros, é apresentado na seção seguinte.

4 MODELO PARA A CONVERSÃO DE REGISTROS

Como abordado, a necessidade de converter registros digitais no domínio bibliográfico relaciona-se à utilização dos dados em uma aplicação de informática diferente daquela da qual os dados proveem, sendo necessário, portanto, considerar na conversão os

¹⁶ Segundo Keenan (2010), o custo estimado para a compra e a inserção dos registros seria de US\$ 25.669,71, enquanto que o custo da reutilização foi de US\$ 1.129,05, incluindo o tempo gasto pela equipe da biblioteca no planejamento, na pesquisa e no desenvolvimento de scripts.

¹⁷ Os estudos sobre a reutilização de metadados, principalmente os nacionais, não empregam essa denominação. Em vez disso, utilizam apenas os termos conversão ou migração.

padrões de metadados e as codificações aceitas pela aplicação de informática que receberá os registros.

Em razão da ampla utilização da ISO 2709 para o intercâmbio de dados no domínio bibliográfico, algumas aplicações de informática, principalmente do tipo sistema de gerenciamento de bibliotecas, possibilitam a importação de registros MARC 21 apenas se codificados de acordo com tal norma. Outras aplicações, no entanto, permitem a importação de registros MARC 21 apenas se codificados com a XML e seguindo o MARCXML.

Considerando essas duas situações, o modelo para a conversão de registros foi elaborado diante da seguinte necessidade: registros de acordo com um padrão de metadados qualquer codificados com a XML precisam ser convertidos em registros em um dos Formatos MARC 21, codificados com a XML (MARCXML) ou com a ISO 2709.

Uma vez que os registros de origem estarão codificados com a XML e os registros de destino poderão também estar codificados com essa tecnologia, foi incluída no modelo uma folha de estilo criada com a *Extensible Stylesheet Language for Transformation* (XSLT).

A linguagem XSLT pode ser utilizada na criação de folhas de estilo (documentos com regras que deverão ser executadas por uma aplicação de informática específica) para a transformação de um documento XML que está de acordo com uma linguagem de marcação em um documento XML de acordo com outra linguagem de marcação ou mesmo em um documento que não seja XML (RAY, 2001; TIDWELL, 2008; W3C, 2007).

Eito Brun (2008, p. 95) destaca a importância da XSLT para as situações em que a XML é utilizada para transferir e intercambiar dados estruturados com diferentes esquemas, ou seja, com diferentes linguagens de marcação. Tais situações requerem uma tecnologia que permita transformar os documentos baseados em um esquema em outro diferente. Para essa transformação uma “folha de estilo XSLT estabelecerá as equivalências entre os elementos e os atributos do esquema em que estão os documentos XML e os elementos e os atributos do esquema ao qual se deseja converter” (EITO BRUN, 2008, p. 95, tradução nossa).

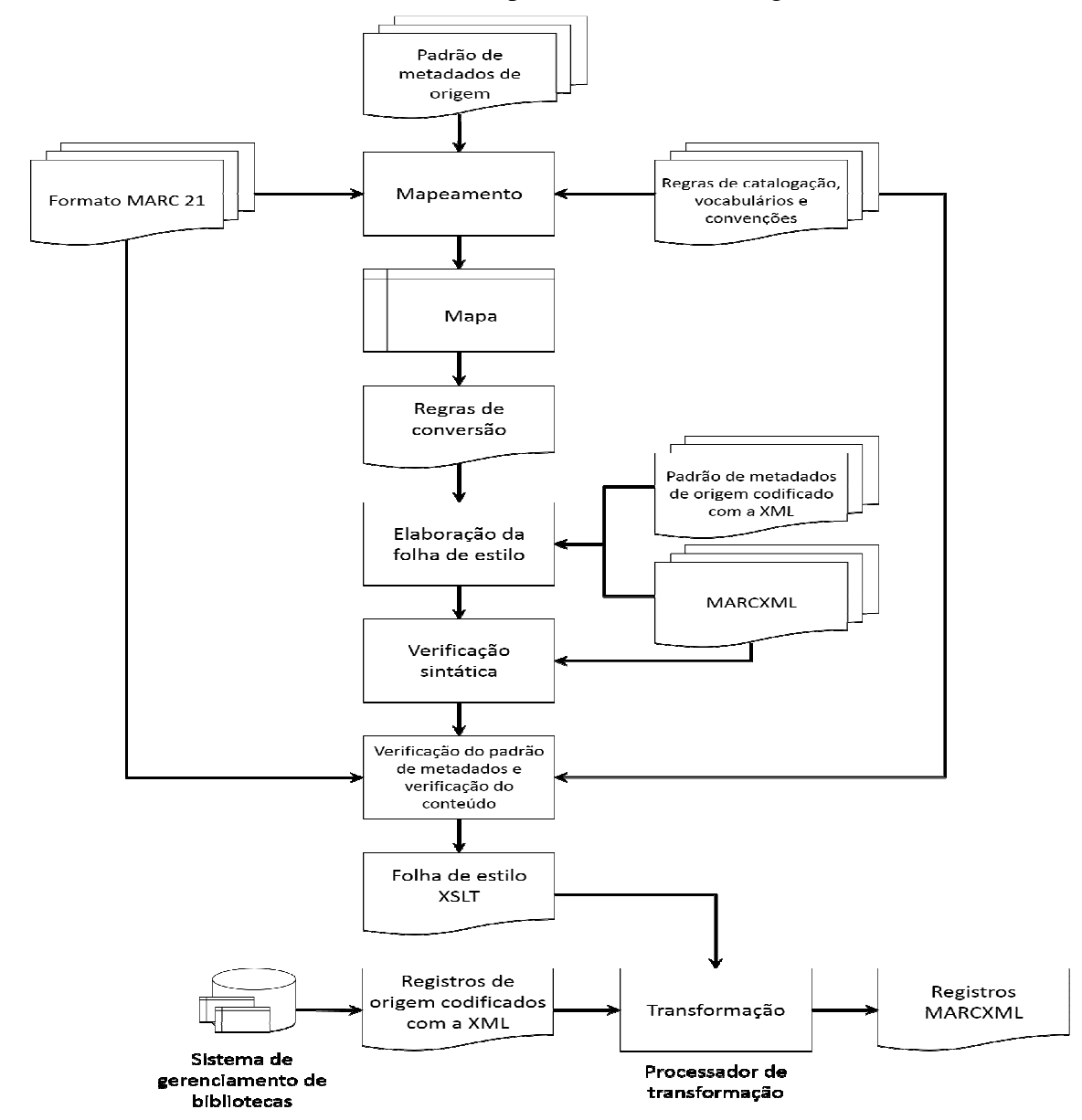
A utilização de folhas de estilo XSLT na conversão de registros é relatada em alguns estudos internacionais que consistem, principalmente, em relatos demonstrando os procedimentos e os instrumentos utilizados (KEITH, 2004; KEENAN, 2010; KURTH; RUDDY; RUPP, 2004; RUDIĆ; SURLA, 2009). Sobre a literatura nacional, concorda-se com os apontamentos realizados por Zafalon (2012, p. 22) de que,

Na literatura nacional, diagnósticos e situações de conversão de bases de dados são relatados no sentido de apresentar as experiências e os

procedimentos metodológicos utilizados em diversas instituições. Estas publicações, de certa forma, tendem a deixar de explicitar as concepções de ordem teórico-metodológica envolvidas nesse processo.

Apesar de, de modo geral, apresentarem seqüências de procedimentos e conjuntos de instrumentos mais ou menos semelhantes, nota-se nos estudos sobre a conversão de registros pouca ou nenhuma sistematização de um modelo ou de uma estrutura teórico-metodológica que compreenda os procedimentos e os instrumentos que comumente relacionam-se à conversão. Assim, para suprir essa lacuna e auxiliar na conversão de registros para os Formatos MARC 21, é apresentado na FIGURA 2 o modelo para a conversão de registros resultado deste estudo.

FIGURA 2 – Modelo para a conversão de registros



O modelo apresentado na FIGURA 2 é sintetizado da seguinte forma:

- a documentação do padrão de metadados de origem, a documentação do Formato MARC 21 de destino, as regras de catalogação, os vocabulários e as convenções são utilizados no mapeamento, que tem como resultado um mapa indicando as correspondências existentes entre os metadados dos padrões de origem e de destino;
- as correspondências indicadas no mapa são redigidas como regras de conversão;
- as regras de conversão, a documentação da codificação do padrão de metadados de origem com a XML e a documentação do MARCXML são utilizadas na elaboração da folha de estilo XSLT;
- durante a elaboração da folha de estilo XSLT são realizadas a verificação sintática, a verificação do padrão de metadados e a verificação do conteúdo:
- um editor XML realiza a verificação sintática com base no esquema do MARCXML, assegurando que os registros obtidos a partir da conversão utilizando a folha de estilo estão em acordo com o MARCXML;
- um agente humano realiza a verificação do padrão de metadados, com base na documentação do Formato MARC 21, e a verificação do conteúdo, com base nas regras de catalogação, nos vocabulários e nas convenções, assegurando, assim, que os registros obtidos a partir da conversão estão de acordo com o padrão de metadados de destino e com os instrumentos de descrição;
- a folha de estilo XSLT já finalizada é inserida no processador de transformação;
- os registros de origem são exportados a partir do sistema de gerenciamento de bibliotecas e inseridos no processador de transformação;
- o processador transforma os registros de origem em registros no Formato MARC 21 codificados de acordo com o MARCXML.

Para uma maior elucidação do modelo apresentado na Figura 2 seus processos e seus componentes são descritos nas seções seguintes.

4.1 Mapeamento e elaboração das regras de conversão

O processo de estabelecer os relacionamentos entre metadados semanticamente equivalentes em diferentes padrões é denominado na literatura como mapeamento (*mapping* ou *crosswalking*). O resultado do mapeamento é chamado de mapa (*map* ou *crosswalk*) e consiste em uma representação visual dos relacionamentos, das equivalências e das lacunas entre os padrões de metadados mapeados (KURTH; RUDDY; RUPP, 2004, p. 154; ST. PIERRE; LAPLANT, 1998; WOODLEY, 2008, p. 3).

St. Pierre e LaPlant (1998) destacam que o mapeamento inclui um mapa semântico e especificações para a conversão que indicam as transformações requeridas para converter o conteúdo do registro do padrão de origem em um registro de acordo com o padrão de destino.

Rudić e Surla (2009, p. 952) apontam que uma descrição de conversão pode conter uma tabela e regras de conversão. Essa tabela de conversão é o mapa, e as regras de conversão são as “especificações para a conversão” (ST. PIERRE; LAPLANT, 1998).

Para Machovec (2002, p. 2), o mapeamento entre quaisquer padrões não será perfeito e será mais ou menos bem sucedido dependendo dos padrões mapeados. Um dos fatores que influenciam o mapeamento e o sucesso da conversão é o nível de granularidade dos padrões de metadados mapeados (MACHOVEC, 2002, p. 1), portanto, esse é um importante aspecto a ser levado em conta.

Segundo Alves, Simionato e Santos (2012, p. 3), o termo granularidade tem origem na Ciência da Computação, porém passa ser utilizado também no universo bibliográfico para referir-se aos níveis de detalhe em que um recurso informacional pode ser descrito. Para as autoras, a granularidade é dividida em dois níveis: a granularidade fina (*fine-granularity*), quando a descrição apresenta um alto nível de detalhamento, e a granularidade grossa (*coarse-granularity*), quando a descrição possui um baixo nível de detalhamento.

Woodley (2008, p. 7) destaca que os metadados no padrão de origem podem não estar bem definidos ou podem conter uma mistura de dados que, no padrão de destino, podem estar em diferentes campos. Nesses casos, a identificação de um dado dentro do conteúdo de um metadado pode não ser possível, pode requerer a manipulação dos dados diversas vezes, ser difícil, demorado e repleto de erros (MACHOVEC, 2002, p. 2; WOODLEY, 2008, p. 7).

Considerando os diferentes níveis de granularidade entre os padrões, as correspondências do tipo *um-para-muitos* e *muitos-para-um* são frequentes. Woodley (2008, p. 7) destaca a falta de correspondências perfeitas entre os metadados, o que requer o estabelecimento de correspondências entre metadados com significados próximos.

Algum metadado do padrão de origem pode não possuir correspondente no padrão de destino, ocasionando no descarte de seu conteúdo durante a conversão. Situações inversas também são possíveis: algum metadado do padrão de destino pode não ter um correspondente no padrão de origem. Nesses casos, se o metadado não é obrigatório, pode ser deixado em branco, caso seja obrigatório, é necessário prover o conteúdo do mesmo, por

exemplo, por meio de um valor padrão (*default*) (MACHOVEC, 2002; ST. PIERRE; LAPLANT, 1998).

Outro ponto a ser observado, segundo St. Pierre e LaPlant (1998), são as propriedades dos metadados do padrão de origem e do padrão de destino, por exemplo, obrigatoriedade, repetitividade e tipo de dado.

Apesar dos padrões de metadados existirem de forma independente dos instrumentos de descrição (regras de catalogação, vocabulários e convenções) e vice-versa, observa-se que os padrões de metadados, em razão do conjunto de metadados de que dispõem e do modo com que tais metadados estão arrançados, condicionam a utilização dos instrumentos de descrição, e, em decorrência da oferta de mais ou de menos possibilidades que aquelas previstas pelos instrumentos de descrição, influenciam o modo com que os dados são registrados. Assim, entendendo que o modo com que um dado é registrado em um padrão pode diferir do modo como que o mesmo dado é registrado em outro, é necessário considerar no mapeamento os instrumentos de descrição utilizados na criação dos registros a serem convertidos.

Durante a elaboração do mapa e após a sua finalização, devem ser redigidas regras de conversão de modo a facilitar a criação da folha de estilo XSLT pelo profissional da Ciência da Computação. Essas regras são redigidas em linguagem natural indicando os relacionamentos fornecidos no mapa e, opcionalmente, provendo orientações sobre os instrumentos de descrição utilizados na criação dos registros de origem.

A importância dessas regras está no fato de que o profissional da Ciência da Computação pode não possuir o conhecimento suficiente dos padrões de metadados e dos instrumentos de descrição para interpretar o mapa.

4.2 Elaboração da folha de estilo XSLT

Além do mapa e das regras de conversão, para a elaboração da folha de estilo é necessário (1) escolher um editor XML, um processador de transformação e a versão da XSLT, e (2) conhecer as especificações das linguagens de marcação dos padrões de metadados de origem e de destino, além do conhecimento de XSLT. Essas especificações definem o modo com que os metadados compreendidos pelos padrões e os valores a eles atribuídos são expressos na XML.

A especificação da linguagem de marcação utilizada com o padrão de metadados de origem nem sempre está disponível em uma *Document Type Definition* (DTD) ou em esquema XML, ou mesmo acessível ao profissional usuário do sistema de gerenciamento de bibliotecas. Nesses casos, uma possibilidade é apreender a estrutura definida pela

linguagem de marcação a partir dos documentos XML que estão de acordo com ela, ou seja, a partir da análise dos registros exportados em XML.

A elaboração da folha de estilo compreende principalmente dois passos:

- a inclusão da declaração XML, a inclusão do elemento do documento dentro do qual será inserido todo o conteúdo da folha de estilo, e a inclusão das declarações dos *namespaces*;
- a elaboração das regras de transformação com base no mapa e nas regras de conversão.

As regras de transformação selecionam e manipulam os elementos e os conteúdos do documento de entrada fazendo com que sejam organizados e/ou transformados para se adequarem aos elementos do documento de saída; algumas das ações realizadas pelas regras de transformação são:

- incluir o conteúdo de um elemento em um elemento de saída;
- reunir o conteúdo de dois ou mais elementos para formar o conteúdo de um elemento de saída;
- separar o conteúdo de um elemento e distribuí-lo entre dois ou mais elementos de saída;
- editar o conteúdo de um elemento (substituir, acrescentar ou remover caracteres) para formar o conteúdo de um elemento de saída;
- criar um elemento de saída com um conteúdo que não pôde ser obtido a partir do documento de entrada;
- incluir, reunir, separar, editar e criar elementos e conteúdos com base em condições.

4.3 Verificação da folha de estilo

Durante a elaboração da folha de estilo, é necessário verificar os registros resultantes da conversão para checar se estão de acordo com o resultado esperado. A verificação busca assegurar que os registros convertidos com a folha de estilo estão de acordo com o MARCXML, com o padrão de metadados (um dos Formatos MARC 21) e com os instrumentos de descrição. Assim, a verificação inclui três etapas que proveem um *feedback* indicando se há ou não a necessidade de modificações no mapa, nas regras de conversão e/ou na folha de estilo:

- **verificação sintática:** realizada por editores XML, é a análise do registro MARCXML para verificar se o mesmo está de acordo com a linguagem de marcação MARCXML;

- **verificação do padrão de metadados:** é a análise das designações do conteúdo (etiquetas, indicadores e códigos de subcampo) presentes nos registros convertidos para verificar se estão de acordo com o padrão de metadados;
- **verificação do conteúdo:** é a confrontação dos dados presentes nos registros convertidos com as regras de catalogação, os vocabulários e as convenções utilizadas pela instituição; tem por objetivo checar a adequação dos registros aos instrumentos de descrição, mostrando se o resultado da conversão atende ou não aos requisitos de qualidade definidos nesses instrumentos.

4.4 Exportação e conversão dos registros

Entre os processos do modelo, a exportação dos registros é o mais dependente do sistema de gerenciamento de bibliotecas. Para a utilização do modelo proposto é necessário que os registros exportados estejam codificados com a XML, não importando qual seja a linguagem de marcação ou o padrão de metadados utilizado com essa codificação. Assim, se os registros não estão codificados com a XML e não há qualquer possibilidade de assim codificá-los, o modelo para a conversão de registros não poderá ser aplicado tal como foi elaborado.

Entende-se que a exportação pode ocorrer de dois modos. O primeiro é por meio de uma interface, seja ela gráfica ou não, provida pelo sistema de gerenciamento de bibliotecas. Um segundo modo ocorre quando o sistema não oferece uma interface de exportação, mas é possível acessar sua base de dados por meio do sistema de gerenciamento de banco de dados e, então, exportar os registros codificados com a XML.

Com a folha de estilo XSLT elaborada e os registros exportados, o próximo processo é converter os registros do padrão de metadados de origem codificados com a XML em registros no Formato MARC 21 de acordo com o MARCXML. Os componentes necessários a essa conversão são: a folha de estilo XSLT, um processador de transformação compatível com a versão da XSLT utilizada na folha de estilo e o documento XML contendo os registros exportados.

De modo geral, a folha de estilo XSLT e o documento XML contendo os registros exportados são inseridos no processador que, a partir de algum comando, inicia a transformação. Ao final da transformação, o processador gera um documento XML contendo os registros convertidos de acordo com o MARCXML.

A conversão dos registros MARCXML para ISO 2709 é um processo opcional, realizado somente quando é necessário obter registros nos Formatos MARC 21 codificados

com a ISO 2709. Para essa segunda conversão pode ser utilizada a aplicação disponibilizada pela LC¹⁸ ou a que acompanha a suíte de aplicativos gratuita MarcEdit¹⁹, entre outras.

Para a validação do modelo elaborado, realizou-se sua aplicação na conversão de registros bibliográficos exportados pelo sistema de gerenciamento de bibliotecas *Personal Home Library* (PHL). Os procedimentos utilizados na aplicação do modelo e os resultados obtidos são detalhados em Assumpção (2013, p. 81-101).

5 CONSIDERAÇÕES FINAIS

A existência de sistemas de gerenciamento de bibliotecas que não possibilitam a exportação dos dados nos Formatos MARC 21 e nem em qualquer padrão de metadados internacionalmente utilizado, fomenta a necessidade de estudos sobre a reutilização de metadados. Essa necessidade está pautada na consideração de que as instituições usuárias desses sistemas, por não disporem de seus registros de acordo com padrões de metadados internacionalmente aceitos, poderão estar impossibilitadas de participarem de programas de catalogação cooperativa e estarão sujeitas a problemas e a retrabalhos durante a migração entre sistemas de gerenciamento de bibliotecas.

Diante dessa consideração, o objetivo deste estudo foi elaborar um modelo para a conversão de registros de distintos padrões de metadados codificados com a XML em registros nos Formatos MARC 21, tendo como um dos pontos centrais do modelo a utilização de folhas de estilo de transformação elaboradas com a linguagem XSLT.

Sobre o modelo elaborado e sua aplicação na conversão de registros bibliográficos do sistema PHL, são traçadas as seguintes considerações:

- o modelo busca apresentar uma generalização e uma formalização dos processos e dos componentes necessários à conversão de registros em XML utilizando folhas de estilo XSLT, processos e componentes estes até então dispersos na literatura e abordados no contexto de realidades individuais;
- nos ambientes digitais, a conversão de registros não deve ser uma tarefa realizada apenas por profissionais da Ciência da Informação ou apenas da Ciência da Computação. A Ciência da Informação possui o conhecimento dos padrões de metadados e dos instrumentos de descrição, portanto, tem a competência necessária para traçar mapas estabelecendo correspondências entre os metadados de diferentes

¹⁸ Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 6 jul. 2014.

¹⁹ Disponível em: <<http://marcedit.reeset.net>>. Acesso em: 6 jul. 2014.

padrões, ao passo que à Ciência da Computação cabe a transposição dos mapas para aplicações de informática que, manipulando os registros do padrão de metadados de origem, possam convertê-los para o padrão de metadados de destino. Nesse sentido, o modelo proposto é uma tentativa de delineamento dos papéis exercidos pelos profissionais dessas áreas durante as atividades de conversão de registros;

- entre os fatores que interferem no tempo demandado na elaboração de uma folha de estilo XSLT estão a complexidade dos padrões de origem e de destino, incluindo o número de campos/subcampos, a especificidade e o nível de granularidade dos padrões, e o quão mínima é a perda de dados tolerada;
- o grau de complexidade das instruções da folha de estilo relaciona-se diretamente à perda de dados na conversão. Quanto maior a manipulação (junção, separação, substituição, etc.) dos dados do documento XML de origem, melhor esses dados poderão ser alocados no padrão de metadados de destino;
- em razão das diferenças entre os padrões de origem e de destino, após a conversão, pode ser necessário completar ou modificar os registros convertidos, o que deve ser levado em conta pelas instituições em seus planos para conversão; e
- ao considerar o tempo demandado na elaboração da folha de estilo, deve ser levada em conta a quantidade de registros que serão convertidos, de modo a melhor ponderar sobre a relação custo-benefício da conversão utilizando o modelo.

A proposição inicial desta pesquisa era a elaboração de um modelo para a conversão que favorecesse o intercâmbio de registros e minimizasse os conflitos durante a migração de registros de sistemas que não utilizam padrões de metadados internacionalmente aceitos.

Os resultados obtidos com a utilização do modelo são registros em um dos Formatos MARC 21 e codificados com a XML seguindo o MARCXML. Tais registros podem ser codificados com a ISO 2709 de modo a atender a uma necessidade específica, por exemplo, a importação em um sistema de gerenciamento de bibliotecas. No entanto, uma vez que os registros convertidos estão de acordo com o MARCXML, uma ampla gama de possibilidades passa a existir, principalmente devido às possibilidades trazidas pela XML (a utilização de folhas de estilo XSLT na transformação de documentos é uma dessas possibilidades) e ao papel que os Formatos MARC 21 podem desempenhar como padrões capazes de intermediar a troca de registros entre diferentes padrões de metadados no domínio bibliográfico.

Ou seja, além de prover uma solução para o problema de como converter os registros exportados em registros nos Formatos MARC 21, esta pesquisa apresentou novas

possibilidades para o intercâmbio de dados que favorecem a interoperabilidade entre diferentes aplicações de informática que comunicam dados sobre recursos informacionais. Nesse sentido, observa-se que, apesar de estar voltado ao contexto dos Formatos MARC 21, o modelo elaborado pode ser aplicado na conversão de registros para outros padrões de metadados, sendo essa possibilidade uma consideração a respeito da aplicação do modelo em outras comunidades além da comunidade de bibliotecas.

Por fim, cabe ressaltar que este estudo não defende que o profissional da Ciência da Informação seja um especialista na construção de folhas de estilo, mas sim que tal profissional tenha conhecimento das tecnologias desenvolvidas na Ciência da Computação que possam ser úteis às suas atividades, favorecendo, assim, o uso estratégico dessas tecnologias em prol do desenvolvimento e do gerenciamento de ambientes informacionais digitais.

REFERÊNCIAS

- ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <<http://base.repositorio.unesp.br/handle/unesp/103361>>. Acesso em: 7 jul. 2014.
- ALVES, R. C. V.; SIMIONATO, A. C.; SANTOS, P. L. V. A. da C. Aspectos de granularidade na representação da informação no universo bibliográfico. In: ENCONTRO NACIONAL DE CATALOGADORES, 1., 2012, Rio de Janeiro. **Trabalhos...** Rio de Janeiro: Fundação Biblioteca Nacional, 2012. Disponível em: <<http://pt.scribd.com/doc/109274547/Aspectos-de-granularidade-na-representacao-da-informacao-no-universo-bibliografico>>. Acesso em: 6 jul. 2014.
- ASENSI ARTIGA, V.; RODRÍGUEZ MUÑOZ, J. V. El proceso de catalogación automatizada. In: PINTO MOLINA, M. (Org.). **Catalogación de documentos: teoría y práctica**. 2. ed. rev. e atual. Madrid: Síntesis, 2001. p. 105-143.
- ASSUMPCÃO, F. S. **Conversão de registros em XML para MARC 21: um modelo baseado em folhas de estilo XSLT**. 135 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013. Disponível em: <<http://base.repositorio.unesp.br/handle/unesp/93658>>. Acesso em: 14 maio 2014.
- AVERKAMP, S.; LEE, J. Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository. **Code4Lib Journal**, n. 7, 2009. Disponível em: <<http://journal.code4lib.org/articles/1647>>. Acesso em: 6 jul. 2014.
- BALBY, C. N. Formatos de intercâmbio de registros bibliográficos: conceitos básicos. **Cadernos da F.F.C.**, v. 4, n. 1, p. 29-35, 1995.
- BOICA, A. L.; OLIVEIRA, L. H. M. de. Conversão de metadados do padrão Dublin Core para o RDF. **Global Science and Technology**, v. 1, n. 2, p. 8-13, dez./mar. 2008.

Disponível em: <<http://rioverde.ifgoiano.edu.br/periodicos/index.php/gst/article/view/13>>. Acesso em: 6 jul. 2014.

BORKO, H. Information Science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, 1968.

BOWMAN, J. H. Retrospective conversion: The early years. **Library History**, v. 23, p. 331-340, Dec. 2007. <<http://dx.doi.org/10.1179/174581607x254811>>. Acesso em: 06 jul. 2014.

CAFÉ, L.; SANTOS, C. dos; MACEDO, F. Proposta de um método para escolha de software de automação de bibliotecas. **Ciência da Informação**, v. 30, n. 2, p. 70-79, maio/ago. 2001. Disponível em: <<http://dx.doi.org/10.1590/S0100-19652001000200009>>. Acesso em: 6 jul. 2014.

CASTRO, F. F. de. Conversão retrospectiva de registros bibliográficos. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 24., 2003, Belo Horizonte. **Anais...** São Paulo: Intercom, 2003. Disponível em: <<http://www.portcom.intercom.org.br/navegacaoDetalhe.php?option=trabalho&id=42503>>. Acesso em: 6 jul. 2014.

CÔRTE, A. R. et al. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de softwares. **Ciência da Informação**, v. 28, n. 3, p. 241-256, set./dez. 1999. Disponível em: <<http://dx.doi.org/10.1590/S0100-19651999000300002>>. Acesso em: 06 jul. 2014.

DIAS, M. do R. I. **Catálogo e qualidade**: breve estudo. Marília: UNESP/CGB, 1999.

EITO BRUN, R. **Lenguajes de marcas para la gestión de recursos digitales**: aproximación técnica, especificaciones y referencia. Gijón: Trea, 2008.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2709:2008**: Abstract. Geneva, 2008. Disponível em: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=41319>. Acesso em: 6 jul. 2014.

KEENAN, T. M. Why Purchase When You Can Repurpose? Using Crosswalks to Enhance User Access. **Code4Lib Journal**, n. 11, 2010. Disponível em: <<http://journal.code4lib.org/articles/3604>>. Acesso em: 6 jul. 2014.

KEITH, C. Using XSLT to manipulate MARC metadata. **Library Hi Tech**, v. 22, n. 2, p. 122-130, 2004. Disponível em: <<http://dx.doi.org/10.1108/07378830410524549>>. Acesso em: 6 jul. 2014.

KURTH, M.; RUDDY, D.; RUPP, N. Repurposing MARC metadata: using digital project experience to develop a metadata management design. **Library Hi Tech**, v. 22, n. 2, p. 153-165, 2004. Disponível em: <<http://dx.doi.org/10.1108/07378830410524585>>. Acesso em: 6 jul. 2014.

LE COADIC, Y-F. **A ciência da informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

LIBRARY OF CONGRESS. **MARC 21 XML Schema**: official web site. Washington, D.C., 2014. Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 6 jul. 2014.

LIBRARY OF CONGRESS. **MARC XML Design Considerations**. Washington, D.C., 2004. Disponível em: <<http://www.loc.gov/standards/marcxml/marcxml-design.html>>. Acesso em: 6 jul. 2014.

LIBRARY OF CONGRESS. **The MARC 21 Formats: Background and Principles**. Revised November 1996. Washington, D.C.: Library of Congress, 1996. Disponível em: <<http://www.loc.gov/marc/96principl.html>>. Acesso em: 6 jul. 2014.

MACHOVEC, G. S. Challenges and issues with metadata crosswalks. **Information Intelligence Online Libraries and Microcomputers**, v. 20, n. 4, p. 1-3, Apr. 2002.

MURAKAMI, T. Migração de registros de uma Tabela para o MARC21. In: **BIBLIOTECÁRIOS sem fronteiras**. [S.l.: s.n.], 2012. Disponível em: <<http://bsf.org.br/2012/01/16/migracao-de-registros-de-uma-tabela-para-o-marc21>>. Acesso em: 6 jul. 2014.

OLIVEIRA, N. M. et al. Compact Disc Cataloging - CatCD: análise de um instrumento para conversão retrospectiva no Sistema de Bibliotecas da UNICAMP. **Perspectivas em Ciência da Informação**, v. 3, n. 1, p. 41-46, jan./jun. 1998. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/608>>. Acesso em: 6 jul. 2014.

PARANHOS, W. M. M. da R. Fragmentos metodológicos para projetos e execução de gestão informatizada de coleções de documentos e serviços em bibliotecas. **Encontros BIBLI**, n. esp., p. 14-32, 2º sem. 2004. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2004v9nesp2p14>><<http://www.loc.gov/marc/96principl.html>>. Acesso em: 6 jul. 2014.

PEREIRA, A. M.; SANTOS, P. L. V. A. da C. O uso estratégico das tecnologias em catalogação. **Cadernos da F.F.C.**, v. 7, n. 1/2, p. 121-131, 1998.

PEREZ, D. R.; LIMA, P. O projeto de conversão retrospectiva de registros bibliográficos: uma experiência do sistema de bibliotecas da Pontifícia Universidade Católica do Rio de Janeiro. In: **SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS**, 12., 2002, Recife. **Anais...** Recife: UFPE, 2002. Disponível em: <<http://alfarrabiosroger.files.wordpress.com/2009/12/37-a.pdf>>. Acesso em: 6 jul. 2014.

PICCO, P.; ORTIZ REPISO, V. RDA, el nuevo código de catalogación: cambios y desafíos para su aplicación. **Revista Española de Documentación Científica**, v. 35, n. 1, p. 145-173, enero-marzo 2012. Disponível em: <<http://dx.doi.org/10.3989/redc.2012.1.848>>. Acesso em: 6 jul. 2014.

RAPOSO, M. de F. P.; OLIVEIRA, V. L. S. de; SHINOTSUKA, F. H. Mudança para o formato CALCO: uma experiência. **Revista de Biblioteconomia de Brasília**, v. 13, n. 1, p. 21-26, jan./jun. 1985.

RAY, E. T. **Aprendendo XML**. Rio de Janeiro: Campus, 2001.

RUDIĆ, G.; SURLA, D. Conversion of bibliographic records to MARC 21 format. **The Electronic Library**, v. 27, n. 6, p. 950-967, 2009. Disponível em: <<http://dx.doi.org/10.1108/02640470911004057>>. Acesso em: 6 jul. 2014.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>>. Acesso em: 6 jul. 2014.

SCHMIDT, N.; PATEL, A. MARC record conversion: a generalized approach. **Computer Standards & Interfaces**, v. 21, p. 287-297, 1999. Disponível em: <[http://dx.doi.org/10.1016/S0920-5489\(99\)00007-0](http://dx.doi.org/10.1016/S0920-5489(99)00007-0)>. Acesso em: 6 jul. 2014.

ST. PIERRE, M.; LAPLANT, W. P. **Issues in Crosswalking Content Metadata Standards**. Baltimore: National Information Standards Organization, 1998. Disponível em: <http://www.niso.org/publications/white_papers/crosswalk>. Acesso em: 6 jul. 2014.

TIDWELL, D. **XSLT**. 2nd ed. Beijing: O'Reilly, 2008.

W3C. **XSL Transformations (XSLT) Version 2.0**: W3C Recommendation 23 January 2007. Cambridge, 2007. Disponível em: <<http://www.w3.org/TR/xslt20>>. Acesso em: 6 jul. 2014.

WOODLEY, M. S. Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information. In: BACA, M. (Org.). **Introduction to Metadata**. 2nd ed. Los Angeles: Getty Research Institute, 2008. Disponível em: <http://www.getty.edu/research/publications/electronic_publications/intrometadata/path.pdf>. Acesso em: 6 jul. 2014.

ZAFALON, Z. R. **Scan for MARC**: princípios sintáticos e semânticos de registros bibliográficos aplicados à conversão de dados analógicos para o Formato MARC21 bibliográfico. 2012. 169 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012. Disponível em: <<http://base.repositorio.unesp.br/handle/unesp/103386>>. Acesso em: 6 jul. 2014.