



Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Filosofia e Ciência
Campus de Marília

Fabício Silva Assumpção

Conversão de registros em XML para MARC 21:
um modelo baseado em folhas de estilo XSLT

Marília
2013



Universidade Estadual Paulista “Júlio de Mesquita Filho”
Faculdade de Filosofia e Ciência
Campus de Marília

Fabício Silva Assumpção

**Conversão de registros em XML para MARC 21:
um modelo baseado em folhas de estilo XSLT**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da UNESP, Faculdade de Filosofia e Ciências, Campus de Marília, como exigência para a obtenção do título de Mestre em Ciência da Informação.

Linha de pesquisa: Informação e Tecnologia.

Orientadora: Dra. Plácida Leopoldina Ventura Amorim da Costa Santos.

Financiamento: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Marília
2013

Assumpção, Fabrício Silva.

A851c Conversão de registros em XML para MARC 21 : um modelo baseado em folhas de estilo XSLT / Fabrício Silva Assumpção. – Marília, 2013. 135 f. ; 30 cm.

Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2013.

Bibliografia: f. 107-114.

Orientadora: Plácida Leopoldina Ventura Amorim da Costa Santos.

Financiamento: CAPES.

1. Registros bibliográficos. 2. Conversão de metadados. 3. Formatos MARC. 4. XML (Linguagem de marcação de documentos). 5. ISO 2709. 6. XSLT (Linguagem de programação de computador). 7. PHL – Personal Home Library. 8. Migração de dados. 9. Catalogação descritiva. I. Título.

CDD 025.302854

Agradecimentos

Agradeço a minha Mãe, aos meus irmãos e a todos meus familiares que sempre me incentivaram em meus estudos.

À Professora Plácida, pela oportunidade de mais uma vez contar com sua orientação e pela confiança depositada no desenvolvimento desta pesquisa.

Aos colegas do Grupo de Pesquisa Novas Tecnologias em Informação (GPNTI), que acompanharam o desenvolvimento desta pesquisa e têm feito parte de minha formação pessoal, acadêmica e profissional.

Aos professores doutores Zaira Regina Zafalon, Silvana Aparecida Borsetti Gregorio Vidotti, Ricardo Cesar Gonçalves Santana e Maria Elisabete Catarino, titulares e suplentes no exame de qualificação e na defesa, pelas importantes contribuições ao desenvolvimento e à conclusão desta pesquisa.

Aos colegas do Programa de Pós-Graduação em Ciência da Informação da UNESP, docentes e discentes, com os quais tive a oportunidade de conviver e aprender.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro para a realização desta pesquisa.

Resumo

Os Formatos MARC 21 são padrões de metadados internacionalmente utilizados no domínio bibliográfico para o intercâmbio de registros. Apesar da importância desses e de outros padrões de metadados internacionalmente aceitos, existem sistemas de gerenciamento de bibliotecas que utilizam padrões de metadados próprios, o que reduz as possibilidades de intercâmbio de registros e pode ocasionar conflitos durante a migração entre sistemas. Diante dos sistemas de gerenciamento de bibliotecas que não utilizam qualquer padrão de metadados internacionalmente aceito, tem-se como problema a conversão dos registros desses sistemas em registros nos Formatos MARC 21. Para que possam ser processados, os registros nos Formatos MARC 21 devem estar codificados de forma que seu conteúdo possa ser identificado por aplicações de informática. As tecnologias para essa codificação têm evoluído da tradicional norma ISO 2709 até a *Extensible Markup Language* (XML). A codificação com a XML trouxe a possibilidade de converter os registros utilizando folhas de estilo de transformação elaboradas com a linguagem *Extensible Stylesheet Language for Transformation* (XSLT). Considerando a necessidade de converter registros para os Formatos MARC 21 e as possibilidades trazidas pela XML, o objetivo geral desta pesquisa é 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 desse modelo a utilização de folhas de estilo de transformação elaboradas com a XSLT. Para esse objetivo, é realizada uma revisão de literatura sobre a conversão de registros, os Formatos MARC 21, a ISO 2709, a XML, a XSLT e o MARCXML. São destacados os principais aspectos da utilização da XML na codificação de documentos e as possibilidades oferecidas pela XSLT. Para discutir a relação entre os Formatos MARC 21 e a XML é apresentada a linguagem de marcação MARCXML. A partir do estudo teórico, é elaborado um modelo para a conversão de registros. São descritas as etapas e os componentes do modelo, assim como sua validação por meio da aplicação na conversão de registros bibliográficos exportados pelo sistema de gerenciamento de bibliotecas *Personal Home Library* (PHL). Como considerações finais, são destacadas as características do modelo e são sumarizadas algumas questões sobre sua utilização e sobre os instrumentos resultantes de sua aplicação na conversão de registros do PHL.

Palavras-chave: Conversão de registros. Formatos MARC 21. MARCXML. ISO 2709. *Extensible Markup Language* (XML). *Extensible Stylesheet Language for Transformation* (XSLT). *Personal Home Library* (PHL). Catalogação descritiva. Informação e Tecnologia.

Abstract

MARC 21 Formats are metadata standards internationally used for records interchange in bibliographic domain. Despite the importance of the MARC 21 Formats and other international metadata standards, there are integrated library systems that use their own metadata standard. The non-use of international metadata standards decreases the possibilities of records interchange and it might cause problems during data migration between systems. In face to the integrated library systems that do not use international metadata standards, our research question is the conversion of the records from these systems to MARC 21 records. In order to be read, MARC 21 records must be coded in such a way that allows informatics applications to recognize the record content. The technologies used for this codification have evolved from traditional ISO 2709 standard to Extensible Markup Language (XML). The use of XML in the codification gave us the possibility to use transformation stylesheets created with Extensible Stylesheet Language for Transformation (XSLT) for the records conversion. Considering the need of records conversion and the possibilities brought by XML, we aim to develop a model for conversion of records that were created with different metadata standards and are coded in XML to MARC 21 records using XSLT stylesheets. For this purpose, we made a literature review about record conversion, MARC 21 Formats, ISO 2709, XML, XSLT and MARCXML. We highlighted the main issues on the use of the XML for documents codification and the possibilities provided by XSLT. In order to discuss the relation between MARC 21 and XML, we presented MARCXML markup language. Starting from the literature review, we developed a model for record conversion and described its components and its steps, as well as its validation through the use in the conversion of bibliographic records from *Personal Home Library* (PHL) integrated library systems. In conclusion, we highlighted the characteristics of the model and we summarized some issues related to the use of the model and to the instruments that came from the application of the model in the PHL records conversion.

Keywords: Record conversion. MARC 21 Formats. MARCXML. ISO 2709. Extensible Markup Language (XML). Extensible Stylesheet Language for Transformation (XSLT). Personal Home Library (PHL). Descriptive cataloging. Information and Technology.

Lista de figuras

Figura 1 – Partes que compõem a catalogação.....	17
Figura 2 – Instrumentos do domínio bibliográfico.....	19
Figura 3 – Organização dos metadados do Formato MARC 21 para Dados Bibliográficos.....	24
Figura 4 – Exemplos de campos do Formato MARC 21 para Dados Bibliográficos	25
Figura 5 – Campo 260 do Formato MARC 21 para Dados Bibliográficos.....	25
Figura 6 – Registro no formato MARC 21 para Dados Bibliográficos.....	26
Figura 7 – Registro no Formato MARC 21 para Dados Bibliográficos codificado com a ISO 2709 ...	29
Figura 8 – Lista de contatos em um documento XML.....	38
Figura 9 – Transformação de documentos XML utilizando folhas de estilo XSLT	44
Figura 10 – Documento XML desejado	46
Figura 11 – Folha de estilo para a transformação do documento Agenda	46
Figura 12 – Fragmento de um registro no Formato MARC 21 para Dados Bibliográficos codificado com a DTD XML	52
Figura 13 – Marcação de registros nos Formatos MARC 21 com o MARCXML	53
Figura 14 – Registro no Formato MARC 21 para Dados Bibliográficos codificado com o MARCXML	54
Figura 15 – Transformação de registros nos Formatos MARC 21	57
Figura 16 – Página HTML criada a partir da transformação de um registro MARCXML com uma folha de estilo XSLT	58
Figura 17 – Componentes e possibilidades de codificação de registros nos Formatos MARC 21	60
Figura 18 – Cenários para a conversão de registros (1)	67
Figura 19 – Cenários para a conversão de registros (2)	69
Figura 20 – Modelo para a conversão de registros.....	71
Figura 21 – Exemplo de campos, etiquetas e definições do Formato PHL.....	82
Figura 22 – Registro bibliográfico no Formato PHL	83
Figura 23 – Marcação de registros no Formato PHL com o PHLXML	84
Figura 24 – Registro no Formato PHL codificado com o PHLXML.....	85
Figura 25 – Fragmento de um registro no Formato PHL de um recurso com dois autores	93
Figura 26 – Fragmento de um registro no Formato MARC 21 para Dados Bibliográficos de um recurso com dois autores.....	94
Figura 27 – Fragmento de um registro no Formato MARC 21 para Dados Bibliográficos obtido a partir da conversão de um registro no Formato PHL.....	94
Figura 28 – Código inicial da folha de estilo	97
Figura 29 – O <i>template bibliographicRecord</i> e algumas regras de transformação	97

Lista de quadros

Quadro 1 – Expressões-chave para a elaboração de regras de conversão	75
Quadro 2 – Mapa do Formato PHL e do Formato MARC 21 para Dados Bibliográficos	86
Quadro 3 – Mapa dos esquemas de codificação utilizados para representar idiomas	95

Lista de siglas

AACR2r	<i>Anglo-American Cataloguing Rules, 2nd ed., 2002 revision</i> (Código de Catalogação Anglo-Americano, 2. ed., revisão de 2002)
ABNT	Associação Brasileira de Normas Técnicas
ANSI	<i>American National Standards Institute</i> (Instituto Nacional Americano de Padrões)
ASCI	<i>American Standard Code for Information Interchange</i> (Código Padrão Americano para o Intercâmbio de Informação)
CALCO	Catálogo Legível por Computador
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CSS	<i>Cascading Style Sheets</i> (Folhas de Estilo em Cascata)
CSV	<i>Comma-Separated Values</i> (Valores Separados por Vírgula)
DC	Dublin Core
DTD	<i>Document Type Definitions</i> (Definição de Tipo de Documento)
FRAD	<i>Functional Requirements for Authority Data</i> (Requisitos Funcionais para Dados de Autoridade)
FRBR	<i>Functional Requirements for Bibliographic Records</i> (Requisitos Funcionais para Registros Bibliográficos)
FRSAD	<i>Functional Requirements for Subject Authority Data</i> (Requisitos Funcionais para Dados de Autoridade de Assunto)
HTML	<i>HyperText Markup Language</i> (Linguagem de Marcação de Hipertexto)
IFLA	<i>International Federation of Library Associations and Institutions</i> (Federação Internacional de Associações e Instituições de Bibliotecas)
ISBD	<i>International Standard Bibliographic Description</i> (Padrão Internacional de Descrição Bibliográfica)
ISO	<i>International Organization for Standardization</i> (Organização Internacional para Padronização)
LC	<i>Library of Congress</i> (Biblioteca do Congresso)
LILACS	Literatura Latino-americana e do Caribe em Ciências da Saúde
MADS	<i>Metadata Authority Description Schema</i> (Esquema de Metadados para a Descrição de Autoridade)
MARC	<i>MACHine-Readable Cataloging</i> (Catalogação Legível por Máquina)
MODS	<i>Metadata Object Description Schema</i> (Esquema de Metadados para a Descrição de Objeto)
OAI	<i>Open Archives Initiative</i> (Iniciativa dos Arquivos Abertos)
OCR	<i>Optical Character Recognition</i> (Reconhecimento Ótico de Caracteres)
PDF	<i>Portable Document Format</i> (Formato de Documento Portável)
PHL	<i>Personal Home Library</i>
PPGCI	Programa de Pós-Graduação em Ciência da Informação
RDA	<i>Resource Description and Access</i> (Descrição de Recursos e Acesso)

SGML	<i>Standard Generalized Markup Language</i> (Linguagem Padronizada de Marcação Genérica)
TDI	Tratamento Descritivo da Informação
TEI	<i>Text Encoding Initiative</i> (Iniciativa de Codificação de Texto)
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
UNIMARC	<i>Universal Machine Readable Cataloging</i> (MARC Universal)
USMARC	<i>United States MARC</i> (MARC dos Estados Unidos)
W3C	<i>World Wide Web Consortium</i> (Consórcio World Wide Web)
XML	<i>Extensible Markup Language</i> (Linguagem de Marcação Extensível)
XSL	<i>Extensible Stylesheet Language</i> (Linguagem Extensível para Folhas de Estilo)
XSL-FO	<i>Extensible Stylesheet Language – Formatting Objects</i> (Linguagem Extensível para Folhas de Estilo – Formatação de Objetos)
XSLT	<i>Extensible Stylesheet Language for Transformation</i> (Linguagem Extensível para Folhas de Estilo de Transformação)

Sumário

1	Introdução.....	10
1.1	Problema.....	11
1.2	Proposição.....	12
1.3	Objetivos.....	12
1.4	Justificativa.....	13
1.5	Metodologia.....	14
1.6	Estrutura do trabalho.....	15
2	Padrões de metadados do domínio bibliográfico: Formatos MARC 21.....	16
2.1	Formatos MARC 21: histórico e características.....	20
2.2	Codificação com a ISO 2709.....	27
2.3	Considerações e críticas aos Formatos MARC 21.....	30
3	Marcação e transformação: XML, XSLT e MARCXML.....	35
3.1	Marcação de documentos: XML.....	36
3.2	Transformação de documentos XML: folhas de estilo XSLT.....	41
3.3	Marcação e transformação de registros MARC 21: MARCXML.....	50
4	Modelo para a conversão de registros.....	63
4.1	Modelo para a conversão de registros.....	66
4.2	Mapeamento e regras de conversão.....	73
4.3	Elaboração da folha de estilo XSLT.....	76
4.4	Verificação da folha de estilo.....	78
4.5	Exportação dos registros e conversão para MARCXML.....	79
4.6	Conversão para ISO 2709.....	80
5	Aplicação do modelo na conversão de registros do Formato PHL.....	81
5.1	<i>Personal Home Library</i> (PHL).....	81
5.2	Mapeamento e regras de conversão.....	86
5.3	Elaboração da folha de estilo XSLT.....	96
5.4	Verificação da folha de estilo.....	99
5.5	Exportação e conversão dos registros.....	99
6	Considerações finais.....	102
	Referências.....	107
	APÊNDICE A – Manual para a conversão dos registros bibliográficos do PHL.....	115

1 Introdução

Esta pesquisa, intitulada “**Conversão de registros em XML para MARC 21: um modelo baseado em folhas de estilo XSLT**”, foi desenvolvida com o financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) na linha de pesquisa Informação e Tecnologia do Programa de Pós-Graduação em Ciência da Informação (PPGCI) da UNESP.

A Ciência da Informação, segundo Le Coadic (2004, p. 25), “tem por objeto o estudo das propriedades gerais da informação (natureza, gênese, efeito), e a análise de seus processos de construção, comunicação e uso”. Para Borko (1968, p. 3, tradução nossa), a Ciência da Informação preocupa-se “com o corpo de conhecimento relacionado à origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação”.

Uma vez que suas preocupações podem ser de naturezas diversas, a Ciência da Informação busca aporte em campos com especificidades que auxiliem 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, juntamente com as estruturas para a codificação, permitiram que os dados resultantes das atividades de catalogação fossem processados e intercambiados entre aplicações de

informática.¹ Atualmente, uns dos 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 2709 da *International Organization for Standardization* (ISO), 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 a codificação tecnologicamente mais vantajosas e mais condizentes com a atualidade, como é caso da *Extensible Markup Language* (XML) que, presente na linguagem de marcação *MARC 21 XML Schema* (MARXML), tem a sua disposição as possibilidades de transformação das folhas de estilo elaboradas com a *Extensible Stylesheet Language for Transformation* (XSLT).

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 como, por exemplo, os sistemas *Personal Home Library* (PHL), OpenBiblio, MiniBiblio, Biblioteca Fácil, Biblioscape e Autec Biblioteca.

Alguns desses sistemas, entre eles o PHL, utilizam padrões de metadados próprios, o que interfere na possibilidade de intercâmbio de registros e pode ocasionar problemas durante a migração entre sistemas de gerenciamento de bibliotecas.

Com base nesse cenário são traçados o problema, a proposição, os objetivos, a justificativa e a metodologia desta pesquisa, descritos nas seções seguintes.

1.1 Problema

Os Formatos MARC 21 são padrões de metadados internacionalmente utilizados para o intercâmbio de registros no domínio bibliográfico. Apesar da importância desses e de outros padrões de metadados internacionalmente aceitos, existem sistemas de gerenciamento de bibliotecas que utilizam padrões de metadados próprios. A não utilização de padrões de metadados internacionalmente aceitos 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

¹ O termo aplicação de informática é utilizado nesta pesquisa para designar qualquer programa de computador. Os termos sistema de gerenciamento de bibliotecas, processador XML, processador de transformação e analisador sintático, também utilizados nesta pesquisa, se referem a tipos de aplicações de informática.

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

1.2 Proposição

Existem sistemas de gerenciamento de bibliotecas que não utilizam padrões de metadados internacionalmente aceitos, por exemplo, os Formatos MARC 21. Ao invés disso, utilizam padrões de metadados próprios, às vezes baseados nas tabelas de seus bancos de dados. Nesses casos, as diferenças entre os bancos de dados dos sistemas de gerenciamento de bibliotecas inviabilizam a criação de uma aplicação de informática que possa ser utilizada na conversão dos registros exportados por 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).

A codificação de documentos utilizando a *Extensible Markup Language* (XML) trouxe benefícios como a possibilidade de uso de folhas de estilo elaboradas com a *Extensible Stylesheet Language for Transformation* (XSLT), que, de modo mais acessível que as linguagens de programação, permitem a transformação de documentos de modo a atender distintos propósitos.

Considerando a provável efemeridade das aplicações de informática específicas para a conversão dos registros de cada sistema de gerenciamento de bibliotecas e a acessibilidade das folhas de estilo XSLT, esta pesquisa tem por proposição a elaboração de um modelo para a conversão de registros de distintos padrões de metadados codificados com a XML em registros nos Formatos MARC 21, de modo a favorecer o intercâmbio de registros e a minimizar os conflitos durante a migração entre sistemas de gerenciamento de bibliotecas.

1.3 Objetivos

O objetivo geral desta pesquisa é 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 de seus pontos centrais a utilização de folhas de estilo de transformação elaboradas com *Extensible Stylesheet Language for Transformation* (XSLT).

Os objetivos específicos são:

- realizar estudo teórico sobre a conversão de registros, os Formatos MARC 21, a codificação com a ISO 2709, a XML, a XSLT e o MARCXML;
- destacar os principais aspectos da utilização da XML na codificação de documentos e as possibilidades oferecidas pela XSLT na transformação de documentos XML;
- elaborar um modelo para a conversão de registros, definindo suas etapas e seus componentes;
- validar o modelo na conversão de registros bibliográficos exportados pelo sistema de gerenciamento de bibliotecas *Personal Home Library* (PHL).

1.4 Justificativa

Em âmbito científico, esta pesquisa provê contribuições à Ciência da Informação por fomentar as discussões sobre a catalogação, sobre os instrumentos de representação do domínio bibliográfico e sobre a conversão de registros no contexto dos ambientes informacionais digitais possibilitados pelas tecnologias de informática. Por abordar a conversão de registros, esta pesquisa contribui também nas discussões sobre a interoperabilidade no âmbito dos sistemas de gerenciamento de bibliotecas.

Em âmbito social e profissional, os resultados desta pesquisa subsidiarão os profissionais da Ciência da Informação e da Ciência da Computação envolvidos nos processos de conversão de registros, favorecendo o intercâmbio de registros, por exemplo, em programas de catalogação cooperativa, e minimizando os conflitos durante a migração entre sistemas de gerenciamento de bibliotecas. A aplicação do modelo na conversão dos registros bibliográficos exportados pelo PHL, além de possibilitar a validação do modelo, apresenta uma solução para as instituições que utilizam esse sistema e que desejam converter seus registros para o Formato MARC 21 para Dados Bibliográficos.

Por fim, cabe destacar como justificativa que o modelo proposto tem como seu ponto forte o delineamento da interação e da integração de profissionais da Ciência da Informação e da Ciência da Computação, cada qual atuando em sua área de competência de forma interdisciplinar na tarefa de conversão de registros.

Entre os fatores motivadores do desenvolvimento desta pesquisa, estão: (1) a observação de situações em que a proferida “impossibilidade” de converter registros ocasionou retrabalhos; (2) a percepção de que a falta de conhecimentos básicos sobre a XML e sobre as tecnologias relacionadas à ela desfavorece o profissional da Ciência da Informação, principalmente o catalogador, no diálogo com profissionais da Ciência da Computação e,

acima de tudo, no uso estratégico das tecnologias de informática em prol das atividades de catalogação; (3) e o interesse pessoal pela XML e pelas tecnologias relacionadas a ela.

1.5 Metodologia

Esta pesquisa caracteriza-se como qualitativa em relação à forma de sua abordagem. Quanto aos seus fins, é descritiva e exploratória. É descritiva porque tem por objetivo “a identificação, registro e análise das características, fatores ou variáveis que se relacionam com o fenômeno ou processo” (OLIVEIRA NETTO, 2008, p. 29) e exploratória porque “objetiva a descoberta, o achado, a elucidação de fenômenos ou a explicação daqueles que não eram aceitos apesar de evidentes” (OLIVEIRA NETTO, 2008, p. 29).

Em razão de seu objetivo geral (elaborar um modelo para a conversão de registros), esta é uma pesquisa metodológica. Segundo Demo (1995, p. 13), a pesquisa metodológica é aquela “dedicada a indagar por instrumentos, por caminhos, por modos de se fazer ciência, ou a produzir técnicas de tratamento da realidade, ou a discutir abordagens teórico-práticas”.

Para o primeiro e segundo objetivos específicos, utilizou-se da pesquisa bibliográfica, que “tem como princípio básico conhecer as diferentes formas de contribuição científica que se realizaram sobre determinado assunto ou fenômeno” (OLIVEIRA NETTO, 2008, p. 30) ou, segundo Macedo (1996, p. 13), procura “identificar, localizar e obter documentos pertinentes ao estudo de um tema bem delimitado, levantando-se a bibliografia básica”.

Para o terceiro objetivo específico, a elaboração do modelo e a definição de suas etapas e componentes, foi utilizada a pesquisa metodológica.

Para o quarto objetivo específico, foram utilizadas a versão 8.2 do PHL, o Manual do PHL 8.2, revisão de 27 de setembro de 2011 (OLIVEIRA, 2011), e a documentação oficial do Formato MARC 21 para Dados Bibliográficos, provida pela *Library of Congress* (2012). Para a criação da folha de estilo XSLT foi utilizado o *Oxygen XML Editor*² e, para a transformação dos registros com a folha de estilo, foi utilizado o *MarcEdit*³.

A escolha pela aplicação do modelo na conversão dos registros bibliográficos exportados pelo PHL ocorreu em razão do número de instituições usuárias desse sistema. Dados de 18 de fevereiro de 2013 apontam a existência de 3176 instituições usuárias do PHL, divididas entre aquelas que disponibilizam o sistema em rede (intranet ou Web) e aquelas que fazem uso no modo monousuário (NOSSOS CLIENTES, 2013). O *PHL.netopac*, projeto que

² Disponível em: <<http://www.oxygenxml.com>>. Acesso em: 11 ago. 2013.

³ Disponível em: <<http://marcedit.reeset.net>>. Acesso em: 11 ago. 2013.

possibilita a busca no catálogo de 345 instituições que utilizam o PHL na Web, possui 5.329.165 registros bibliográficos (dados de 18 de fevereiro de 2013) (PHL.NETOPAC, 2013).

1.6 Estrutura do trabalho

Este capítulo reservou-se à apresentação do problema de pesquisa, da proposição, dos objetivos, da justificativa e da metodologia. Os demais capítulos deste trabalho estão arranjados como descrito a seguir.

O Capítulo 2, **Padrões de metadados do domínio bibliográfico: Formatos MARC 21**, contextualiza a catalogação descritiva e os padrões de metadados do domínio bibliográfico. Apresenta um breve histórico dos Formatos MARC, as características do Formato MARC 21 para Dados Bibliográficos e a codificação de registros utilizando a norma ISO 2709. Por fim, destaca considerações e críticas sobre os Formatos MARC 21 encontradas na literatura.

O Capítulo 3, **Marcação e transformação: XML, XSLT e MARCXML**, descreve a XML como uma forma de marcar documentos de modo estruturado, permitindo o posterior processamento de seu conteúdo. Aborda as possibilidades para a transformação de documentos XML utilizando folhas de estilo XSLT e finaliza com um resgate histórico sobre os principais esforços para a utilização da XML na marcação (codificação) de registros nos Formatos MARC 21.

O Capítulo 4, **Modelo para a conversão de registros**, inicia com uma revisão de literatura sobre a conversão de registros no domínio bibliográfico. Em seguida é apresentado um modelo para a conversão de registros de distintos padrões de metadados codificados com a XML em registros nos Formatos MARC 21. Os processos compreendidos pelo modelo, assim como seus componentes, são descritos nesse capítulo.

O Capítulo 5, **Aplicação do modelo na conversão de registros do Formato PHL**, descreve a aplicação do modelo proposto na conversão dos registros bibliográficos exportados pelo sistema de gerenciamento de bibliotecas PHL.

O Capítulo 6, Considerações finais, traz considerações sobre a conversão de registros no domínio bibliográfico, sobre o modelo apresentado no Capítulo 4 e sobre sua aplicação na conversão dos registros bibliográficos exportados pelo PHL, descrita no Capítulo 5.

2 Padrões de metadados do domínio bibliográfico: Formatos MARC 21

Os componentes-chave do modelo para a conversão de registros elaborado nesta pesquisa são os padrões de metadados, tanto o padrão utilizado nos registros a serem convertidos (padrão de metadados de origem) quanto o padrão para o qual os registros serão convertidos (padrão de metadados de destino). Os Formatos MARC 21 têm se tornado padrões de metadados internacionalmente utilizados no domínio bibliográfico, assim, no modelo para a conversão de registros, o padrão de metadados de destino poderá ser qualquer um dos Formatos MARC 21. Essa família de padrões de metadados, seu surgimento, características e tradicional codificação são os objetos de estudo deste capítulo.

Inerente aos mais distintos sistemas de informação, a atividade de representação da informação ou de representação de recursos informacionais tem sido realizada e estudada em diversos domínios, entre eles o domínio bibliográfico, atualmente abarcado pela Ciência da Informação.

Para Alves e Santos (2009), a construção de representações padronizadas dos recursos informacionais sempre foi uma preocupação da Ciência da Informação e, em especial, da Biblioteconomia. As autoras observam também que, no decorrer da história dessas áreas, foram desenvolvidas e aperfeiçoadas metodologias, técnicas e ferramentas para a construção e a padronização das representações dos recursos informacionais.

No domínio bibliográfico, a representação ou a descrição de recursos informacionais tem sido frequentemente pautada na atividade de catalogação, esta orientada pelas especificidades das disciplinas de catalogação descritiva e de catalogação de assunto, oriundas da prática biblioteconômica.

Segundo Mey (1995, p. 5), a catalogação é “o estudo, preparação e organização de mensagens codificadas, com base em itens existentes ou passíveis de inclusão em um ou vários acervos, de forma a permitir a interseção entre as mensagens contidas nos itens e as mensagens internas dos usuários”.

Santos e Corrêa (2009) destacam que a catalogação lida com representações sintéticas e codificadas de recursos informacionais de modo a torná-los únicos e a possibilitar sua recuperação e sua utilização.

Para Alves (2010, p. 12) a catalogação é entendida como

um processo de representação informacional que garante a identificação única do recurso informacional para fins de recuperação e, portanto, essencial para proporcionar a recuperação eficiente dos recursos nos variados tipos de sistemas de informação. Desenvolvida na área da

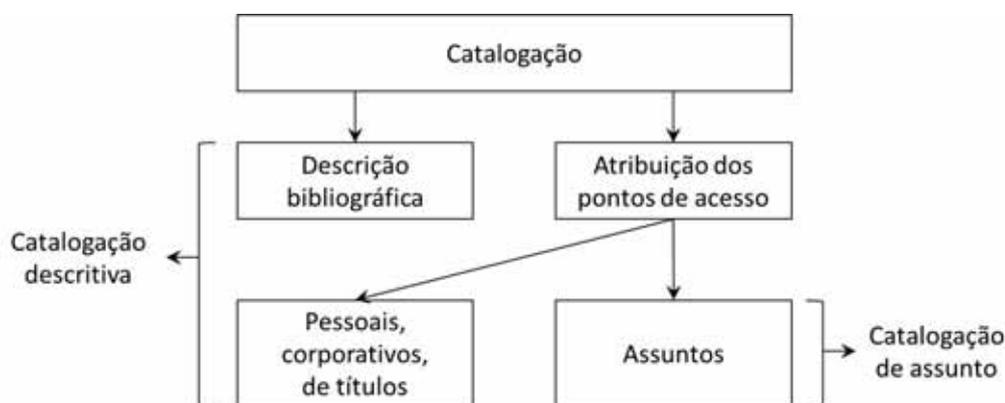
Biblioteconomia, a catalogação é um método para construção de representações do qual a Ciência da Informação se utiliza de forma interdisciplinar.

Garrido Arilla (1996, p. 25, tradução nossa) concebe a catalogação como

um processo unitário que, partindo da descrição do documento (descrição bibliográfica), passa pela escolha dos pontos de acesso (pessoais, corporativos, de títulos, temáticos e sistemáticos) e se encerra, em nível local, com a formação e a transcrição dos dados locais (atribuição e registro).

A autora destaca também que a descrição bibliográfica, juntamente com o estabelecimento dos pontos de acesso pessoais, corporativos e de título, é chamada de catalogação descritiva (GARRIDO ARILLA, 1996, p. 26), tal como apresentado na Figura 1.

Figura 1 – Partes que compõem a catalogação



Fonte: Garrido Arilla (1996, p. 26, tradução nossa).

Para referir-se à catalogação descritiva, Alves (2010, p. 13) tem utilizado o termo Tratamento Descritivo da Informação (TDI), que “designa o processo de representação informacional do recurso, ou seja, individualização e caracterização de um item documentário, diferenciando-o do tratamento temático também realizado na área de Biblioteconomia”.

A catalogação descritiva desenvolveu-se quase que exclusivamente sobre seus próprios aspectos teóricos e práticos e tomou para si também a responsabilidade pelo intercâmbio, tanto no ambiente analógico quanto no digital, das representações dos recursos informacionais, representações estas contidas nos registros bibliográficos e complementadas pelos registros de autoridade e pelos registros de itens.

Visando a possibilitar o intercâmbio dessas representações, diversas tecnologias têm sido utilizadas desde a antiguidade, de modo que os vários instrumentos da catalogação descritiva, construídos sobre princípios, modelos e aspectos teóricos e práticos, estão apoiados também em tecnologias. Entre essas tecnologias, destacam-se a ficha catalográfica,

padronizada no início do século XX, e, a partir da década de 1960, a fita magnética e as tecnologias de informática (BARBOSA, 1978, p. 72).

Entende-se que um dos pilares que serviram de base para o desenvolvimento da catalogação até os dias de hoje foi a proposta de que um recurso informacional fosse catalogado uma única vez e que todos os esforços fossem realizados no sentido de evitar que uma instituição precisasse catalogar um recurso já catalogado por outra instituição (BALBY, 1995, p. 30). Essa proposta foi mantida e, em parte, realizada pelos programas de catalogação na publicação, de catalogação centralizada e de catalogação cooperativa. No entanto, o estabelecimento de qualquer um desses programas, ou de qualquer atividade de intercâmbio dos registros provenientes dos processos da catalogação, trouxe a necessidade de padronizar as formas de representação oriundas de tais processos, o que culminou no estabelecimento de regras de catalogação (SANTOS; CORRÊA, 2009, p. 16), utilizadas em conjunto com vocabulários e convenções.

As **regras de catalogação** têm evoluído desde pequenos conjuntos de regras para a listagem de livros até complexos conjuntos de regras que, projetados e arranjados com base nos mais recentes modelos conceituais, se propõem a prover diretrizes e instruções para o registro dos atributos e dos relacionamentos dos mais variados tipos de recursos informacionais (RESOURCE..., 2013). Os **vocabulários**, enquanto instrumentos dos processos de catalogação, são entendidos não apenas como os conjuntos de pontos de acesso destinados a representar conceitos durante a catalogação de assunto, mas sim como quaisquer conjuntos de termos padronizados destinados a representar dados das mais diversas naturezas, tais como nomes de países, idiomas, tipos de conteúdo, de suporte, etc. As **convenções**, por sua vez, têm sido expressas em políticas e em manuais de catalogação, podendo ter abrangência tanto local quanto internacional.

A utilização das tecnologias de informática a partir da década de 1960 trouxe consigo os ambientes digitais e, assim, a necessidade de estruturas que tornassem as representações provenientes da catalogação, até então suficientemente padronizadas para o ambiente analógico, processáveis nos ambientes digitais. A partir do desenvolvimento dessas estruturas surgiram os formatos para o intercâmbio de dados catalográficos (BALBY, 1995, p. 30), nos últimos anos considerados padrões de metadados (ALVES, 2010), e as codificações.

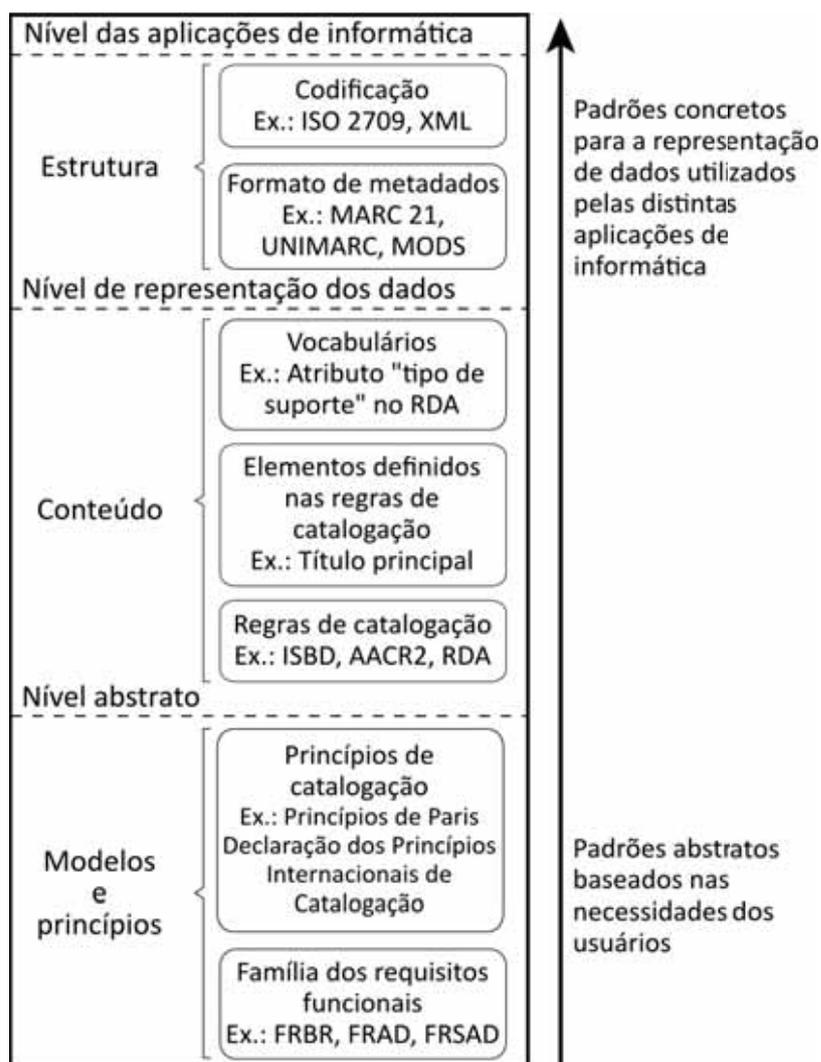
Os metadados são elementos descritivos ou atributos que representam características próprias ou atribuídas a um recurso, ao passo que os **padrões de metadados** são conjuntos predeterminados de metadados, metodologicamente construídos e padronizados (ALVES,

2010, p. 47-48). As **codificações**, nos ambientes digitais, permitem que os metadados e os valores a eles atribuídos sejam processados por aplicações de informática.

Em síntese, as codificações permitem que os dados indicados pelos metadados de um padrão e provenientes do uso de regras de catalogação, de vocabulários e de convenções sejam processados por aplicações de informática.

Os elementos destacados nesta seção – regras de catalogação, vocabulários, convenções, padrões de metadados e codificações – fazem parte do cenário sobre o qual é elaborado o modelo para a conversão de registros, objetivo central desta pesquisa. Esses e outros instrumentos encontrados no domínio bibliográfico, tais como os princípios e modelos conceituais, podem ser categorizados de diversas formas (ALVES, 2010; PICCO; ORTIZ REPISO, 2012). Entre as categorizações existentes, é utilizada aqui a categorização apresentada na Figura 2, desenvolvida por Picco e Ortiz Repiso (2012, p. 149).

Figura 2 – 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 princípios do domínio bibliográfico; um nível de representação dos dados, que compreende 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.

Cabe notar que, conforme a Figura 2, tanto as codificações quanto os padrões de metadados podem ser entendidos como estruturas. As codificações são estruturas que permitem o processamento dos registros por aplicações de informática. Os padrões de metadados são estruturas que permitem a identificação dos dados nos registros durante seu processamento. Assim, dadas essas possibilidades de interpretação da palavra estrutura, nesta pesquisa serão utilizados, sempre que possível, os termos codificação e padrão de metadados, ao invés de estrutura.

Além dos instrumentos elencados pelas autoras como relacionados à representação dos dados, são consideradas aqui também as convenções, anteriormente mencionadas, uma vez que, apesar de estarem intrinsicamente relacionadas às necessidades locais de cada instituição, devem ser consideradas na conversão de registros.

A adoção dessa categorização justifica-se por ela favorecer uma visualização clara dos instrumentos do domínio bibliográfico relacionados ao modelo para a conversão de registros proposto nesta pesquisa. Essa categorização, no delineamento do nível das aplicações de informática, explicita também a relação que o domínio bibliográfico tem mantido com a Ciência da Computação, relação esta que, como já mencionado, tem sido uma característica da Ciência da Informação.

Partindo dos conceitos até então apresentados, da categorização de Picco e Ortiz Repiso (2012) e servindo de base para a elaboração do modelo para a conversão de registros, as seções seguintes abordam os padrões de metadados Formatos MARC 21, seu histórico e características (seção 2.1), a tradicional codificação de registros com a norma ISO 2709 (seção 2.2) e algumas das considerações sobre e críticas aos Formatos MARC 21 (seção 2.3).

2.1 Formatos MARC 21: histórico e características

No domínio bibliográfico, as questões sobre a representação de recursos informacionais no ambiente digital estavam em debate já na década de 1960, mesmo que tal

ambiente fosse pensado apenas como ferramenta para potencializar as atividades dos ambientes analógicos. Nesse contexto, Barbosa (1978, p. 196) destaca que

A necessidade de prover serviços em maior profundidade e de forma mais rápida a um maior número de usuários, bem como o aumento quantitativo dos materiais tradicionais, acrescentado ao aparecimento de novas formas de materiais, levaram as bibliotecas dos países desenvolvidos a optar pelo uso de computadores para processamento de suas operações internas. Por meio de processos simplesmente manuais tornava-se impossível garantir o tratamento técnico atualizado das coleções e o atendimento, em tempo hábil, aos usuários. Exigia-se portanto, um melhor nível de serviços.

Nos Estados Unidos, em 1964, o *Council on Library Resources* financiou um estudo dos possíveis métodos para a conversão de fichas catalográficas da *Library of Congress* (LC) para uma forma legível por máquina (*machine-readable*) com o objetivo de utilizar computadores na impressão de bibliografias. O estudo deu origem a um relatório que levou a realização da *First Conference on Machine-Readable Catalog Copy* em janeiro de 1965 (AVRAM, 1968, p. 3).

Entre os tópicos de discussão dessa conferência estavam o formato para registros catalográficos legíveis por máquina, os elementos bibliográficos compreendidos por esse formato, a distribuição de dados catalográficos legíveis por máquina e o futuro dos catálogos em fichas e impressos (AVRAM, 1968, p. 3). Entre as conclusões da conferência estavam:

3. A *Library of Congress* deve provavelmente incluir em seu registro legível por máquina todos os dados impressos na ficha catalográfica e informações adicionais. A maioria dos conferencistas foi favorável à codificação de tantos dados quanto possível para assegurar uma máxima recuperação no futuro. Parece ser desejável para a *Library of Congress* ir adiante com suas próprias necessidades e que outras bibliotecas utilizem o que quiserem de um registro legível por máquina da LC.

4. O registro legível por máquina seria utilizado para uma variedade de produtos bibliográficos, tais como catálogos em fichas, catálogos impressos, bibliografias, listas de aquisição, etc. (AVRAM, 1968, p. 3, tradução nossa).

A discussão desses tópicos, a realização de análises dos dados catalográficos na LC e a revisão do formato proposto conduziram a uma segunda conferência em novembro de 1965. Em fevereiro de 1966 foi realizada a terceira conferência, marcando o início oficial do *MARC (MACHINE-Readable Cataloging) Pilot Project*. Em novembro do mesmo ano teve início o serviço semanal de distribuição de registros que, ao ser finalizado em junho de 1968, havia distribuído aproximadamente 50 mil registros em fitas magnéticas (AVRAM, 1968, p. 4-7).

Durante o projeto piloto foi utilizado o formato MARC I. Com base na experiência obtida nesse projeto, foi desenvolvido o MARC II, que passou a ser utilizado a partir de

março de 1969 na distribuição de registros de publicações monográficas em língua inglesa (AVRAM, 2003, p. 1714).

Tendo como base o MARC II, formatos com características semelhantes surgiram em diversos lugares, como o CAN/MARC (Canadá), o IBERMARC (Espanha), o MARCAL (América Latina) e o UKMARC (Reino Unido), além do UNIMARC, desenvolvido e mantido pela *International Federation of Library Associations and Institutions* (IFLA) (Federação Internacional de Associações e Instituições de Bibliotecas) (BARBOSA, 1978, p. 212; EITO BRUN, 2008, p. 147; MORENO, BRASCHER, 2007, p. 15), fazendo com que, na década de 1970, fossem encontradas mais de 50 variações do formato original (FLAMINO, 2006, p. 83).

No Brasil destaca-se a iniciativa da bibliotecária Alice Príncipe Barbosa que, em 1972, defendeu em seu mestrado o Formato Catalogação Legível por Computador (CALCO), baseado no Formato MARC II. Apesar da iniciativa, o CALCO não foi atualizado e caiu em desuso (SANTOS; CORRÊA, 2009, p. 49).

Na década de 1980 o MARC II tornou-se USMARC (LIBRARY OF CONGRESS, 2006). Na década seguinte surgiu a denominação MARC 21 a partir da harmonização entre os Formatos USMARC e CAN/MARC, que passaram a ser publicados sob o novo nome: MARC 21.

O MARC 21 não é um novo formato. De 1994 a 1997 as comunidades de usuário do USMARC e do CAN/MARC trabalharam para eliminar todas as diferenças remanescentes em seus dois formatos já similares. A compatibilidade tem sido uma característica do processo de desenvolvimento de ambos os formatos por muitos anos. Em 1997 e no início de 1998, atualizações dos formatos foram emitidas para tornar as especificações idênticas. MARC 21, a continuação do USMARC e do CAN/MARC, publica os formatos em uma edição sob um novo nome. (LIBRARY OF CONGRESS, 1998, tradução nossa).

Desde a harmonização entre os formatos, o MARC 21 tem sido mantido e atualizado, sendo que as atualizações mais recentes destinaram-se à inclusão e/ou à modificação de campos e subcampos para apoiar a utilização do *Resource Description and Access* (RDA) (Descrição de Recursos e Acesso), diretrizes e instruções desenvolvidas para substituir o *Anglo-American Cataloguing Rules, 2nd ed., 2002 revision* (AACR2r) (Código de Catalogação Anglo-Americano, 2. ed., revisão de 2002) (LIBRARY OF CONGRESS, 2012b).

Os Formatos MARC 21 têm sido descritos como padrões de metadados do domínio bibliográfico (ALVES, 2010). Enquanto padrões 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). Os cinco formatos que são coletivamente chamados de Formatos MARC 21 e os tipos de dados a que seus metadados se destinam são:

- *MARC 21 Format for Bibliographic Data* (Formato MARC 21 para Dados Bibliográficos): compreende os metadados necessários para descrever, recuperar e controlar diversos tipos de materiais bibliográficos: livros, publicações seriadas, arquivos de computador, mapas, músicas, matérias visuais e materiais mistos;
- *MARC 21 Format for Holdings Data* (Formato MARC 21 para Dados de Itens): compreende os metadados relacionados aos itens e a sua localização;
- *MARC 21 Format for Authority Data* (Formato MARC 21 para Dados de Autoridade): compreende os metadados que identificam ou controlam o conteúdo e a designação do conteúdo das partes de um registro bibliográfico que podem estar sujeitas ao controle de autoridade;
- *MARC 21 Format for Classification Data* (Formato MARC 21 para Dados de Classificação): compreende os metadados relacionados aos números de classificação e às rubricas associadas a eles, favorecendo a manutenção e o desenvolvimento de esquemas de classificação;
- *MARC 21 Format for Community Information* (Formato MARC 21 para Informação Comunitária): compreende os metadados sobre eventos, programas e serviços voltados à comunidade, etc. de modo que estes possam ser descritos e constar nos catálogos (LIBRARY OF CONGRESS, 1996).

Um registro MARC 21, em qualquer um dos cinco formatos, envolve três componentes: a estrutura do registro, a designação do conteúdo e o conteúdo (LIBRARY OF CONGRESS, 1996). Esses três componentes têm estado presentes desde o Formato MARC II:

A filosofia por detrás do MARC II foi o projeto de uma estrutura de formato (a representação física sobre um meio legível por máquina) capaz de conter informações bibliográficas de todos os tipos de materiais (livros, publicações seriadas, mapas, músicas, artigos de periódicos, etc.) e os registros relacionados (registros de remissivas de nomes e assuntos, etc.). A estrutura, ou “container vazio”, os designadores do conteúdo (etiquetas, indicadores e códigos de subcampo) usados para identificar explicitamente ou adicionalmente caracterizar os elementos de dados, e o conteúdo, os próprios dados (nomes dos autores, títulos, etc.), são os três componentes do formato. (AVRAM, 2003, p. 1713-1714, tradução nossa).

Seguindo a categorização de Picco e Ortiz Repiso (2012, p. 49), apresentada no início deste capítulo, 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, vocabulários e convenções.

O padrão de metadados Formato MARC 21 para Dados Bibliográficos, destinado a prover a designação do conteúdo para registros bibliográficos, será apresentado nesta seção. A codificação usualmente empregada com os registros criados com esse padrão de metadados será abordada na seção seguinte, enquanto que outras codificações alternativas serão apresentadas no próximo capítulo. As regras de catalogação, os vocabulários e as convenções que proveem o conteúdo de um registro bibliográfico não serão abordados nesta pesquisa de forma extensa.

Nos registros de qualquer um dos Formatos MARC 21, os metadados são representados pela designação do conteúdo, podendo esta ser de três tipos: etiqueta de campo (composta por três caracteres numéricos), indicador (composto por um espaço de caractere) ou código de subcampo (composto por um caractere alfanumérico). Essa designação do conteúdo é apresentada na documentação de cada padrão de metadados constituinte dos Formatos MARC 21.

Em razão da ampla utilização e estudo do Formato MARC 21 para Dados Bibliográficos, a designação de conteúdo utilizada para indicar os metadados compreendidos por tal padrão é brevemente exemplificada nesta pesquisa.

Os metadados do Formato MARC 21 para Dados Bibliográficos estão organizados em grupos, como mostrado na Figura 3.

Figura 3 – Organização dos metadados do Formato MARC 21 para Dados Bibliográficos

00X	Campos de controle
01X–09X	Campos de números e códigos
1XX	Campos do ponto de acesso principal
20X–24X	Campos de título e títulos relacionados
25X–28X	Campos de edição, imprensa, etc.
3XX	Campos de descrição física
4XX	Campos de indicação de série
5XX	Campos de nota
6XX	Campos dos pontos de acesso de assunto
70X–75X	Campos dos pontos de acesso secundários
76X–78X	Campos dos pontos de acesso de ligação
80X–83X	Campos dos pontos de acesso secundários de série
841–88X	Campos de itens, localização, gráficos alternativos, etc.

Fonte: Adaptado de <<http://loc.gov/marc/bibliographic>>. Acesso em: 24 out. 2012.

Cada grupo é composto por um conjunto de campos, como exemplificado na Figura 4. Cada campo, exceto os campos de controle, compreende dois indicadores e um número de subcampos que varia de campo para campo, como mostrado na Figura 5.

Figura 4 – Exemplos de campos do Formato MARC 21 para Dados Bibliográficos

25X-28X: Campos de edição, imprensa, etc.

250	Indicação de edição (NR)
254	Indicação de apresentação musical (NR)
255	Dados cartográficos matemáticos (R)
256	Características de arquivo de computador (NR)
257	País da entidade produtora (R)
258	Dados sobre material filatélico (R)
260	Publicação, distribuição, etc. (imprensa) (R)
263	Data planejada para a publicação (NR)
264	Produção, distribuição, manufatura e copyright (R)
270	Endereço (R)

Fonte: Adaptado de <<http://loc.gov/marc/bibliographic/bd25x28x.html>>. Acesso em: 24 out. 2012.

Figura 5 – Campo 260 do Formato MARC 21 para Dados Bibliográficos

260: Publicação, distribuição, etc. (imprensa) (R)

Primeiro indicador

Sequência das indicações de publicação

– Não aplicável/Nenhuma informação provida/Publicador mais antigo disponível
2 – Publicador interveniente
3 – Atual/último publicador

Segundo indicador

Indefinido

– Indefinido

Códigos dos subcampos

\$a – Local de publicação, distribuição, etc. (R)

\$b – Nome do publicador, distribuidor, etc. (R)

\$c – Data de publicação, distribuição, etc. (R)

\$e – Local de manufatura (R)

\$f – Manufaturador (R)

\$g – Data de manufatura (R)

\$3 – Materiais especificados (NR)

\$6 – Vínculo (NR)

\$8 – Vínculo de campo e sequência numérica (R)

Fonte: Adaptado de <<http://www.loc.gov/marc/bibliographic/bd260.html>>. Acesso em: 24 out. 2012.

As indicações (R) e (NR) presentes na Figura 4 e na Figura 5 indicam se o campo ou o subcampo é repetível ou não repetível, respectivamente. Se repetível, um campo pode ocorrer mais de uma vez em um registro; se não repetível, sua ocorrência está limitada a uma única vez. Se o subcampo é repetível, pode ocorrer mais de uma vez dentro daquele subcampo, se não repetível pode apresentar no máximo uma ocorrência em cada instância do campo ao qual pertence.

De modo geral, os indicadores são utilizados para prover dados adicionais sobre o conteúdo registrado em um determinado campo ou subcampo, por exemplo, indicar se uma nota é um resumo, uma resenha ou um alerta sobre o conteúdo do recurso, e indicar o vocabulário ao qual um ponto de acesso de assunto pertence. Apesar de estarem disponíveis em todos os campos, exceto nos campos de controle, nem sempre os indicadores são utilizados, nesses casos são declarados como “Indefinidos” e não são preenchidos, ou seja, os espaços destinados a eles são deixados em branco.

Além desses campos, indicadores e subcampos, cada registro contém um líder, elemento obrigatório composto por 24 caracteres que comporta metadados relacionados ao processamento do registro, tal como seu tamanho, o tipo do recurso descrito, o nível bibliográfico, o esquema de codificação dos caracteres do registro e a forma da catalogação descritiva.

A Figura 6 traz um registro no Formato MARC 21 para Dados Bibliográficos que faz uso de alguns dos metadados disponíveis em tal padrão.

Figura 6 – Registro no formato MARC 21 para Dados Bibliográficos

```

LDR 00822nam#a2200229#a#4500
005 20121108132540.0
008 121107s2007####bl#a###gr#####000#0#por#d
020 ## $a857642004X
082 0 4 $a001.42$222
100 1 # $aPescuma, Derna
245 1 0 $aProjeto de pesquisa :$bO que é? como fazer? : um guia para sua elaboração
    /$cDerna Pescuma, Antonio Paulo F. de Castilho ; colaboração de Paulo Angelo
    Lorandi. -
250 ## $a4. ed. -
260 ## $aSão Paulo :$bOlho d'Água,$c2007
300 ## $a96 p. :$bil. ;$c21 cm
490 1 # $aMétodo ;$v3
500 ## $aSegue as mais recentes normas da ABNT
650 0 4 $aRedação acadêmica
650 0 4 $aTrabalhos científicos$xMetodologia
650 0 4 $aRelatórios$xRedação
700 1 # $aCastilho, Antonio Paulo Ferreira de,$d1953-
700 1 # $aLorandi, Paulo Angelo
830 # 0 $aMétodo
  
```

Fonte: Elaborada pelo autor.

Legenda: cada # indica uma posição de caractere em branco ou um indicador indefinido (em branco); o \$ é utilizado para identificar os códigos dos subcampos.

O registro da Figura 6 foi elaborado com base na documentação oficial do Formato MARC 21 para Dados Bibliográficos (LIBRARY OF CONGRESS, 2012a) e no AACR2r, sendo o número de classificação (campo 082) e os pontos de acesso de assunto (campos 650) registrados a partir dos dados presentes na catalogação na fonte que acompanha o recurso representado pelo registro.

A forma de apresentação com que o registro é exibido na Figura 6 é convencionalmente utilizada para propósitos de leitura por agentes humanos e envolve a designação do conteúdo e o conteúdo do registro, não sendo utilizada com o propósito de possibilitar o processamento por uma aplicação de informática. Para que isso ocorra é necessário dotar o registro de uma estrutura que o torne processável por aplicações de informática, tal estrutura é a codificação do registro. A codificação tradicionalmente utilizada com os registros nos Formatos MARC 21 é especificada pela norma ISO 2709, abordada na seção seguinte.

2.2 Codificação com a ISO 2709

Como anteriormente mencionado, um dos componentes dos registros nos Formatos MARC 21 é a estrutura, provida pela codificação. Para Barbosa (1978, p. 209), essa “estrutura é o arranjo da informação bibliográfica em fita magnética, ou seja, é a moldura básica do formato. Contém certas informações de controle, necessárias à transmissão de informações e define o ‘layout’ dos campos de dados”.

Balby (1995, p. 31) entende a codificação 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 legíveis por máquina desenvolvidos pela LC na década de 1960 visavam à transmissão de registros em fitas magnéticas, surgiu assim o caráter sequencial da codificação utilizada originalmente e ainda atualmente na transmissão de registros MARC.

Essa codificação foi reconhecida nacionalmente nos Estados Unidos pelo *American National Standards Institute* (ANSI) (Instituto Nacional Americano de Padrões) em 1971 como a norma Z39.2, e internacionalmente em 1981 pela *International Organization for Standardization* (ISO) (Organização Internacional para Padronização) como a 2709, *Documentation – Format for bibliographic information interchange on magnetic tape* (Documentação – Formato para intercâmbio de informação bibliográfica em fita magnética).

No decorrer dos anos essas normas foram atualizadas, sendo de 1994 a última versão da ANSI Z39.2 e de 2008 a última versão da ISO 2709. Em razão das semelhanças entre as duas normas, esta pesquisa fará menções somente a norma ISO, devido ao seu caráter internacional e por ser a mais recente.

A ISO 2709 “descreve uma estrutura geral, um *framework* projetado especialmente para comunicações entre sistemas de processamento de dados e não para o uso como um formato de processamento dentro dos sistemas” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008, tradução nossa). Essa norma não define o tamanho ou o conteúdo de registros individuais, também não atribui qualquer significado às etiquetas, aos indicadores ou aos códigos de subcampos, uma vez que prover tais especificações é função dos formatos (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008). Assim, os elementos de um registro definidos por essa norma são:

- líder: compreende os 24 primeiros caracteres de um registro trazendo informações sobre o mesmo, tais como o tamanho do registro e o número de indicadores utilizado em cada campo;
- diretório: compreende um número variado de partes, cada uma indicando um campo, o tamanho do campo e a posição do caractere inicial do campo; o diretório termina com um caractere separador de campo;
- campos: compreendem todo o conteúdo dos campos presentes no registro (campo identificador do registro – 001, campos de referência – 002-009 e 00A-00Z, e campos de dados – 010-999 e 0AA-ZZZ), os indicadores e os identificadores (códigos dos subcampos); cada campo termina com um separador de campo; ao final do último campo é adicionado um separador de registros (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011).

O exemplo de um registro no Formato MARC 21 para Dados Bibliográficos codificado com a ISO 2709 é apresentado na Figura 7.

Figura 7 – Registro no Formato MARC 21 para Dados Bibliográficos codificado com a ISO 2709

```

00822nam a2200229 a 450000500170000000800410001702000150005808200150
00731000019000882450171001072500013002782600038002913000026003294900
01700355500004200372650002500414650004000439650002700479700004800506
700002600554830001200580RS20121108132540.0RS121107s2007 bl a gr
000 0 por dRS USa857642004XRS04USa001.42US222RS1 USaPescuma,
DernaRS10USaProjeto de pesquisa :USbo que é? como fazer? : um guia
para sua elaboração /UScDerna Pescuma, Antonio Paulo F. de Castilho
; colaboração de Paulo Angelo Lorandi. -RS USa4. ed. -RS USaSã
Paulo :USbOlho d'Água,USc2007RS USa96 p. :USbil. ;USc21 cmRS1
USaMétodo ;USv3RS USaSegue as mais recentes normas da
ABNTRS04USaRedação acadêmicaRS04USaTrabalhos
científicosUSxMetodologiaRS04USaRelatóriosUSxRedaçãoRS1 USaCastilho,
Antonio Paulo Ferreira de,USd1953-RS1 USaLorandi, Paulo AngeloRS
0USaMétodoRSGS

```

Fonte: Elaborado pelo autor.

O registro mostrado na Figura 7 constitui-se de uma linha contínua de caracteres, no entanto, é apresentado com quebras de linha com vistas a possibilitar uma melhor exibição. As indicações **RS**, **US** e **GS** representam os caracteres ASCII não imprimíveis⁴ utilizados para separar, respectivamente, campos, subcampos e registros.

Esse registro é intercambiado entre aplicações de informática e seu conteúdo é processado de acordo com os dados presentes no líder e no diretório e com os indicadores, as etiquetas dos campos e os códigos dos subcampos definidos no Formato MARC 21 para Dados Bibliográficos.

Apesar das normas para a codificação terem sido atualizadas 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 Z39.50⁵, de importação entre sistemas de gerenciamento de bibliotecas ou de armazenamento em bancos de dados.

⁴ No contexto do *American Standard Code for Information Interchange* (ASCII) (Código Padrão Americano para o Intercâmbio de Informação), esses caracteres significam, respectivamente, *record separator* (separador de registro), *unit separator* (separador de unidade) e *group separator* (separador de grupo). Nota-se que cada registro ISO 2709 é entendido como um grupo de dados, dentro do qual os campos são separados por separadores de registros (RS) e os subcampos por separadores de unidades (US). Mais informações em: <<http://pt.wikipedia.org/wiki/ASCII>>. Acesso em: 18 fev. 2013.

⁵ O protocolo Z39.50 é amplamente utilizado nos sistemas de gerenciamento de bibliotecas internacionais e é definido pelo norma ANSI/NISO *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*. A norma estabelece um modelo genérico para o processo de recuperação em bases de dados entre sistemas informatizados conectados em rede: o

2.3 Considerações e críticas aos Formatos MARC 21

García Melero e García Camarero (1999⁶ apud EITO BRUN, 2008, p. 148, tradução nossa) sintetizam os objetivos básicos que levaram à criação do MARC:

1. A criação de descrições bibliográficas reutilizáveis em distintos processos (aquisição, catalogação, recuperação e circulação ou empréstimo).
2. Normalizar a estrutura dos registros bibliográficos para facilitar seu intercâmbio entre bibliotecas.
3. Alcançar a independência dos dados bibliográficos frente aos sistemas de informação utilizados para seu tratamento automatizado.

Sobre a importância dos Formatos MARC, Moreno e Brascher (2007, p. 14) destacam que a necessidade de intercâmbio de informações de forma padronizada, o planejamento e a implantação da catalogação cooperativa para redução de custos e retrabalhos, ganhou impulso com tais formatos.

Alves (2010, p. 33) relaciona a ampla utilização dos formatos MARC com a consonância existente entre esses formatos e as regras de catalogação. Nas palavras da autora,

é importante destacar que o formato passou a ser amplamente utilizado pela comunidade biblioteconômica por refletir a lógica de descrição contemplada nas estruturas descritivas dos códigos de catalogação, possibilitando, dessa forma, uma facilidade na importação e exportação de dados bibliográficos. (ALVES, 2010, p. 33).

Alves (2010, p. 70) entende também que, apesar dos termos metadados e padrões de metadados não terem sido utilizados na época de seu desenvolvimento, o Formato MARC 21 para Dados Bibliográficos sem dúvida se constitui como um padrão de metadados do domínio bibliográfico.

O formato MARC 21 apresenta como escopo criar representações padronizadas dos recursos informacionais de uma unidade de informação, por meio de seu esquema de metadados e esquemas associados (princípios e códigos de catalogação), com o intuito de intercambiar metadados descritivos ou mais especificamente metadados bibliográficos. (ALVES, 2010, p. 70).

Para Zafalon (2012, p. 23), o Formato MARC 21 para Dados Bibliográficos “é um dos instrumentos indispensáveis ao catalogador que intenciona prover de meios automatizados de acesso aos, e compartilhamento de, registros bibliográficos na unidade de informação em que

sistema cliente envia consultas ao sistema servidor, que recupera os registros de uma ou mais bases de dados e os envia ao sistema cliente (EITO BRUN, 2008, p. 303).

⁶ GARCÍA MELERO, L. A.; GARCÍA CAMARERO, E. **Automatización de bibliotecas**. Madrid: ArcoLibros, 1999.

atua”. Para a autora, esse padrão ainda não é utilizado por grande parte das instituições nacionais, sendo que alguns dos motivos que levam à não utilização desse padrão são:

- o desconhecimento dos padrões ou das vantagens em adotá-los;
- as limitações oriundas de infra-estrutura tecnológica;
- as restrições orçamentárias para o tratamento dos dados, por requerer pessoal capacitado;
- a falta de habilidades de gestão para o gerenciamento de diferenciados recursos necessários ao processo de mudança ou implantação de sistemas automatizados;
- a inexperiência no processo de seleção de software de gerenciamento que atenda aos requisitos de compartilhamento de dados e uso do formato MARC21 no delineamento das bases de dados; e
- o desconhecimento dos processos necessários para o compartilhamento, a conversão e a migração de dados bibliográficos, apesar de se ter conhecimento da existência de padrões e das vantagens na adoção. (ZAFALON, 2012, p. 26-27).

Além das considerações sobre sua importância e sobre os benefícios de sua utilização, também são encontradas na literatura diversas críticas aos Formatos MARC 21. Um ponto destacado por diversos autores é que, em razão dos Formatos MARC terem sido criados na década de 1960 para possibilitar principalmente a produção de fichas catalográficas, um registro MARC 21, ainda hoje, simula uma ficha catalográfica no ambiente digital.

Segundo Coyle (2004, p. 166), o registro MARC foi criado como um reflexo digital das regras de catalogação de seu tempo, regras estas que não são tão diferentes das regras de catalogação atuais. Essas regras foram originalmente projetadas para a produção de fichas e ainda mantêm essa característica, o que é possível observar por meio dos pontos de acesso das entradas principais, da forma invertida dos nomes e do agrupamento de elementos em segmentos semelhantes a parágrafos.

Danskin (2004, p. 115) acrescenta à discussão outros aspectos. Para o autor, o registro MARC moderno é uma “reencarnação” da ficha catalográfica e tão ineficiente quanto esta, pois, toda vez que um catalogador cria um registro para uma nova edição de uma obra que a biblioteca já possui um exemplar, ele digita (ou copia) dados que já estão presentes na base de dados, mas não estão hierarquicamente estruturados de modo que possam ser reutilizados. Para Danskin, isso é ignorar a capacidade das bases de dados relacionais presentes na maioria dos modernos sistemas de gerenciamento de bibliotecas.

Mönch e Aalberg (2003, p. 405) descrevem algo semelhante. Os autores exemplificam que uma busca por uma obra literária que existe em numerosas edições e traduções está suscetível a resultar em um grande conjunto de resultados, pois cada edição ou tradução está

representada por um registro individual que não está relacionado aos outros registros que descrevem a mesma obra.

Segundo Gorman (1997, tradução nossa), “[...] o MARC tem suas origens na ficha catalográfica. Isso não é somente visto na ordem de seus campos, que preserva exatamente a ordem dos dados na ficha catalográfica, incluindo a separação do cabeçalho da ‘entrada principal’ (1XX) de outros pontos de acesso (7XX)”. Gorman (1997, tradução nossa) complementa que

Todos os problemas que temos com o MARC derivam da violação inicial da lei fundamental da automação de bibliotecas – “nunca automatize apenas o que você tem”. Anos atrás, clamei por uma reforma completa do MARC que deveria, essencialmente, substituir por simples registros (nomes, descrições, assuntos) com muitas e complexas conexões o que temos hoje – complexos registros com poucas conexões.⁷ É evidente que meu chamado não foi ouvido. Como uma consequência, estamos lidando com os efeitos de milhões de registros MARC e centenas de sistemas baseados nesses registros sem a capacidade de tirar vantagem da sofisticação dos modernos sistemas online.

Do ponto de vista da construção de catálogos digitais, Fusco (2011) destaca os problemas resultantes da modelagem dos bancos de dados utilizando os Formatos MARC 21 como base. Para o autor,

É importante destacar que o MARC21 é um formato para intercâmbio de dados, que informa como um registro bibliográfico e catalográfico deve estar descrito para que sua importação ou exportação ocorra com sucesso, porém o modelo de dados do catálogo não necessariamente deve estar organizado estruturalmente no mesmo formato de um registro MARC21. (FUSCO, 2011, p. 57).

Fusco expõe que o uso indevido dos Formatos MARC 21 resulta em anomalias nos bancos de dados:

Isso [anomalias no banco de dados] ocorre porque padrões de metadados como o formato MARC têm uma estrutura linear em suas estruturas de registros, ou seja, um modelo bidimensional de atributo-valor e determinam o formato de entrada e de intercâmbio de dados, sem determinar o conceito de domínio dos objetos representados com seus relacionamentos. (FUSCO, 2011, p. 119).

Miller e Clarke (2003) apontam alguns problemas com os Formatos MARC 21, em síntese:

⁷ A proposta de Gorman é descrita em: GORMAN, Michael. Authority files in a developed machine system (with particular reference to AACR II). In: FURUYA, Natsuko Y. (Ed.). **What’s in a name: control of catalogue records through automated authority files**. Toronto: University of Toronto Press, 1978. p. 179-202.

- inconsistência dos dados – um mesmo tipo de dado é registrado em diferentes campos/subcampos de diferentes formas (por exemplo, datas);
- redundância de dados – um mesmo dado é registrado em mais de um campo/subcampo, ora de modo codificado, ora literalmente (por exemplo, idiomas);
- mistura de dados e seus atributos;
- dispersão e irregularidade dos relacionamentos; e
- extrema complexidade na codificação.

Os autores acrescentam que,

Onde há informação hierárquica envolvida, a estrutura relativamente plana do MARC a torna difícil de ser refletida. As limitações do tamanho do campo variam de sistema para sistema. A granularidade injustificada existe, enquanto que a funcionalidade indicando a granularidade está faltando. (MILLER; CLARKE, 2003, p. 116, tradução nossa).

Miller e Clarke (2003, p. 116) finalizam que os problemas são tantos que o MARC deve ser completamente reavaliado e reconstruído. Afirmam também que o MARC efetivamente evita que as bibliotecas tirem total vantagem da XML e das tecnologias relacionadas e coloca as bibliotecas em desvantagem na arena competitiva do gerenciamento informacional. “Apesar de sua vida longa e de suas contribuições úteis, o MARC representa agora mais uma desvantagem que uma vantagem” (MILLER; CLARKE, 2003, p. 116, tradução nossa).

Com o objetivo de explorar os vários problemas que alguns autores têm associado aos Formatos MARC 21, Yee (2004) analisa e define quatro tipos de problemas:

- problemas que na verdade não são culpa do MARC 21, mas sim das regras e das práticas de catalogação que proveem o conteúdo dos registros MARC 21, por exemplo, o AACR2r;
- problemas que realmente não são problemas, ao invés disso são soluções para problemas imperfeitamente compreendidos pelos autores;
- problemas devido ao ambiente de catalogação compartilhada para o qual o MARC 21 foi projetado;
- problemas causados ou parcialmente causados pelo MARC 21 e que talvez possam ser solucionados nos processos de migração dos dados para um novo padrão de estrutura de dados no futuro.

Sobre os problemas do quarto tipo, Yee (2004, p. 166, tradução nossa) destaca que uma parte desses problemas poderia ser resolvida nos *software* e não no próprio Formato MARC 21.

Infelizmente, muitos dos problemas que são atribuídos ao MARC 21 são, na verdade, problemas que derivam do insucesso dos distribuidores de *software* em suportar as capacidades totais do MARC 21. Às vezes isso se deve às considerações de ordem financeira (o desenvolvimento é feito somente quando um número significativo de clientes será beneficiado); às vezes isso ocorre devido à falta de entendimento que os distribuidores de softwares têm do MARC 21, dos registros catalográficos, dos problemas que surgem em complexas bases de dados de registros bibliográficos e dos problemas de face pública no acesso aos catálogos online de acesso público. (YEE, 2004, p. 166, tradução nossa).

Apesar das críticas que têm recebido, os Formatos MARC 21 ocupam ainda o status de padrões de metadados consolidados para o intercâmbio de registros no domínio bibliográfico. Uma discussão exaustiva sobre a adequação ou não de tal padrão de metadados não cabe como um dos objetivos desta pesquisa.

Nascidos no início da automação das unidades de informação, os Formatos MARC e seus derivados se constituem como um dos resultados do uso estratégico das tecnologias em prol do processo de catalogação (PEREIRA; SANTOS, 1998, p. 124). Com o desenvolvimento da Ciência da Computação, uma gama de tecnologias de informática tem surgido, oferecendo à Ciência da Informação e à catalogação novas oportunidades para a melhoria de seus processos. Algumas dessas tecnologias e o modo com que elas foram, são e podem ser utilizadas a favor da catalogação são os objetos de discussão do capítulo seguinte.

3 Marcação e transformação: XML, XSLT e MARCXML

Os padrões de metadados de origem e de destino são elementos-chave do modelo para a conversão de registros elaborado nesta pesquisa. Os registros criados de acordo com esses padrões precisam, no entanto, de uma estrutura que os tornem processáveis por aplicações de informática. Essa estrutura é chamada de codificação. A tradicional codificação utilizada com os registros nos Formatos MARC 21, a ISO 2709, foi apresentada no capítulo anterior. Além dela, o modelo para a conversão de registros proposto nesta pesquisa faz uso da XML. A codificação de registros com a XML e a transformação de documento XML com folhas de estilo XSLT são os objetos de estudo deste capítulo.

O desenvolvimento da Ciência da Computação tem trazido uma gama de tecnologias de informática, oferecendo à Ciência da Informação novos ambientes e ferramentas para a discussão das questões relacionadas à origem, à coleção, à organização, ao armazenamento, à recuperação, à interpretação, à transmissão, à transformação e ao uso da informação. Nesse contexto, surgem novas possibilidades para a catalogação no que diz respeito à representação, ao uso e ao intercâmbio de dados em ambientes digitais.

Para Siqueira (2003, p. 33), “a trajetória histórica da catalogação mostra que esta, sempre utilizou as tecnologias vigentes em cada época em busca de métodos mais eficientes e econômicos para facilitar a representação, a recuperação e a disseminação da informação”.

Segundo Santos e Alves (2009), “a organização, a representação e a recuperação dos recursos informacionais sempre estiveram relacionadas com a tecnologia vigente em cada época e, atualmente, com as tecnologias de informática”. É nesse sentido que Balby (1995, p. 29) aponta que os próprios Formatos MARC 21, apresentados no capítulo anterior, “são um desenvolvimento da Biblioteconomia cuja origem está ligada à adoção de inovações tecnológicas no trabalho de catalogação”.

Alves (2010, p. 98) aponta que, em razão de sua atuação como agentes de mudança no tratamento descritivo de recursos informacionais,

as tecnologias de informática vêm contribuindo para uma reavaliação da teoria, dos princípios, dos fundamentos, dos métodos e dos instrumentos de representação do domínio bibliográfico. Essa reavaliação vem sendo desenvolvida com o objetivo de tornar mais consistente o processo de TDI em relação aos novos ambientes informacionais digitais e às novas necessidades dos usuários diante das tecnologias.

A *Extensible Markup Language* (XML) e a *Extensible Stylesheet Language for Transformation* (XSLT) são algumas das tecnologias de informática que podem contribuir

para uma reavaliação dos instrumentos de representação do domínio bibliográfico e oferecer métodos mais eficientes e econômicos para os processos de catalogação, facilitando, assim, a representação, a recuperação e a disseminação da informação. Tais tecnologias são apresentadas neste capítulo.

Primeiramente, a XML é descrita como um meio de marcar documentos de modo estruturado, permitindo o posterior processamento de seu conteúdo (seção 3.1). Em seguida, aborda-se a XSLT e as possibilidades para a transformação de documentos XML (seção 3.2). Por fim, é realizado um resgate histórico sobre os principais esforços empreendidos para a utilização da XML na codificação de registros nos Formatos MARC 21 e sobre algumas possibilidades de uso da XSLT na conversão de tais registros (seção 3.3).

3.1 Marcação de documentos: XML

Em 1986 a ISO publicou a norma 8879 *Information processing – Text and office systems – Standard Generalized Markup Language (SGML)* (Linguagem Padronizada de Marcação Genérica), provendo uma linguagem – a SGML – para facilitar a produção e o compartilhamento de documentos eletrônicos (MILLER; CLARKE, 2004, p. 2) ou, segundo Moreno e Brascher (2007, p. 15), “uma linguagem para descrever a estrutura dos documentos, ou qualquer tipo de dado textual, deixando a interpretação dos dados para outros programas”.

Com base na SGML, em 1998, o *World Wide Web Consortium (W3C)* (Consórcio World Wide Web) lançou a *Extensible Markup Language (XML)* (Linguagem de Marcação Extensível). Como afirma o próprio W3C (2012b, tradução nossa), “originalmente projetada para vencer os desafios da publicação eletrônica em larga escala, a XML está desempenhando um papel de crescente importância na troca de uma ampla variedade de dados na Web e em outros lugares”.

Segundo Ray (2001, p. 2),

Em um nível, a XML é um protocolo para conter e gerenciar informações. Em outro, é uma família de tecnologias que pode fazer de tudo, desde formatar documentos até filtrar dados. E, no nível mais alto, é uma filosofia para o tratamento de informações, que busca o máximo de utilidade e flexibilidade para os dados, definindo-os à sua forma mais pura e mais estruturada.

O autor destaca também que a XML “é um conjunto de regras para a criação de linguagens de marcação” (RAY, 2001, p. 2). Segundo Almeida (2002, p. 6) a palavra marcação tem sido utilizada “para descrever anotações ou marcas em um texto, que tem por objetivo dar instruções ao desenhista ou datilógrafo sobre a maneira como uma parte do texto

deveria ser representada”. Ray (2001, p. 2) apresenta aspectos que complementam a visão de Almeida. Para o autor, a marcação é entendida como “a informação incluída em um documento para melhorar seu significado de certas maneiras, por identificar as partes e como elas se relacionam umas com as outras” (RAY, 2001, p. 2).

Uma linguagem de marcação, por sua vez, é “um conjunto de símbolos que pode ser colocado no texto de um documento para demarcar e rotular as partes desse documento” (RAY, 2001, p. 2) ou um conjunto de convenções utilizadas para a codificação de textos que especificam quais marcas são permitidas, quais são exigidas, como se deve fazer distinção entre as marcas e o texto e qual o significado da marcação (ALMEIDA, 2002, p. 6).

No contexto da marcação de documentos digitais, Almeida (2002, p. 6) aponta que “Codificar ou ‘marcar’ um texto para processamento por computadores é também um processo de tornar explícito o que é conjetural”, indicando como seu conteúdo deve ser interpretado. Para Ray (2001, p. 2),

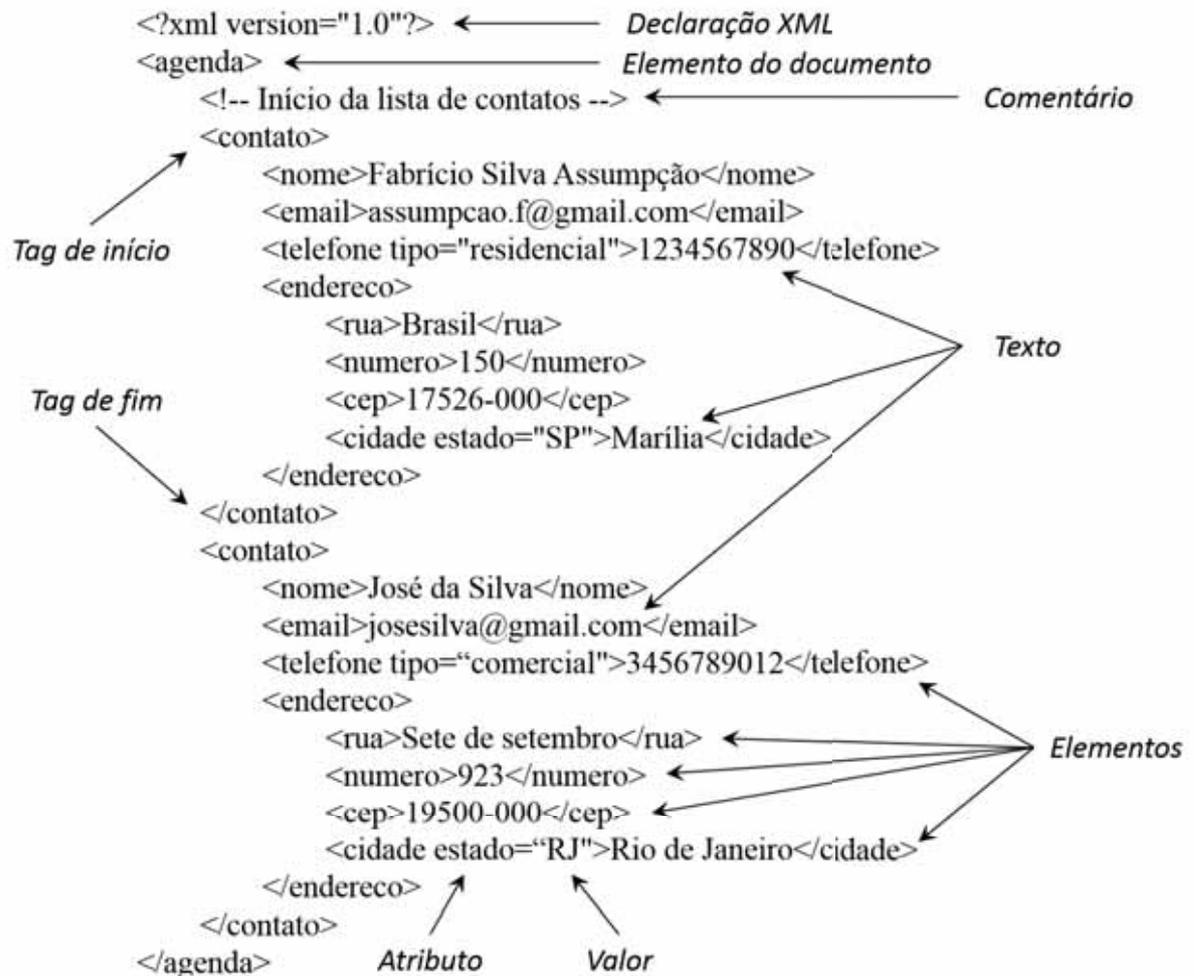
A marcação é importante para os documentos eletrônicos porque eles são processados por programas de computador. Se um documento não tiver rótulos ou limites, então um programa não saberá como tratar uma parte do texto para distingui-la de outra parte.

A XML não especifica qualquer símbolo ou marca, melhor dizendo, elemento, para a marcação de documentos. O que ela especifica são regras ou restrições sintáticas para a marcação de documentos, e, com base nessas regras é possível criar linguagens e utilizá-las para marcar documentos. As principais regras da XML são:

- um elemento que não seja vazio deve conter *tags* (etiquetas) de início e de fim;
- a *tag* de um elemento vazio precisa ter uma barra antes do sinal de maior;
- o valor de cada atributo deve estar entre aspas;
- os elementos não podem se sobrepor;
- caracteres de marcação isolados não podem aparecer no conteúdo analisado (aqui são incluídos os sinais <, > e &);
- os nomes dos elementos podem começar apenas com letras e com o caractere de sublinhado, e podem conter apenas letras, números, hifens, pontos e sublinhados. Os sinais de dois pontos são reservados para indicar *namespaces* (RAY, 2001, p. 60).

Quando um documento XML satisfaz essas regras, como é o caso do documento Lista de contatos, exemplificado na Figura 8, é considerado um documento bem formado (RAY, 2001, p. 60).

Figura 8 – Lista de contatos em um documento XML



Fonte: Elaborada pelo autor.

Em razão de sua estrutura hierárquica, um documento XML é entendido como uma árvore; seus componentes, chamados de nós, dividem-se em sete tipos:

- **elementos** – são os blocos de um documento XML; os nós do tipo elemento e o nó raiz (descrito a seguir) são os únicos nós que podem conter outros nós;
- **atributos** – acompanham os elementos descrevendo seus detalhes, tal como uma propriedade, um comportamento específico ou um subtipo; por mais que os atributos dependam dos elementos para existir em um documento XML, eles são considerados como nós separados;
- **textos** – são cadeias de caracteres que podem compor o conteúdo de um elemento;
- **comentários** – são notas no documento que não são interpretadas pelo processador de XML;

- **instruções de processamento** – são indicações para o processamento do documento XML por uma aplicação de informática específica;
- **namespaces** – os documentos XML podem conter elementos e atributos de diferentes vocabulários, nesses casos, para especificar de qual vocabulário provém cada elemento ou atributo, são adicionados prefixos de *namespaces* aos seus nomes; um *namespace* representa um grupo de elementos e atributos que compõem o vocabulário de uma linguagem de marcação, assim, o prefixo de *namespace* é uma cadeia de caracteres que precede o nome do elemento ou do atributo indicando a qual *namespace* ele pertence;
- **raiz** – é um ponto abstrato acima do elemento do documento que abrange todo o documento (RAY, 2001).

O documento da Figura 8 utiliza a linguagem de marcação Agenda, criada apenas para o propósito de exemplificação nesta pesquisa. A linguagem Agenda compreende os elementos *agenda*, *contato*, *nome*, *email*, *telefone*, *endereco*, *rua*, *numero*, *CEP* e *cidade* e seus atributos (*tipo* e *estado*), as regras sobre sua localização, sobre seu número mínimo ou máximo de ocorrência e sobre seu conteúdo, entre outros. Algumas dessas regras são: o elemento *contato* pode ocorrer quantas vezes for necessário dentro do elemento *agenda*; *cidade* pode ocorrer somente dentro do *endereco* e uma única vez; o conteúdo de cada *telefone* deve conter dez dígitos numéricos, etc.

No documento da Figura 8 nada impediria, por exemplo, que fosse acrescentado após o elemento *telefone* o elemento *ISBN*, de forma semelhante, nada impediria que, dentro do *endereco*, fossem adicionados dois elementos *cidade*. Esses acréscimos não tornariam o documento mal formado, uma vez que estariam de acordo com as restrições sintáticas da XML, no entanto, o tornaria inválido em relação à linguagem Agenda, uma vez que ela não permite a existência de um *ISBN* e nem a existência, dentro de um *endereco*, de mais de um elemento *cidade*.

Assim como o processamento, a validação de documentos XML é desempenhada por aplicações de informática, o que requer a especificação das linguagens de marcação (seus elementos, as regras sobre sua localização, etc.) de modo processável por essas aplicações.

O processo de especificação de uma linguagem de marcação é chamado de modelagem de documentos (RAY, 2001, p. 6), sendo que distintas tecnologias podem ser utilizadas para representar o resultado dessa modelagem.

Uma dessas tecnologias é a *Document Type Definition* (DTD) (Definição de Tipo de Documento), que compreende “regras ou declarações que especificam quais *tags*⁸ podem ser usadas e o que podem conter” (RAY, 2001, p. 6). Originalmente, as DTDs foram criadas para validar documentos SGML, no entanto, passaram a ser utilizadas também para a validação de documentos XML, embora ainda preservem características da SGML. Ou seja, uma DTD é um documento que contém especificações sobre uma linguagem para a marcação de documentos XML, no entanto, uma DTD não é um documento XML.

Como uma alternativa ao uso de DTDs, o W3C aprovou em 2001 a linguagem *XML Schema*, também utilizada para a especificação de linguagens de marcação. Essa linguagem, além de prover maiores possibilidades que a DTD, foi projetada para e seguindo a XML. Assim, um esquema XML, ou seja, um documento que especifica uma linguagem de marcação construído com a *XML Schema*, é também um documento XML, podendo fazer uso das mesmas ferramentas utilizadas na criação e na edição de documentos XML, o que é visto como uma vantagem da utilização de esquemas XML em vez de DTDs (RAY, 2001, p. 7).

Segundo o W3C (2012a, tradução nossa), os “esquemas XML expressam vocabulários⁹ compartilhados e permitem que máquinas executem regras feitas por pessoas. Eles provêm meios para definir a estrutura, o conteúdo e a semântica de documentos XML”. Os esquemas XML “também apresentam uma verificação mais poderosa quanto ao tipo de dados, permitindo a descoberta de erros no conteúdo e também no uso das *tags*” (RAY, 2001, p. 7). Para Tidwell (2008, p. 8), tanto a DTD quanto a *XML Schema* são metalinguagens que permitem a definição das características de um vocabulário XML.

Após a confrontação de um documento XML com a especificação da linguagem nele utilizada, diz-se que o documento é válido, se está de acordo com a linguagem, ou inválido, se não está de acordo.

Cabe notar que a especificação da linguagem de marcação por meio de uma DTD ou de um esquema XML e a validação dos documentos XML não são ações obrigatórias ao uso da XML, embora garantam a adequação do documento aos propósitos para os quais ele foi criado.

Por serem arquivos de texto, os documentos XML podem ser criados e editados em simples editores de texto, por exemplo, no aplicativo Bloco de notas que acompanha as diferentes versões do sistema operacional *Microsoft Windows*. Existem também editores de

⁸ Nessa citação o termo “*tags*” é utilizado como sinônimo de “elementos”.

⁹ Um vocabulário, como descrito pelo W3C, é entendido como o conjunto de termos de uma linguagem de marcação.

texto livres e gratuitos que oferecem maiores possibilidades que o Bloco de notas na criação e na edição de documentos XML, como é caso do *Notepad++*¹⁰. No entanto, esses editores de texto nem sempre estão aprimorados para trabalhar com documentos XML e podem não possibilitar a validação ou a transformação (abordada na seção seguinte) desses documentos. Para suprir essa necessidade, existem aplicações direcionadas e otimizadas para a criação e a edição de documentos XML, são algumas dessas aplicações: *Liquid XML Studio*, *Oxygen XML Editor*, *Stylus Studio*, *XML Notepad*, *XMLBlueprint* e *XMLSpy*.¹¹

A XML tem sido objeto de diversos estudos na Ciência da Informação, inclusive no cenário nacional. Desse modo, não será conduzida nesta pesquisa uma extensa apresentação sobre suas possibilidades de uso e seus aspectos técnicos.

Santos e Flamino (2004, p. 128) atentam para o fato de que “são as tecnologias associadas a XML que ampliam seu potencial”. A DTD e a *XML Schema* são apenas duas dessas tecnologias. Enquanto essas tecnologias destinam-se à validação de documentos, outras tecnologias foram criadas para atender a outros propósitos. A linguagem *Extensible Stylesheet Language for Transformation* (XSLT), tecnologia criada para a transformação de documentos XML e que possui um papel de destaque no modelo para a conversão de registros proposto nesta pesquisa, é apresentada na seção seguinte.

3.2 Transformação de documentos XML: folhas de estilo XSLT

Uma das características da XML é a diferenciação de estrutura e conteúdo do documento da sua apresentação (EITO BRUN, 2008, p. 89). A XML lida com a estruturação do conteúdo, ao passo que sua apresentação fica a cargo de outras tecnologias: as folhas de estilo (*stylesheets*). Em linhas gerais, uma folha de estilo é um conjunto de regras para a apresentação de um documento XML de modo a atender uma dada necessidade.

Para Eito Brun (2008, p. 90), as folhas de estilo permitem gerar múltiplas apresentações para um mesmo conteúdo. Segundo o autor, “se associamos um documento XML a distintas folhas de estilo, sua apresentação na tela ou impressa mudará, sem a necessidade de realizar alterações no próprio documento” (EITO BRUN, 2008, p. 90, tradução nossa). Ou seja, por meio de folhas de estilo é possível, por exemplo, fazer com que

¹⁰ O Notepad++ é um aplicativo livre e gratuito. Requer a instalação de um complemento (*plugin*) para executar a validação e a transformação de documentos XML. Mais informações podem ser obtidas em: <<http://notepad-plus-plus.org>>.

¹¹ Uma comparação entre alguns dos editores XML pode ser consultada em <http://en.wikipedia.org/wiki/Comparison_of_XML_editors>. Acesso em: 11 ago. 2013.

um determinado documento XML seja apresentado em um telefone celular de modo diferente do modo como é apresentado em um computador *desktop*. Podem ser criadas também uma apresentação para um *tablet* e uma versão para a impressão, sem que seja necessário alterar o documento XML original.

Atualmente, o W3C possui especificações de três linguagens para a criação de folhas de estilo. A primeira delas é a *Cascading Style Sheets* (CSS) (Folhas de Estilo em Cascata), que define as propriedades de um elemento de marcação e que, embora possa ser utilizada em documentos XML, é mais frequentemente utilizada para definir os estilos em um documento HTML. As propriedades da CSS podem ser usadas para definir, por exemplo, que o conteúdo de um determinado elemento será apresentado em azul, com uma fonte de tamanho 58 ou em negrito (TIDWELL, 2008, p. 1-2).

A segunda linguagem é a *Extensible Stylesheet Language for Transformation* (XSLT) (Linguagem Extensível para Folhas de Estilo de Transformação), criada para ser uma linguagem mais poderosa e flexível para a transformação de documentos XML (TIDWELL, 2008, p. 2). A especificação mais recente da XSLT (W3C, 2007) define, entre outros, um conjunto de elementos e atributos para a criação de regras de transformação que, em folhas de estilo, são utilizadas principalmente para transformar documentos XML em documentos HTML, visando sua apresentação em *browsers*, e para transformar documentos XML criados com uma linguagem de marcação em documentos de acordo com outra linguagem de marcação.

A terceira linguagem é a *Extensible Stylesheet Language* (XSL) (Linguagem Extensível para Folhas de Estilo), conhecida também como *Formatting Objects* (XSL-FO) (Formatação de Objetos), que lida com a transformação de documentos XML. Normalmente está associada à criação de documentos para exibição em *browsers* ou para a impressão, e de documentos paginados, por exemplo, documentos em *Portable Document Format* (PDF). No entanto, é utilizada também para transformar documentos XML como documentos em Braille ou como arquivos de áudio (EITO BRUN, 2008, p. 90; TIDWELL, 2008, p. 18).¹²

Entre as três linguagens para a criação de folhas de estilo, a presente pesquisa focará apenas na XSLT, pois, por possibilitar a transformação de documentos XML criados com uma linguagem de marcação em documentos de acordo com outra linguagem de marcação, a XSLT desempenha um papel de destaque na elaboração do modelo para a conversão de registros, principal objetivo desta pesquisa.

¹² O nome *Extensible Stylesheet Language* é também utilizado para denominar a família de linguagens para folhas de estilo que inclui a XSLT e a XSL-FO.

A primeira versão da XSLT foi publicada pelo W3C em 1999 (W3C, 1999). Em 2007 foi publicada a versão 2.0 (W3C, 2007), trazendo um conjunto maior de funções para a construção das regras de transformação. Atualmente o W3C trabalha no rascunho da versão 3.0 (W3C, 2012).

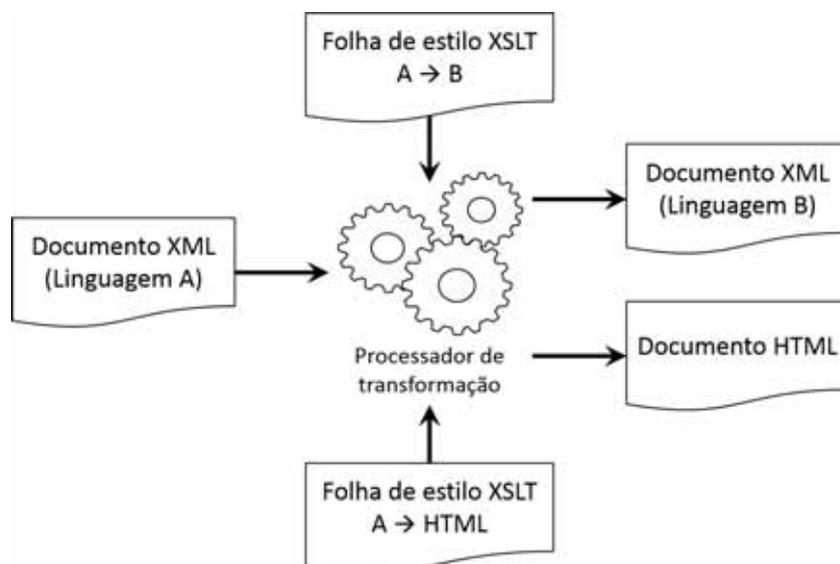
Eito Brun (2008, p. 95) destaca a importância da XSLT para as situações em que a XML é utilizada para transferir e intercambiar dados estruturados com diferentes esquemas, ou seja, estruturados de acordo com diferentes linguagens de marcação. Tais situações requerem uma tecnologia que permita transformar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

Bradley (2001) destaca a transformação de documentos XML de uma linguagem de marcação para outra como uma das principais aplicações da XSLT, sendo usualmente conduzida com vistas a possibilitar a troca de dados entre sistemas e organizações.

Ray (2001) aponta alguns motivos que levam à necessidade de transformar um documento em outro: armazenar em um formato, exibir em outro; converter para um formato mais útil; tornar o documento mais compacto; e usar o documento como um *front-end* para consultas a bancos de dados. Transformações como essas podem ser desempenhadas por aplicações de informática, por exemplo, por analisadores sintáticos. No entanto, Ray (2001, p. 203) destaca que “para muitos tipos de transformação, a escrita de um programa é desnecessariamente complicada”. O autor acrescenta que “a XSLT foi projetada especificamente para realizar transformações e nada mais, tornando-a mais fácil de aprender, mais simples de se ler e otimizada para a sua tarefa em particular”.

Como representado pela Figura 9, as transformações são realizadas por processadores, também referidos como processadores de transformação. Esses processadores recebem um documento XML e, com base no conjunto de regras de transformação presentes em uma folha de estilo XSLT, o transformam em um documento de saída, seja ele um documento XML ou outro tipo de documento (BRADLEY, 2001).

Figura 9 – Transformação de documentos XML utilizando folhas de estilo XSLT



Fonte: Elaborada pelo autor.

De modo geral, as regras de transformação de uma folha de estilo XSLT selecionam nós (elementos, atributos, textos, comentários, instruções de processamento, *namespaces* e/ou a raiz) do documento de entrada e indicam ao processador o que fazer com eles para gerar o documento de saída. A seleção desses nós e a indicação ao processador ocorrem por meio da utilização dos elementos e atributos da XSLT e de recursos de outras tecnologias, com a XPath. Entre as ações que as regras de transformação de uma folha de estilo XSLT podem realizar, estão:

- adicionar cadeias de caracteres antes ou depois do conteúdo de um elemento;
- remover, criar, reordenar e classificar o conteúdo de um elemento;
- converter o conteúdo do elemento em atributo e vice-versa;
- apresentar o conteúdo de elementos situados em qualquer parte do documento;
- substituir os elementos do documento de entrada por novos elementos no documento de saída (BRADLEY, 2001).

Os elementos e atributos da XSLT e o modo como as regras de transformação estão estruturadas ficarão mais evidentes após o exemplo de folha de estilo apresentado mais adiante nesta seção.

Com as especificações da XSLT, têm sido publicadas pelo W3C as especificações da XPath. A XPath é uma linguagem utilizada para indicar os nós de um documento XML (W3C, 2010). Ainda utilizando como exemplo o documento Lista de endereços da Figura 8,

por meio da XPath é possível referir-se ao primeiro elemento *contato*, ao valor do atributo *estado* do elemento *cidade*, a todos os elementos *email* que possuem em seu conteúdo a cadeia de caracteres *gmail.com*, e assim por diante.

Nas folhas de estilo XSLT, as expressões em XPath são utilizadas para indicar quais nós do documento de origem serão transformados, para selecionar um elemento, ou para construir uma condição a ser verificada, por exemplo, a existência de um elemento.

A XPath consiste em uma mistura de expressões básicas de linguagens de programação, por exemplo, `[^d]` (seleciona qualquer caractere que não seja numérico), e expressões de caminhos, ou seja, de localização, por exemplo, `/agenda/contato/endereco/rua`. Além dessa sintaxe básica, a XPath provê um conjunto de funções que permite descobrir diversas coisas sobre um documento XML (TIDWELL, 2008, p. 45).

Segundo Tidwell (2008, p. 55), um dos conceitos mais importantes na utilização da XPath é o de contexto, pois todas as expressões em XPath são interpretadas em relação ao contexto. Um documento XML pode ser pensado como a hierarquia de diretórios em um sistema de arquivos. Utilizando como exemplo o documento Lista de contatos da Figura 8, pode-se imaginar que na raiz do sistema de arquivos há o diretório *agenda*, dentro dele há um diretório *contato* contendo os diretórios *nome*, *email*, *telefone* e *endereco*. Nesse exemplo, o contexto é o diretório atual. Se um comando é executado, os resultados obtidos podem variar dependendo do diretório atual, ou seja, dependendo do contexto. De modo similar, os resultados de uma expressão em XPath provavelmente variarão em razão do contexto, ou seja, do nó em que o processador de transformação se encontra.

Para melhor explicar as regras de transformação, os elementos da XSLT e as expressões em XPath, segue o exemplo da transformação do documento Lista de contatos, que foi apresentado na Figura 8 e está de acordo com a linguagem de marcação Agenda, criada para propósitos de exemplificação nesta pesquisa. Neste exemplo, deseja-se obter como documento de saída um documento XML de acordo com a linguagem *Address Book*, também criada para exemplificação. O documento XML de saída deve parecer com o documento da Figura 10.

Figura 10 – Documento XML desejado

```

<?xml version="1.0" encoding="UTF-8"?>
  <addressBook>
    <person>
      <name> </name>
      <e-mail> </e-mail>
      <city> </city>
      <phone type=" ">
        <prefix> </prefix>
        <number> </number>
      </phone>
    </person>
  </addressBook>

```

Fonte: Elaborada pelo autor.

Para transformar o documento Lista de endereços em um documento semelhante ao documento da Figura 10 foi elaborada a folha de estilo da Figura 11.

Figura 11 – Folha de estilo para a transformação do documento Agenda

```

<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:output method="xml" indent="yes" version="1.0" encoding="UTF-8"/>
  <xsl:template match="/agenda">
    <addressBook>
      <xsl:for-each select="contato">
        <person>
          <name>
            <xsl:value-of select="nome" />
          </name>
          <xsl:for-each select="email">
            <e-mail>
              <xsl:value-of select="."/>
            </e-mail>
          </xsl:for-each>
          <xsl:if test="endereco/cidade">
            <city>
              <xsl:value-of select="endereco[1]/cidade"/>
            </city>
          </xsl:if>
        </person>
      </xsl:for-each>
    </addressBook>
  </xsl:template>
</xsl:stylesheet>

```

1º bloco

2º bloco

3º bloco

4º bloco

5º bloco

```

<xsl:for-each select="telefone">
  <xsl:variable name="phoneType">
    <xsl:choose>
      <xsl:when test="@tipo='residencial'">home</xsl:when>
      <xsl:when test="@tipo='comercial'">business</xsl:when>
      <xsl:otherwise>other</xsl:otherwise>
    </xsl:choose>
  </xsl:variable>
  <phone type="{ $phoneType }">
    <prefix>
      <xsl:value-of select="substring(.,1,2)"/>
    </prefix>
    <number>
      <xsl:value-of select="substring(.,3,8)"/>
    </number>
  </phone>
</xsl:for-each>
</person>
</xsl:for-each>
</addressBook>
</xsl:template>
</xsl:stylesheet>

```

} 6º bloco

} 7º bloco

Fonte: Elaborada pelo autor.

A folha de estilo da Figura 11 inicia com a declaração XML, uma vez que ela é também um documento XML, e com os elementos *xsl:stylesheet* e *xsl:output*. O elemento *xsl:stylesheet* é o elemento do documento, por isso compreende todas as regras de transformação da folha de estilo. Seu atributo *version* indica a versão da XSLT em uso na folha de estilo. O *xsl:output* indica que o documento resultante da transformação será um documento XML (*method="xml"*), que ele estará endentado (*indent="yes"*), que a versão da XML utilizada nele é a 1.0 (*version="1.0"*) e que o esquema de caracteres é o UTF-8 (*encoding="UTF-8"*).

O elemento *xsl:stylesheet* indica também o prefixo de *namespace* dos elementos da XSLT (*xsl*). Esse prefixo é utilizado pelo processador de transformação para determinar quais elementos fazem parte da estrutura da folha de estilo e quais devem ser criados no documento resultante da transformação. Assim, se o elemento está acompanhado de *xsl:*, ele faz parte de uma regra de transformação, caso contrário, fará parte do documento resultante (RAY, 2001, p. 210).

A primeira linha do 2º bloco indica a partir de qual elemento do documento de entrada terá início a transformação (*<xsl:template match="/agenda">*). Os elementos *xsl:template* contêm conjuntos de regras e podem aparecer mais de uma vez em uma folha de estilo. Um *xsl:template* pode receber um nome e ser chamado por tal nome, ou possuir uma indicação de

qual elemento do documento de entrada ele destina-se. Na Figura 11, o *xsl:template* destina-se ao elemento *agenda*, indicado no valor do atributo *match* pela expressão em XPath */agenda*.

A segunda linha do 2º bloco apresenta apenas o elemento *addressBook* que, por não estar acompanhado do prefixo *xsl:*, será inserido no documento resultante. Dentro dele, para cada elemento *contato* do documento de entrada (`<xsl:for-each select="contato">`) será criado um elemento *person* no documento de saída. Assim, se o documento de entrada possui dez elementos *contato*, o documento de saída possuirá dez elementos *person*. O conteúdo de cada um dos elementos *person* é indicado pelo 3º, 4º, 5º e 6º blocos.

O 3º bloco contém uma regra simples indicando que será criado o elemento *name* e que, para o seu conteúdo, será selecionado o conteúdo do elemento *nome* (`<xsl:value-of select="nome" />`).

O 4º bloco indica que, para cada elemento *email* (`<xsl:for-each select="email">`), será criado o elemento *e-mail* contendo o valor de tal elemento (`<xsl:value-of select="."/>`). O atributo *select* tendo como valor apenas um ponto indica o elemento atual, assim, se a regra anterior diz “para cada elemento *email*”, o elemento ao qual o ponto refere-se é o elemento *email*.

O 5º bloco indica que, se existir o elemento *endereco* e, dentro dele, o elemento *cidade* (`<xsl:if test="endereco/cidade">`), será criado o elemento *city* e, para seu conteúdo, será selecionado o conteúdo do elemento *cidade* da primeira ocorrência do elemento *endereco* (`<xsl:value-of select="endereco[1]/cidade"/>`). Assim, se um elemento *contato* possui mais de um elemento *endereco*, será incluído no documento de saída apenas a cidade indicada no primeiro elemento *endereco*.

O 6º bloco traz um conjunto de regras que será aplicado a cada elemento *telefone* (`<xsl:for-each select="telefone">`). Primeiramente estão as regras para definir o valor da variável *phoneType*. Uma variável recebe um nome e um valor. Seu valor pode ser uma cadeia de caracteres, um nó ou um conjunto de nós (W3C, 2007). Apesar de seu nome, o valor de uma variável é constante, ou seja, uma vez definido, não pode ser alterado até o fim da transformação (RAY, 2001, p. 253). Uma variável pode ser global, quando pode ser mencionada em qualquer parte da folha de estilo, ou local, quando pode ser mencionada somente em uma determinada regra (W3C, 2007). A cada vez que uma variável é mencionada, o processador de transformação substitui sua menção por seu valor.

A variável *phoneType* é uma variável local e seu valor é definido com o auxílio do elemento *xsl:choose*. Esse elemento contém um número qualquer de elementos *xsl:when* e um

elemento *xsl:otherwise*. Cada *xsl:when* indica uma condição a ser verificada e, se a condição for verdadeira, como o processador de transformação deve prosseguir. Caso nenhum dos *xsl:when* apresente uma condição verdadeira, o processador dirige-se ao *xsl:otherwise*, se presente, se não presente, nada é realizado pelo processador, que continuará o processamento do elemento seguinte da folha de estilo. Assim, no exemplo da Figura 11, é dito ao processador “escolha: quando o atributo *tipo* tiver como valor *residencial*, selecione o texto *home*; quando o atributo *tipo* tiver o valor *comercial*, selecione o texto *business*; de outro modo, selecione o texto *other*”.

Após a definição da variável e de seu valor, é criado o elemento *phone* e, em seu atributo *type*, é inserido o valor da variável *phoneType* (*home*, *business* ou *other*) (`<phone type="{ $phoneType }">`).

Dentro do elemento *phone* são criados os elementos *prefix* e *number*. Para selecionar o conteúdo do *prefix* é utilizado o elemento *xsl:value-of* com a função *substring*. Essa função permite selecionar um fragmento de uma cadeia de caracteres (*string*) e consiste em três argumentos: a cadeia de caracteres da qual se deseja selecionar um fragmento; a posição inicial do fragmento, indicada por um número; e o tamanho do fragmento em número de caracteres (WALMSLEY, 2007). Assim, a função *substring*(.,1,2) selecionará do nó atual (indicado pelo ponto) um fragmento que inicia na primeira posição de caractere e de tamanho igual a dois caracteres, ou seja, selecionará os dois primeiros dígitos do número de telefone. A função *substring* é utilizada também na seleção do conteúdo do elemento *number*: “selecione do nó atual um fragmento iniciando na terceira posição e de tamanho igual a oito” (`<xsl:value-of select="substring(.,3,8)" />`).

A *substring* e outras funções destinadas à manipulação de cadeias de caracteres são bastante úteis para a separação de uma cadeia de caracteres em diversos fragmentos que podem ser utilizados para os mais distintos propósitos (compor o conteúdo de um elemento, o valor de uma variável ou de um atributo, etc.).

O 7º bloco destina-se às *tags* de fim dos elementos *person*, *xsl:for-each*, *addressBook*, *xsl:template* e *xsl:stylesheet*.

A folha de estilo da Figura 11 exemplifica apenas alguns dos elementos da XSLT e algumas combinações desses elementos na elaboração de regras para a transformação de documentos XML, seja em documentos XML de acordo com outras linguagens de marcação ou outros tipos de documentos, por exemplo, uma página HTML.

Uma vez que a folha de estilo está pronta, existem diversas formas para realizar a transformação. Ela pode ser realizada em um processador local destinado à transformação ou, se o documento a ser transformado estiver *online*, pode ser utilizado o processador do servidor Web ou ainda o *browser* pelo qual o documento será acessado. Se a transformação for realizada no servidor Web ou no *browser*, é necessário indicar no documento XML a folha de estilo a ser utilizada para a transformação. Essa indicação deve ocorrer com uma instrução de processamento semelhante a `<?xml-stylesheet type="text/xml" href="endereço-da-folha-de-estilo.xsl"?>` (RAY, 2001, p. 211).

Tidwell (2008, p. 20-24) lista quatro processadores XSLT populares: *Xalan*, *Saxon*, *Microsoft XSLT Processor* e *Altova XSLT Engine*. Esses e outros processadores XSLT podem ser instalados isoladamente ou em conjunto com outras aplicações, por exemplo, com editores XML ou *browsers*.

Devido suas características, considera-se a XSLT uma tecnologia acessível que amplia as possibilidades de uso da XML, permitindo a transformação de documentos de modo a atender diferentes propósitos, tanto na apresentação quanto no intercâmbio de dados.

Em razão de sua ampla utilização nas aplicações de informática, surgiram esforços para viabilizar a utilização da XML também na codificação de registros catalográficos. A esses esforços somam-se as potencialidades da XSLT, que podem minimizar os conflitos resultantes da adoção de diferentes padrões de metadados, seja durante o intercâmbio de dados, seja no processo de migração entre sistemas de gerenciamento de bibliotecas. A utilização da XML na codificação de registros, alguns de seus marcos históricos, características e vantagens, assim como a transformação de registros utilizando a XSLT, são descritas na seção seguinte.

3.3 Marcação e transformação de registros MARC 21: MARCXML

Originalmente desenvolvida para suportar o intercâmbio de registros MARC em fitas magnéticas, a codificação definida pela ISO 2709 apresenta o caráter sequencial necessário ao processamento dos dados disponíveis em tal suporte. O desenvolvimento das tecnologias da informática trouxe, além de outros suportes para o armazenamento e posterior intercâmbio de registros, outros ambientes digitais, como é o caso da *World Wide Web*, ou simplesmente 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. Alguns dos esforços envolvidos nessa busca são destacados nesta seção.

Os esforços da LC na tentativa de codificar registros nos Formatos MARC de forma mais condizente com as tecnologias de informática de cada época iniciaram-se na década de 1990 com estudos sobre a utilização da SGML. Como resultado foram criadas duas DTDs para a codificação (marcação) de registros nos Formatos MARC utilizando a SGML. A primeira DTD era utilizada para registros bibliográficos, de itens e de informação comunitária; a segunda destinava-se aos registros de autoridade e de classificação.

As duas DTDs SGML definiam um elemento para cada campo e subcampo dos Formatos MARC. Por exemplo, para o campo 245 foram criados os elementos *mrcb245*, *mrcb245-a*, *mrcb245-b*, etc. Esses elementos foram agrupados em elementos como *mrcb-title-and-title-related*, *mrcb-subject-access*, etc. Para os campos de controle foram definidos elementos para cada dado. Essa abordagem utilizada na codificação de registros nos Formatos MARC utilizando a SGML resultou em DTDs extremamente largas e complexas (EITO BRUN, 2008, p. 153). Posteriormente, com o lançamento da XML, a LC voltou seus esforços para essa nova possibilidade de codificação.

Eito Brun (2008, p. 152) aponta que um dos objetivos que levaram à codificação de registros utilizando a XML foi evitar a segregação do MARC frente a outros padrões de metadados projetados para a Web e que prometiam se tornarem grandes protagonistas no ambiente das bibliotecas digitais. Ainda segundo o autor, manter a codificação tradicional do MARC como o único meio para a transferência de registros influenciaria negativamente nas possibilidades de intercambiar informações e de expor as descrições a outros serviços de informação.

Taylor e Dickmeiss (2005, p. 2) destacam que, com o aumento da ubiquidade da XML como um metaformato para o intercâmbio de dados, houve a percepção de uma necessidade de reprojeter os Formatos MARC 21 nos termos da XML. Desse modo, foram criadas duas DTDs para a codificação de registros nos Formatos MARC 21 utilizando a XML. Essas DTDs XML definiam todos os elementos que poderiam aparecer em um registro MARC 21 e especificavam como esses elementos seriam rotulados e representados com a XML (TAYLOR; JOUDREY, 2009, p. 153).

Embora essas DTDs fossem destinadas à codificação de registros com a XML, ainda mantinham as características das DTDs SGML. Ou seja, definiam elementos para cada campo e subcampo dos Formatos MARC 21, como mostrado na Figura 12, e eram documentos demasiado extensos (KEITH, 2004, p. 122). Segundo Siqueira (2003, p. 83), se impressas, as duas DTDs XML corresponderiam a um documento com mais de 600 páginas.

Figura 12 – Fragmento de um registro no Formato MARC 21 para Dados Bibliográficos codificado com a DTD XML

```

<mrcb-title-and-title-related>
  <mrcb245 i1="i1-1" i2="i2-2">
    <mrcb245-a>A cura de Schopenhauer :</mrcb245-a>
    <mrcb245-b>romance </mrcb245-b>
    <mrcb245-c>Irvin D. Yalom ; tradução de Beatriz Horta. -</mrcb245-c>
  </mrcb245>
</mrcb-title-and-title-related>
<mrcb-edition-imprint-etc>
  <mrcb250 i1="i1-blank" i2="i2-blank">
    <mrcb250-a>2. ed. rev. -</mrcb250-a>
  </mrcb250>
  <mrcb260 i1="i1-blank" i2="i2-blank">
    <mrcb260-a>Rio de Janeiro :</mrcb260-a>
    <mrcb260-b>Ediouro,</mrcb260-b>
    <mrcb260-c>2006</mrcb260-c>
  </mrcb260>
</mrcb-edition-imprint-etc>

```

Fonte: Elaborada pelo autor.

O surgimento da *XML Schema* levou a LC à criação de uma nova forma para a codificação de registros MARC 21 com a XML, surgiu assim o *MARC 21 XML Schema*, mais conhecido como MARCXML.

Para Eito Brun (2008, p. 155), o MARCXML tem como objetivos:

- facilitar a codificação de registros nos Formatos MARC 21 utilizando a XML;
- facilitar a captura e a integração de registros nos Formatos MARC 21 aos processos de coleta (*harvesting*) de metadados, característicos dos projetos *Open Archives Initiative* (OAI) (Iniciativa dos Arquivos Abertos) e similares;¹³
- permitir a descrição original de recursos utilizando os Formatos MARC 21 com a XML, ou seja, permitir que registros MARC 21 sejam criados diretamente como documentos XML usando uma ferramenta de criação/edição adequada;
- facilitar a inclusão de registros nos Formatos MARC 21 em textos completos também codificados com a XML, sendo essa uma possibilidade oferecida por padrões como o TEI P5, da *Text Encoding Initiative* (Iniciativa de Codificação de Texto), e permitida mediante o uso de *namespaces*.

¹³ Flaminio (2006) aborda a adequação do MARCXML à descrição de recursos informacionais nas iniciativas de arquivos abertos (*open archives*).

No desenvolvimento do MARCXML foi adotada uma abordagem diferente da utilizada nas DTDs. Em vez de criar um elemento para cada campo e subcampo, foram criados elementos XML para os diferentes tipos de campo do MARC 21 (*leader*, *controlfield*, *datafield* e *subfield*), sendo que as etiquetas dos campos, os indicadores e os códigos de subcampos passaram a ser indicados por meio de atributos XML. O resultado dessa abordagem foi um modo mais simples para codificar registros em qualquer um dos Formatos MARC 21 (bibliográfico, de autoridade, etc.) (EITO BRUN, 2008, p. 154). Uma representação do MARCXML é provida pela Figura 13.

Figura 13 – Marcação de registros nos Formatos MARC 21 com o MARCXML

```
<?xml version="1.0" encoding="UTF-8" ?>
<collection>
  <record type=" ">
    <leader> </leader>
    <controlfield tag="00X"> </controlfield>
    <datafield tag="XXX" ind1=" " ind2=" ">
      <subfield code="x"> </subfield>
    </datafield>
  </record>
</collection>
```

Fonte: Siqueira (2003, p. 88).

No MARCXML, o elemento do documento é o elemento *collection* que pode conter um número qualquer de elementos *record*. O tipo do registro é especificado pelo valor do atributo *type* que pode ser *Bibliographic*, *Authority*, *Holdings*, *Classification* ou *Community*.

Cada registro contém um *leader* e um número qualquer dos elementos *controlfield* e *datafield*. Cada *controlfield* requer o atributo *tag*, utilizado para indicar a etiqueta do campo de controle; o valor desse atributo é composto de dois zeros e um terceiro caractere alfanumérico. Cada *datafield* requer os atributos *tag*, *ind1* e *ind2*. O valor do atributo *tag* é composto por três caracteres alfanuméricos; o valor do *ind1* e do *ind2* pode ser um espaço em branco ou um caractere alfanumérico. Um *datafield* deve conter ao menos um *subfield*, sendo este especificado pelo atributo *code* e seu valor (um caractere alfanumérico indicando o código do subcampo) (LIBRARY OF CONGRESS, 2009). A Figura 14 exemplifica um registro no Formato MARC 21 para Dados Bibliográficos codificado de acordo com o MARCXML.

Figura 14 – Registro no Formato MARC 21 para Dados Bibliográficos codificado com o MARCXML

```

<?xml version="1.0" encoding="UTF-8" ?>
<marc:collection xmlns:marc="http://www.loc.gov/MARC21/slim"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.loc.gov/MARC21/slim
  http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  <marc:record>
    <marc:leader>00822nam a2200229 a 4500</marc:leader>
    <marc:controlfield tag="005">20121108132540.0</marc:controlfield>
    <marc:controlfield tag="008">121107s2007 bl a gr 000 0 por d</marc:controlfield>
    <marc:datafield tag="020" ind1=" " ind2=" ">
      <marc:subfield code="a">857642004X</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="082" ind1="0" ind2="4">
      <marc:subfield code="a">001.42</marc:subfield>
      <marc:subfield code="2">22</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="100" ind1="1" ind2=" ">
      <marc:subfield code="a">Pescuma, Derna</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="245" ind1="1" ind2="0">
      <marc:subfield code="a">Projeto de pesquisa :</marc:subfield>
      <marc:subfield code="b">o que é? como fazer? : um guia para sua elaboração
    </marc:subfield>
      <marc:subfield code="c">Derna Pescuma, Antonio Paulo F. de Castilho ; colaboração de
    Paulo Angelo Lorandi. -</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="250" ind1=" " ind2=" ">
      <marc:subfield code="a">4. ed. -</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="260" ind1=" " ind2=" ">
      <marc:subfield code="a">São Paulo :</marc:subfield>
      <marc:subfield code="b">Olho d'Água,</marc:subfield>
      <marc:subfield code="c">2007</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="300" ind1=" " ind2=" ">
      <marc:subfield code="a">96 p. :</marc:subfield>
      <marc:subfield code="b">il. ;</marc:subfield>
      <marc:subfield code="c">21 cm</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="490" ind1="1" ind2=" ">
      <marc:subfield code="a">Método ;</marc:subfield>
      <marc:subfield code="v">3</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="500" ind1=" " ind2=" ">
      <marc:subfield code="a">Segue as mais recentes normas da ABNT</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="650" ind1="0" ind2="4">
      <marc:subfield code="a">Redação acadêmica</marc:subfield>
    </marc:datafield>
    <marc:datafield tag="650" ind1="0" ind2="4">
      <marc:subfield code="a">Trabalhos científicos</marc:subfield>
  </marc:record>
</marc:collection>

```

```

    <marc:subfield code="x">Metodologia</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="650" ind1="0" ind2="4">
    <marc:subfield code="a">Relatórios</marc:subfield>
    <marc:subfield code="x">Redação</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="700" ind1="1" ind2=" ">
    <marc:subfield code="a">Castilho, Antonio Paulo Ferreira de,</marc:subfield>
    <marc:subfield code="d">1953-</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="700" ind1="1" ind2=" ">
    <marc:subfield code="a">Lorandi, Paulo Angelo</marc:subfield>
  </marc:datafield>
  <marc:datafield tag="830" ind1=" " ind2="0">
    <marc:subfield code="a">Método</marc:subfield>
  </marc:datafield>
</marc:record>
</marc:collection>

```

Fonte: Elaborada pelo autor.

O MARCXML foi pautado 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).

Para Keith (2004, p. 122), o MARCXML incorpora o aprendizado obtido no desenvolvimento das DTDs e tira proveito dos padrões tecnológicos atuais para prover uma abordagem melhor na codificação de registros com a XML. Segundo Eito Brun (2008, p. 152), a codificação de registros MARC com a XML traria como vantagens a utilização de uma codificação universalmente aceita, mais simples que a provida pela ISO 2709 e de processamento mais fácil, o que resultaria em uma menor complexidade no desenvolvimento de aplicações de informática para o processamento de registros MARC.

Keith (2004, p. 125) destaca que a codificação tradicionalmente utilizada nos registros MARC 21 não possibilita o fácil desenvolvimento de aplicações de informática, pois, decifrar as partes do diretório definido pela ISO 2709 e lidar com questões de conversão de caracteres são tarefas no desenvolvimento de aplicações de informática que requerem uma longa curva de aprendizado. O autor acrescenta que a codificação de registros com a XML facilita drasticamente o acesso aos desenvolvedores de hoje em dia, uma vez que, diferentemente dos

desenvolvedores que lidam com a XML, um desenvolvedor que lide com registros MARC 21 codificados com a ISO 2709 é algo relativamente raro.

Do ponto de vista de um profissional da Ciência da Computação, a facilidade em lidar com registros nos Formatos MARC 21 codificados com a XML é abordada por Siqueira (2003). Segundo o autor, a codificação de um registro MARC 21 com a XML torna mais simples o desenvolvimento de programas para a importação de dados nesse padrão, principalmente considerando que (1) a XML é de conhecimento dos programadores; (2) as ferramentas existentes trazem recursos que facilitam a recuperação das etiquetas, posições, indicadores, subcampos e seus respectivos conteúdos; (3) a própria estrutura da XML provê um meio de validação dos dados; e (4) um registro MARC 21 com a XML pode ser apresentado de várias formas por meios de folhas de estilo (SIQUEIRA, 2003, p. 89-90).

Em posse do MARCXML, existem duas formas de se obter registros codificados com a XML. A primeira delas é a criação de registros já codificados com a XML, o que pode ocorrer, por exemplo, por meio de editores XML. A segunda forma envolve a conversão de registros codificados com a ISO 2709 em registros codificados com a XML.

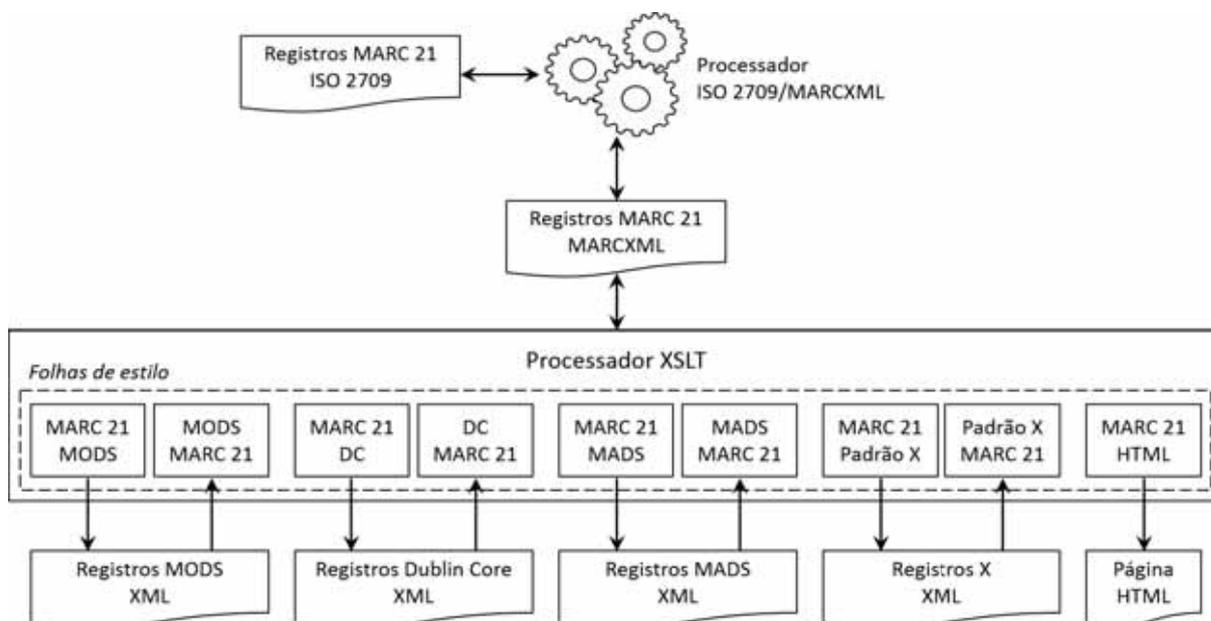
Para essa conversão existem aplicações de informática que processam registros codificados com a ISO 2709 e criam um documento XML contendo os registros de acordo com o MARCXML. Entre essas aplicações estão a disponibilizada pela LC¹⁴ e a que acompanha a suíte de aplicativos livre MarcEdit¹⁵.

Uma vez que os registros nos Formatos MARC 21 estão codificados com a XML é possível convertê-los para registros em uma variedade de padrões de metadados ou gerar distintas formas de apresentação por meio de folhas de estilo XSLT, como mostrado na Figura 15.

¹⁴ Mais informações disponíveis em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 11 ago. 2013.

¹⁵ Mais informações disponíveis em: <<http://people.oregonstate.edu/~reaset/marcedit/html/index.php>>. Acesso em: 11 ago. 2013.

Figura 15 – Transformação de registros nos Formatos MARC 21



Fonte: Elaborada pelo autor com base em Keith (2004) e Flaminio (2006).

Os registros codificados com a ISO 2709 são processados por uma aplicação de informática que os transforma em registros codificados com a XML seguindo o MARCXML. Os registros em MARCXML são inseridos em um processador XSLT que, com uma folha de estilo, os transforma em registros de acordo com outro padrão de metadados. A conversão no sentido oposto também é possível: um registro de acordo com determinado padrão de metadados e codificado com a XML pode ser transformado em um registro MARCXML e, então, em um registro MARC 21 com a ISO 2709. Para a conversão entre os principais padrões de metadados já existem folhas de estilo disponíveis.¹⁶ Cabe notar que uma folha de estilo que realiza a transformação de *A para B* não realiza a transformação de *B para A*, sendo necessária, para isso, outra folha de estilo.

Além da conversão entre padrões de metadados, as folhas de estilo podem ser utilizadas também para a transformação de registros em páginas HTML, de modo que possam ser apresentados em *browsers*, ou de acordo com outras formas de apresentação destinadas a atender a objetivos específicos. Uma das utilidades da transformação de registros MARCXML em páginas HTML é a apresentação em catálogos, assim, folhas de estilo XSLT

¹⁶ As folhas de estilo MARCXML-MODS, MODS-MARCXML, MARCXML-Dublin Core, Dublin Core-MARCXML, ONIX-MARCXML e OAI MARC-MARCXML estão disponíveis em: <<http://www.loc.gov/standards/marcxml>>. A folha de estilo MARCXML-MADS está disponível em: <<http://www.loc.gov/standards/mads>>. Acesso em: 11 ago. 2013.

que realizam essa transformação podem ser encontradas em sistemas de gerenciamento de bibliotecas, como é o caso do Koha¹⁷. Um exemplo de registro no Formato MARC 21 para Dados Bibliográficos transformado em uma página HTML para exibição no Koha é mostrado na Figura 16.

Figura 16 – Página HTML criada a partir da transformação de um registro MARCXML com uma folha de estilo XSLT



Fonte: Registro bibliográfico elaborado pelo autor no Koha.

Segundo Keith (2004, p. 124-125), uma meta que motivou o desenvolvimento do MARCXML foi permitir o fácil acesso a pequenos conjuntos de dados de um registro. Esse fácil acesso, conseguido por expressões em XPath que permitem a seleção do conteúdo de um registro no nível dos subcampos, favorece a criação de folhas de estilo para a transformação dos registros. Ainda segundo o autor, embora a XSLT não disponha de todas as características e do controle presente em uma linguagem de programação como a *Java*, é bastante surpreendente o que pode ser realizado com ela. Além disso, transformar documentos XML utilizando folhas de estilo XSLT traz diversos benefícios: as folhas de estilo podem ser facilmente modificadas nos editores de texto mais simples, sem que haja a necessidade de uma aplicação de informática específica para isso, e os profissionais que atuam em bibliotecas

¹⁷ Mais informações disponíveis em: <<http://koha-community.org>>. Acesso em: 11 ago. 2013.

e que não são desenvolvedores de aplicações de informática podem ser capazes de realizar modificações em folhas de estilo requerendo pouco auxílio.

Alves (2010, p. 74) destaca que

O MARC XML trouxe maior flexibilidade ao formato MARC 21, outro requisito importante nos princípios gerais para a construção de padrões de metadados atualmente. Dessa forma, bibliotecas digitais que compartilham metadados mesmo que em formatos distintos (padrão DC e padrão MARC 21, por exemplo), conseguem promover mais adequadamente a interoperabilidade dos dados, por compartilharem metadados em uma sintaxe de codificação compatível.

Assim com os Formatos MARC 21, o MARCXML também recebe críticas. Miller e Clarke (2003, p. 103, tradução nossa) destacam que

Os recentes esforços para modernizar o MARC tem, em sua maioria, envolvido uma tradução literal e sem crítica dos campos, indicadores e subcampos para elementos e atributos XML. Infelizmente, isso incorpora no novo meio os problemas inerentes ao MARC e não permite que se faça uso das reais vantagens da XML.

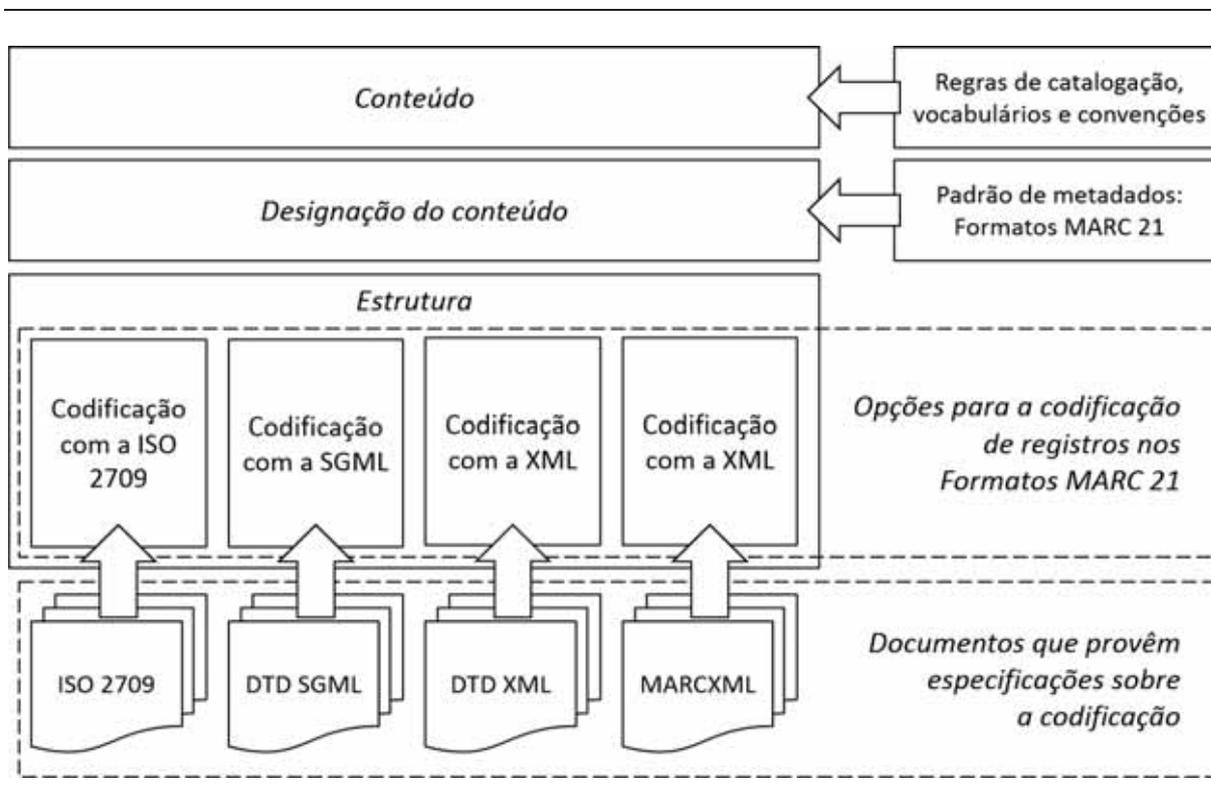
Eito Brun (2008, p. 151-152) apresenta uma visão semelhante. Para o autor, as iniciativas para a codificação de registros MARC com a XML, frequentemente chamadas de “x-emelificação do MARC”, propõem o uso da XML para representar e transferir registros, no entanto sem questionar a organização lógica do registro ou o significado dos campos, dos subcampos e dos indicadores.

Apesar das críticas, com o MARCXML há a possibilidade de codificar registros nos Formatos MARC 21 utilizando a XML e obter maior consonância com as tecnologias de informática atuais. Mesmo com essa possibilidade, nota-se que a comunidade de bibliotecas ainda está presa à codificação utilizando a estrutura desenvolvida para fitas magnéticas. Observa-se também que, como destacam Hillmann et al. (2010), o MARCXML é utilizado por poucos sistemas de gerenciamento de bibliotecas e apenas na saída de dados visando posteriores transformações.

A título de sumarização, um registro em qualquer um dos Formatos MARC 21 é composto por: um conteúdo, registrado de acordo com regras de catalogação, vocabulários e convenções; por uma designação do conteúdo, provida por algum dos Formatos MARC 21; e por uma estrutura, a codificação que permite ao registro ser processado por aplicações de informática. Ao longo dos anos buscaram-se diferentes formas para a codificação de registros, iniciando pela codificação de caráter sequencial voltada às fitas magnéticas, mais tarde reconhecida como a ISO 2709, passando pela SGML e chegando à XML, num primeiro

momento com as DTDs e posteriormente com o MARCXML. Os componentes e as possibilidades de codificação de registros MARC 21 estão sintetizados na Figura 17.

Figura 17 – Componentes e possibilidades de codificação de registros nos Formatos MARC 21



Fonte: Elaborada pelo autor.

Cabe observar que a codificação especificada pela ISO 2709 teve sua origem junto ao padrão de metadados Formato MARC, o que fez com que tal padrão e os padrões que dele derivaram (MARC II, UNIMARC, MARC 21, etc.) se tornassem bastantes dependentes dessa codificação. Como resultado, ainda hoje, os registros estão codificados de modo a atender as necessidades das fitas magnéticas, sendo a codificação com a XML utilizada apenas para propósitos de transformação com folhas de estilo e não como um veículo para o intercâmbio de registros entre instituições na Web.

Tendo como base o MARCXML, foi criada a norma ISO 25577 *Information and documentation – MarcXchange* (Informação e documentação – *MarcXchange*), que especifica um esquema XML genérico utilizado para a codificação de qualquer registro que possa estar codificado com a ISO 2709 (TAYLOR; DICKMEISS, 2005, p. 2), por exemplo, um registro bibliográfico no Formato UNIMARC.

Nota-se também que foram conduzidos outros projetos para a codificação de registros MARC com a XML além dos realizados pela LC, como é o caso do XMLMARC

desenvolvido pela *Lane Medical Library* da Universidade de Stanford, nos Estados Unidos, e o BiblioML, subsidiado pelo Ministério da Cultura e Comunicação da França (EITO BRUN, 2008, 161). No entanto, o MARCXML, desenvolvido pela instituição mantenedora dos Formatos MARC 21, a LC, tornou-se o mais conhecido.

Junto aos esforços para a utilização da XML na codificação de registros, a LC tem projetado padrões de metadados que apresentam alternativas ao uso dos tradicionais Formatos MARC 21. Entre esses padrões estão o *Metadata Object Description Schema* (MODS) (Esquema de Metadados para a Descrição de Objeto) e o *Metadata Authority Description Schema* (MADS) (Esquema de Metadados para a Descrição de Autoridade), ambos criados para o uso com a XML e especificados por esquemas XML.

O MODS e o MADS apresentam grande compatibilidade com os tradicionais Formatos MARC 21, embora, de modo geral, não possibilitem o registro de dados com o nível de especificidade possibilitado pelos Formatos MARC 21.

Apesar do MODS já existir há uma década e o MADS há mais de sete anos, observa-se, principalmente na literatura nacional, a falta de estudos e de relatos sobre a implantação desses padrões, ao passo que é possível encontrar diversos estudos sobre a utilização, no domínio bibliográfico, de um padrão não projetado para atender as necessidades de tal domínio, que é o caso do Dublin Core (ASSUMPCÃO et al., 2013).

Em maio de 2011 a LC fez um anúncio sobre sua iniciativa para uma nova estrutura bibliográfica. Afirmando a inadequação dos Formatos MARC 21 às necessidades atuais, inclusive àquelas relacionadas à utilização do *Resource Description and Access* (RDA) (Descrição de recursos e acesso), a LC conduz uma iniciativa cuja atividade central é o desenvolvimento de um novo meio para a captura e o compartilhamento de dados bibliográficos e que incluirá a busca por um substituto para os Formatos MARC 21 (LIBRARY OF CONGRESS, 2011).

Embora a iniciativa da LC e os padrões de metadados MODS e MADS estejam relacionados à situação atual e futura dos Formatos MARC 21, não se pretende abordá-los nesta pesquisa, uma vez que não se enquadram nos objetivos desta.

Apesar das críticas e dos esforços para sua substituição ou, ao menos, para a substituições de sua tradicional codificação, os Formatos MARC 21 ainda estão consolidados no domínio bibliográfico e são utilizados em âmbito internacional, tanto por agências bibliográficas nacionais quanto por programas de catalogação cooperativa.

Com base nessa consideração, esta pesquisa defende e propõe a conversão dos registros que não estão de acordo com qualquer padrão de metadados internacionalmente utilizado em registros nos Formatos MARC 21, de modo que as instituições que não utilizam tais padrões possam inserir-se no cenário internacional do intercâmbio de registros e minimizar a perda de dados e o retrabalho durante a migração entre sistemas de gerenciamento de bibliotecas.

Entende-se também que, uma vez que os registros estiverem de acordo com um padrão de metadados internacionalmente utilizado, mais fácil será a transição para um cenário com uma nova estrutura bibliográfica e com um novo padrão, considerando que as instituições que utilizarem os Formatos MARC 21 poderão acompanhar e tirar proveito dos estudos e das ferramentas para a transição desenvolvidas em âmbito internacional.

Com base na ampla utilização dos Formatos MARC 21 no domínio bibliográfico, nas vantagens da utilização desse padrão e nos benefícios da marcação e da transformação trazidos pelas tecnologias de informática XML e XSLT, é elaborado um modelo para a conversão de registros de distintos padrões de metadados e codificados com a XML em registros nos Formatos MARC 21, codificados com a XML ou com a ISO 2709. Esse modelo é descrito no capítulo seguinte.

4 Modelo para a conversão de registros

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, também denominada de retroconversão, reconversão ou RECON¹⁸, envolve a inserção dos dados de registros analógicos, de modo geral presentes em fichas catalográficas, em registros processáveis por aplicações de informática (ASENSI ARTIGA; RODRÍGUEZ MUÑOZ, 2001; MARTINELLI, 1998¹⁹ apud 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.

Para a realização da conversão retrospectiva podem ser adotados diversos métodos. Um método frequentemente apresentado na literatura é a importação de registros a partir de bases de dados externas à instituição que deseja realizar a conversão. São realizadas consultas a essas bases de dados para saber se dispõem de registros digitais correspondentes aos registros analógicos da instituição. Os registros correspondentes são importados para uma base de dados local, onde são realizadas as modificações necessárias para sua adequação às necessidades da instituição (BOWMAN, 2007; CASTRO, 2003; DARKO-AMPEM, 2006; ETUBI, 2009; OLIVEIRA et al., 1998).

Para Martinelli (1998, p. 34 apud Castro, 2003, p. 2), o princípio desse método “é o aproveitamento de registros existentes em outras bases na formação da base de dados local, diminuindo o esforço de catalogar em máquina o material da biblioteca”.

Perez e Lima (2002) relatam a conversão retrospectiva utilizando como método a inserção dos dados de fichas catalográficas diretamente em registros no Formato MARC 21 para Dados Bibliográficos, ou seja, sem a importação a partir de uma base de dados externa. Para os autores, “a conversão de registros bibliográficos antigos é uma tarefa onerosa e de

¹⁸ RECON refere-se a *Retrospective Conversion*, um projeto de conversão retrospectiva iniciado em 1969 na LC (AVRAM, 2003). Hoje, a sigla é frequentemente encontrada na literatura como um termo genérico para designar qualquer projeto de conversão retrospectiva.

¹⁹ MARTINELLI, A. T. S. A base de dados bibliográficos de acervos como suporte para o processo de automação: uma experiência na UNESP-Universidade Estadual Paulista. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITARIAS, 10., Fortaleza, 1998. *Anais...* Fortaleza: FINEP, 1998. p. 2-34.

trabalho intenso porque, em geral, esses registros devem ser atualizados de acordo com novas regras de catalogação e classificação” (PEREZ; LIMA, 2002, p. 2).

Outro método utilizado na conversão retrospectiva é a captura dos dados dos registros analógicos por meio do *Optical Character Recognition* (OCR) (reconhecimento ótico de caracteres) (BOWMAN, 2007; ZAFALON, 2012).

Bowman (2007, p. 333-334) conta que, em 1978, um dispositivo de OCR produzido pela empresa *Kurzweil Computer Products* foi visto como um meio para a conversão retrospectiva. Poucos anos depois, foi publicado um relatório sobre o uso desse dispositivo na *Glasgow University Library*. O relatório apontava que o reconhecimento ótico de caracteres alcançou uma precisão de aproximadamente 85% na identificação dos dados e permitiu a conversão da média de dez registros por hora, sendo necessário, no entanto, editar os registros convertidos e incluir campos.

Ainda segundo o autor, apesar de muitos experimentos, o método de conversão utilizando o OCR permaneceu não comprovado e somente nos últimos anos foi utilizado com maior sucesso (BOWMAN, 2007, p. 334). Um dos casos de sucesso da utilização do OCR na conversão retrospectiva é descrito por Zafalon (2012).

A autora fez uso do OCR na identificação dos dados presentes em fichas catalográficas. Após sua identificação, os dados foram processados por um analisador sintático e convertidos em registros no Formato MARC 21 para Dados Bibliográficos. O analisador sintático conseguiu processar os dados extraídos das fichas porque foi criado com base em um modelo teórico-conceitual de sintaxe e semântica em registros bibliográficos, desenvolvido pela autora a partir de estudos linguísticos saussureanos e hjelmslevianos das manifestações da linguagem humana (ZAFALON, 2012, p. 25).

Sobre os diferentes métodos para a conversão retrospectiva, Oliveira et al. (1998) destacam que

A escolha de um método eficaz que seja adequado às características e necessidades das bibliotecas e que assegure a conversão de catálogos manuais para informatizados, com prazos de execução razoáveis, custo baixo e garantindo a qualidade das informações, são critérios que devem prevalecer na definição de um modelo de conversão retrospectiva.

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, a literatura sobre conversão passou a preocupar-se não somente com a conversão retrospectiva (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 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 largo 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. A autora aponta também que uma das razões que pode levar à necessidade desse tipo de conversão é a atualização para um novo sistema de gerenciamento de bibliotecas que utilize padrões diferentes dos utilizados pelo sistema anterior.

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 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 21.²⁰

Na literatura brasileira são encontrados estudos e relatos sobre a conversão retrospectiva (CASTRO, 2003; DIAS, 1999; PEREZ; LIMA, 2002) e sobre a reutilização de

²⁰ 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.

metadados (BOICA; OLIVEIRA, 2008; MURAKAMI, 2012; RAPOSO; OLIVEIRA; SHINOTSUKA, 1985).²¹

Relacionado à conversão de registros, a literatura tem apresentado também o termo migração. Antes de prosseguir, é necessário distinguir migração de conversão. A migração é a transposição dos registros de uma aplicação de informática para outra, por exemplo, de um sistema de gerenciamento de bibliotecas para outro. A migração pode ou não requerer a conversão dos registros.

Zafalon (2012, p. 62) entende que a conversão ocorre no nível da estrutura do registro, sem que haja alterações em seu conteúdo. No entanto, para os objetivos desta pesquisa, adota-se a concepção de que a conversão, na vertente da reutilização de metadados, requer alterações em qualquer um dos componentes do registro: na codificação, na designação do conteúdo e/ou no próprio conteúdo.

As alterações na codificação incluem a mudança de uma codificação para outra, por exemplo, de ISO 2709 para XML e de XML para *Comma-Separated Values* (CSV). Considerando que, de modo geral, cada padrão de metadados possui um conjunto próprio de etiquetas, de códigos de subcampos, de elementos, etc., a alteração da designação do conteúdo é algo característico da conversão. As alterações no conteúdo, por sua vez, envolvem principalmente a modificação dos valores de modo a melhor se adequarem ao padrão de metadados de destino.

Partindo da necessidade de converter registros presentes em ambientes digitais, na ampla utilização dos Formatos MARC 21 e nas possibilidades oferecidas pelas tecnologias de informática XML e XSLT, esta pesquisa, considerada do tipo reutilização de metadados, apresenta neste capítulo um modelo para a conversão de registros (seção 4.1), junto aos seus processos e componentes (seções 4.2 à 4.6).

4.1 Modelo para a conversão de registros

Como abordado anteriormente, 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 padrões de metadados e as codificações aceitas pela aplicação de informática que receberá os registros.

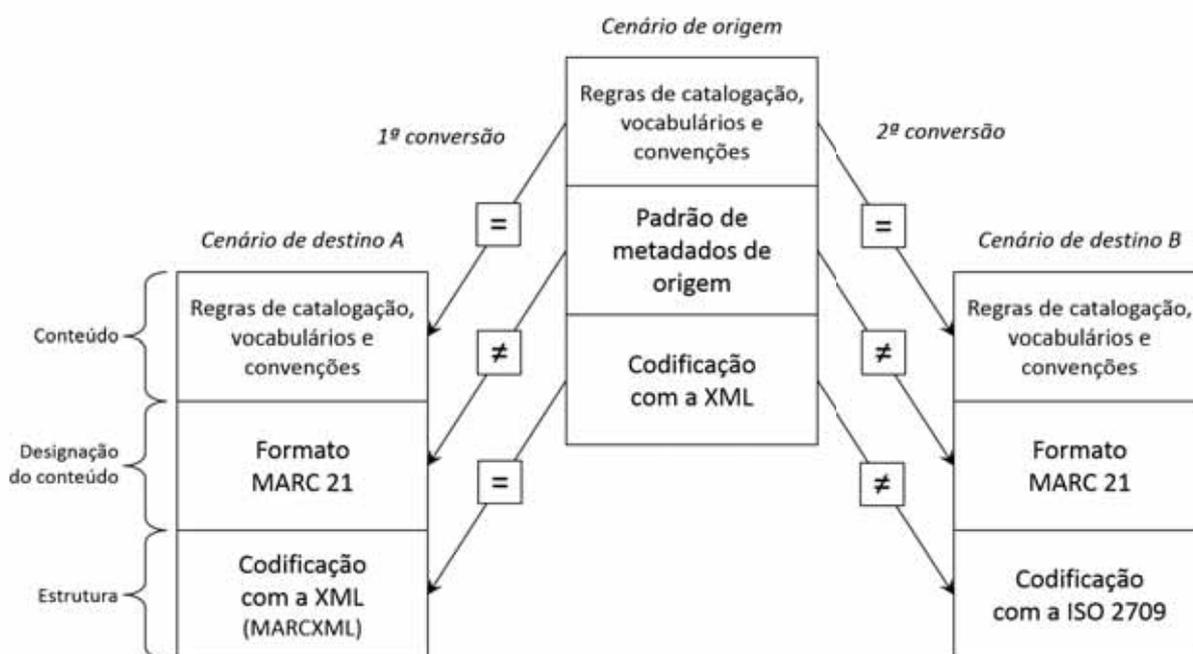
²¹ Os estudos sobre a reutilização de metadados, principalmente os nacionais, não tem empregado essa denominação. Em vez disso, utilizam apenas os termos conversão ou migração.

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 ou com a ISO 2709.

A partir dessa necessidade foram delineados um **cenário de origem**, situação em que se encontram os registros a serem convertidos (padrão de metadados qualquer/XML), e dois cenários de destino: **cenário de destino A** (MARC 21/MARCXML) e **cenário de destino B** (MARC 21/ISO 2709). Os três cenários estão representados na Figura 18.

Figura 18 – Cenários para a conversão de registros (1)



Fonte: Elaborada pelo autor.

A Figura 18 indica duas conversões: do cenário de origem para o cenário de destino A, e do cenário de origem para o cenário de destino B.

Na primeira conversão, do cenário de origem para o cenário de destino A, a designação do conteúdo empregada no padrão de metadados de origem é diferente da

empregada no Formato MARC 21, assim, essa conversão requer uma modificação no nível da designação do conteúdo. Nos Formatos MARC 21 a designação é composta por códigos numéricos e alfanuméricos. Alguns padrões de metadados, principalmente aqueles derivados de algum dos Formatos MARC, também utilizam códigos numéricos e alfanuméricos. Outros padrões de metadados criados mais recentemente, no entanto, têm empregado termos como *title*, *creator*, *publisher*, *typeOfResource*, etc., em vez de códigos como 245, 260, 300, etc.

Na primeira conversão, a codificação tanto do cenário de origem quanto do cenário de destino A é a XML. Como apresentado na seção 3.1, a XML não é uma linguagem de marcação, portanto não define quaisquer elementos que possam ser utilizados para a marcação de um documento. Ou seja, para a codificação de registros com a XML são necessárias linguagens de marcação. No cenário de origem a linguagem de marcação condiz com o padrão de metadados de origem; no cenário de destino A, a linguagem de marcação é o MARCXML, abordado na seção 3.3.

Observa-se que os padrões de metadados utilizados no domínio bibliográfico e, de modo mais geral, na Ciência da Informação, trazem junto às linguagens de marcação a designação do conteúdo. Por exemplo, para indicar o tipo do recurso informacional representado no registro, o padrão de metadados MODS estabelece a designação do conteúdo *typeOfResource* e, em sua linguagem de marcação, especifica o uso dessa designação na codificação de registros com a XML (LIBRARY OF CONGRESS, 2013). Essa característica faz com que as linguagens de marcação para a codificação de registros sejam resultantes da junção do padrão de metadados com a codificação. Outro exemplo dessa junção é a linguagem de marcação MARCXML.

Tais observações permitem concluir que a primeira conversão consiste em transformar um documento XML que está de acordo com uma dada linguagem de marcação (resultado da junção do padrão de metadados de origem com a XML) em um documento de acordo com a linguagem de marcação MARCXML (junção do Formato MARC 21 com a XML). A transformação de documentos XML, como discutido na seção 3.2, é uma tarefa que pode ser realizada por folhas de estilo XSLT, assim, tais folhas podem ser utilizadas na primeira conversão.

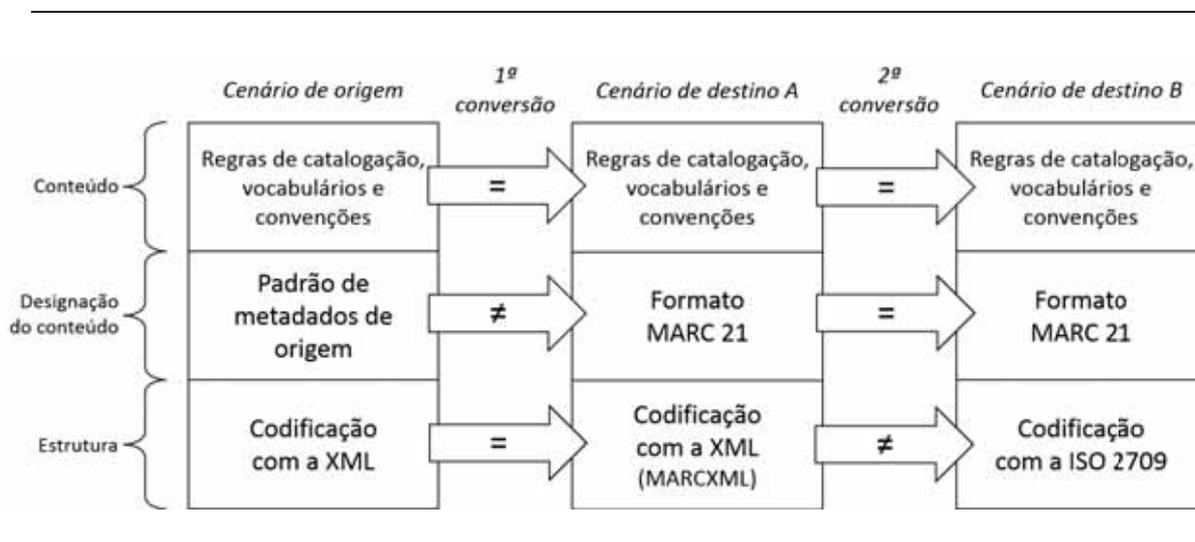
A segunda conversão mostrada na Figura 18, do cenário de origem para o cenário de destino B, requer a alteração da designação do conteúdo (do padrão de metadados de origem para o Formato MARC 21) e da codificação (da XML para a ISO 2709). Entende-se que para essa conversão, principalmente em razão da alteração da codificação, seria necessário criar,

utilizando alguma linguagem de programação, um analisador sintático para processar o registro de origem, reorganizar seu conteúdo adicionando a designação definida pelo Formato MARC 21 e codificar os registros de acordo com os componentes líder, diretório e campos, definidos na ISO 2709.

Considerando que existem sistemas de gerenciamento de bibliotecas que utilizam padrões de metadados próprios, um analisador sintático criado para a conversão dos registros exportados por um dado sistema não poderia ser utilizado na conversão dos registros de um sistema que utilizasse um padrão de metadados diferente. Além disso, como mencionado na revisão de literatura da seção 3.3, encontrar um profissional da Ciência da Computação que lide com a codificação definida na ISO 2709 é algo relativamente raro. Assim, a segunda conversão apresenta dois principais empecilhos: a necessidade de analisadores sintáticos para cada padrão de metadados e a dificuldade em encontrar profissionais que possam desenvolver tais analisadores.

Superar esses empecilhos e obter registros MARC 21 codificados com a ISO 2709 de modo acessível é possível por meio da adoção de uma abordagem diferenciada para a conversão dos registros. Essa abordagem é apresentada na Figura 19.

Figura 19 – Cenários para a conversão de registros (2)



Fonte: Elaborada pelo autor.

A abordagem apresentada na Figura 19 consiste na utilização do cenário de destino A com um intermediário entre o cenário de origem e o cenário de destino B. A primeira conversão, do cenário de origem para o cenário de destino A, discutida anteriormente, pode ser realizada por uma folha de estilo XSLT. O resultado dessa conversão são registros MARC 21 de acordo com o MARXML. A segunda conversão, do cenário de destino A para o

cenário de destino B, requer a alteração da codificação dos registros, passando de MARCXML para ISO 2709. Como apresentado na seção 3.3, para essa alteração estão disponíveis ferramentas, inclusive gratuitas, que são de fácil utilização pelos profissionais da Ciência da Informação.

Essa abordagem elimina a necessidade de um analisador sintático desenvolvido como uma aplicação específica, que, em termos gerais, seria de elaboração dispendiosa. A utilização de uma folha de estilo XSLT, por sua vez, está pautada nas vantagens apresentadas pela XML e pelas tecnologias relacionadas a ela, vantagens essas discutidas no capítulo anterior:

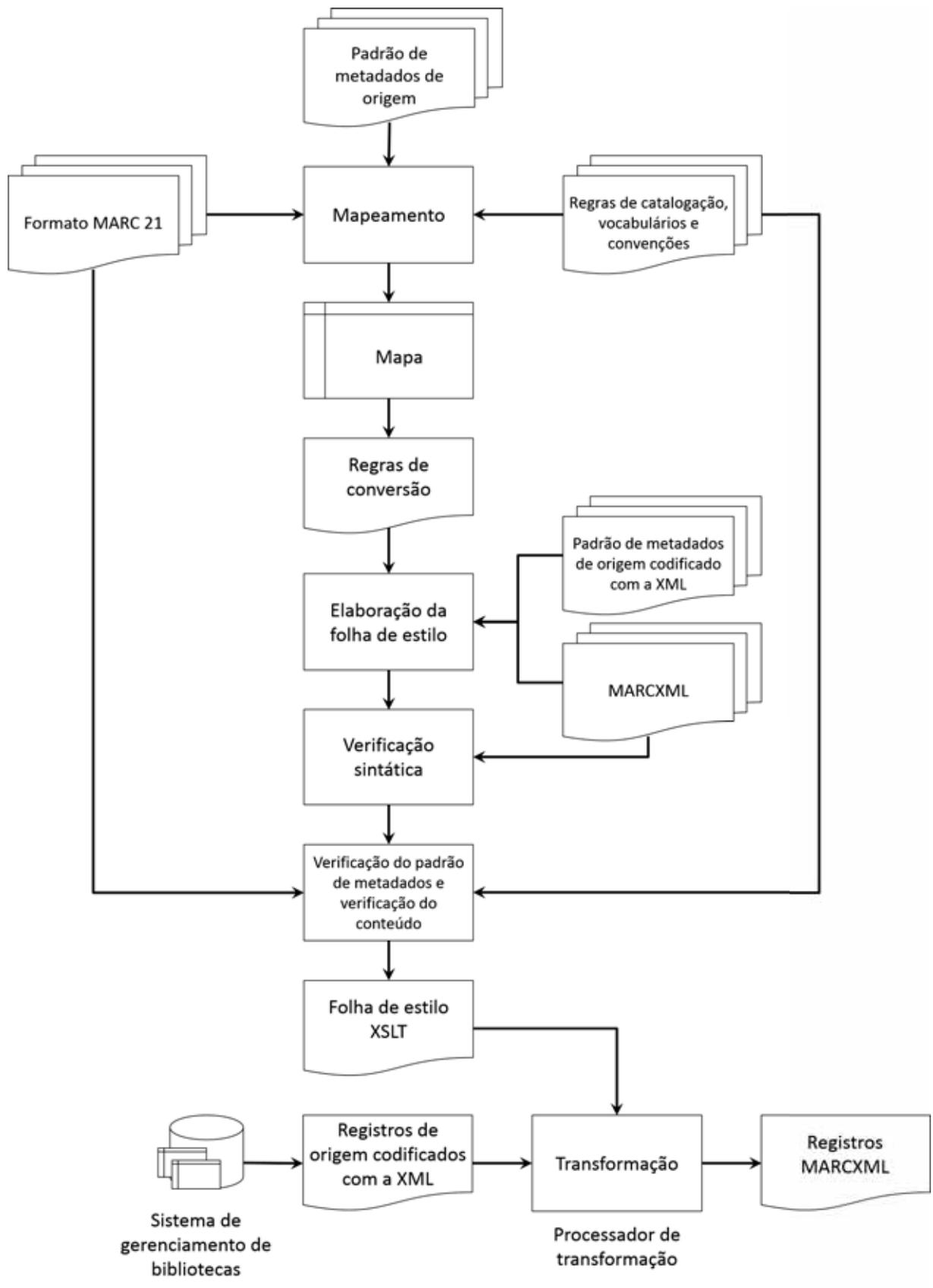
- a utilização e conhecimento da XML pelos profissionais da Ciência da Computação;
- a disponibilidade de aplicações de informática para a criação e a edição de documentos XML;
- as possibilidades de transformação providas pelas folhas de estilo XSLT;
- a facilidade da XSLT em relação às linguagens de programação; e
- a disponibilidade de processadores que permitem a transformação de documentos XML por meio de folhas de estilo XSLT.

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; RUDIC; 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 sequências 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 20 um modelo para a conversão de registros construído tendo como um de seus componentes centrais folhas de estilo XSLT.

Figura 20 – Modelo para a conversão de registros



Fonte: Elaborada pelo autor

O modelo para a conversão de registros é descrito 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.

Como abordado na seção 3.3, os registros codificados de acordo com o MARCXML podem ser convertidos em registros codificados com a ISO 2709 ou em registros de acordo com outros padrões de metadados.

Embora não explicitado na Figura 20, ao final de cada um dos processos do modelo devem ser realizadas ações de avaliação, o que garantirá maior segurança aos processos e permitirá realizar os ajustes necessários à realização dos processos posteriores.

Os processos e os componentes do modelo para a conversão de registros representado na Figura 20 são descritos de forma pormenorizada nas seções seguintes.

4.2 Mapeamento e 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 para um mapeamento significativo são necessárias definições claras e precisas dos metadados dos padrões a serem mapeados. Os autores pontuam também que o mapeamento inclui um mapa semântico e especificações para a conversão. Essas especificações indicam as transformações requeridas para converter o conteúdo do registro do padrão de origem em um registro de acordo com o padrão de destino.

Para Kurth, Ruddy e Rupp (2004, p. 156-157) há pouco acordo na literatura sobre a separação entre o mapeamento e a conversão, uma vez que a fronteira entre esses dois processos é imprecisa. Diferentemente de St. Pierre e LaPlant (1998), os autores entendem que apenas a elaboração do mapa faz parte do mapeamento, já a conversão, também chamada de transformação, é o projeto e a implementação de *scripts* e de outras ferramentas que moverão os dados de acordo com o mapa.

Rudić e Surla (2009, p. 952), por sua vez, apontam que uma descrição de conversão pode conter uma tabela e regras de conversão. Essa tabela de conversão é o mapa resultante do mapeamento entre os padrões de metadados. Já as regras de conversão são o mesmo que as “especificações para a conversão” (ST. PIERRE; LAPLANT, 1998) e que as “decisões de mapeamento escritas como instruções em linguagem natural” (KURTH; RUDDY; RUPP, 2004, p. 156-157), sendo esses três termos tratados aqui como sinônimos.

Segundo Machovec (2002, p. 2), o mapeamento entre quaisquer padrões não será perfeito e será mais ou menos bem sucedido dependendo dos padrões mapeados. Na literatura são discutidos diversos pontos que influenciam no mapeamento dos padrões de metadados. Alguns desses pontos são brevemente destacados a seguir.

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, na prática, frequentemente os padrões de metadados mapeados não possuem o mesmo nível de granularidade, o que torna a conversão mais complexa. 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, ou seja, um metadado do padrão de origem tem como correspondente mais de um metadado no padrão de destino e vice-versa. Woodley (2008, p. 7) destaca a existência de casos em que não há correspondências perfeitas entre os metadados, o que requer o estabelecimento de correspondências entre metadados com significados próximos.

Alguns metadados do padrão de origem podem 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 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. Essas propriedades indicam, por exemplo, se os metadados são obrigatórios ou opcionais, repetíveis ou não repetíveis, o tipo de dado que devem possuir em seu conteúdo (data, texto, URL), se o conteúdo deve ser extraído de um vocabulário (pontos de acesso, códigos de idiomas, de países, etc.), entre outros.

Rudić e Surla (2009, p. 958) chamam a atenção para a necessidade de levar em conta a pontuação que deve ser adicionada ao conteúdo dos metadados.

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 arranjados, 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. A elaboração dessas regras é de responsabilidade do profissional da Ciência da Informação, que deve redigi-las em linguagem natural de forma mais clara e precisa possível, indicando os relacionamentos dados no mapa e, opcionalmente, provendo orientações sobre os instrumentos de descrição utilizados na criação dos registros de origem.

A importância das regras de conversão reside 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, uma vez que essa não é sua área de competência, mas sim do profissional da Ciência da Informação

Na etapa seguinte do modelo, as regras de conversão serão inseridas na folha de estilo XSLT como regras de transformação. Assim, quanto maior a semelhança entre as regras de conversão e as regras de transformação, mais fácil poderá ser a elaboração da folha de estilo. Para favorecer essa semelhança, são sugeridas para a redação das regras de conversão as expressões-chave do Quadro 1.

Quadro 1 – Expressões-chave para a elaboração de regras de conversão

Expressão-chave	Equivalente na XSLT
Se presente o <i>nome-do-metadado</i> ...	<xsl:if test=" <i>nome-do-metadado</i> "> ... </xsl:if>
Se <i>condição</i> for verdadeira...	<xsl:if test=" <i>condição</i> "> ... </xsl:if>
Crie um <i>nome-do-metadado</i> ...	< <i>nome-do-metadado</i> > ... </ <i>nome-do-metadado</i> >
Inclua o conteúdo do <i>nome-do-metadado</i> .	<xsl:value-of select=" <i>nome-do-metadado</i> "/>
Escolha: quando <i>condição</i> for verdadeira... de outro modo...	<xsl:choose> <xsl:when test=" <i>condição</i> "> ... </xsl:when> <xsl:otherwise> ... </xsl:otherwise> </xsl:choose>
Para cada ocorrência do <i>nome-do-metadado</i> ...	<xsl:for-each select=" <i>nome-do-elemento</i> "> ... </xsl:for-each>

Expressão-chave	Equivalente na XSLT
Para a primeira ocorrência do <i>nome-do-metadado</i> ...	<code><xsl:if test="nome-do-metadado[1]"> ... </xsl:if></code>
Para cada ocorrência do <i>nome-do-metadado</i> que não seja a primeira...	<code><xsl:for-each select="nome-do-elemento[position()>1]"> ... </xsl:for-each></code>
Junte o conteúdo do <i>nome-do-metadado-A</i> , <i>nome-do-metadado-B</i> , <i>nome-do-metadado-C</i> , etc.	<code><xsl:value-of select="concat(nome-do-elemento-A,nome-do-elemento-B,nome-do-elemento-C)"/></code>

Fonte: Elaborado pelo autor.

Obviamente, essas expressões-chave não cobrem todas as necessidades de um mapeamento, no entanto, auxiliam na redação das regras mais frequentes, tais como:

- se presente o *metadado-ISBN*, crie um 020; crie um \$a; inclua o conteúdo do *metadado-ISBN*;
- crie um 260; se presente o *metadado-Local-de-publicação*, crie um \$a; inclua o conteúdo do *metadado-Local-de-publicação*; se presente o *metadado-Publicador*, crie um \$b; inclua o conteúdo do *metadado-Publicador*; se presente o *metadado-Data-de-publicação*, crie um \$c; inclua o conteúdo do *metadado-Data-de-publicação*.
- para cada ocorrência do *metadado-Contribuidor*, crie um 700; crie um \$a; inclua o conteúdo do *metadado-Contribuidor*;
- para a primeira ocorrência do *metadado-Autor*, crie um 100; crie um \$a; inclua o conteúdo do primeiro *metadado-Autor*;
- para cada ocorrência do *metadado-Autor* que não seja a primeira, crie um 700; crie um \$a; inclua o conteúdo do *metadado-Autor*.

Em síntese, os componentes necessários ao mapeamento são: a documentação do padrão de metadados de origem, a documentação do Formato MARC 21 adequado ao tipo de registro que será convertido (bibliográfico, de autoridade, etc.), as regras de catalogação, as instruções sobre a utilização dos vocabulários e as convenções. Os componentes resultantes do mapeamento são: o mapa e as regras de conversão. As regras de conversão serão utilizadas na elaboração da folha de estilo XSLT, abordada na seção seguinte.

4.3 Elaboração da folha de estilo XSLT

Além do mapa e das regras de conversão, para a elaboração da folha de estilo são necessárias as especificações das linguagens de marcação dos padrões de metadados de origem e de destino. 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 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 linguagem de marcação utilizada com o padrão de metadados de destino é o MARCXML (LIBRARY OF CONGRESS, 2011b). A especificação dessa linguagem foi descrita na seção 3.3.

Além desses componentes, devem ser escolhidos um editor XML, um processador de transformação e a versão da XSLT. Os conhecimentos sobre XSLT podem ser obtidos ou aprimorados por meio das especificações publicadas pelo W3C (1999, 2007, 2010, 2012) e de livros (TIDWELL, 2008), guias, manuais e tutoriais (WALMSLEY, 2007) sobre essa linguagem e, de forma mais geral, sobre a XML (BRADLEY, 2001; EITO BRUN, 2008; MILLER; CLARKE, 2004; RAY, 2001).

Tendo por base esses componentes, inicia-se a elaboração da folha de estilo, que compreende principalmente dois passos:

- a inclusão da declaração XML, a inclusão do elemento do documento (*xsl:stylesheet*) dentro do qual será inserido todo o conteúdo da folha de estilo, e a inclusão das declarações dos *namespaces*. Uma vez que a folha de estilo incluía elementos de diferentes linguagens de marcação (padrão de metadados de origem, padrão de metadados de destino e a própria linguagem XSLT), é necessário indicar a qual linguagem pertence cada elemento, sendo essa a função dos *namespaces*; e
- 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 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/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.

4.4 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 com a conversão.

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 (regras de catalogação, vocabulários e convenções). Assim, a verificação é composta por três etapas: verificação sintática, verificação do padrão de metadados e verificação do conteúdo.

A verificação sintática, realizada por editores XML, é a análise do registro MARCXML para verificar se o mesmo está de acordo com o esquema da linguagem de marcação MARCXML.

A 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, no caso, um dos Formatos MARC 21.

A 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. Essa verificação 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.

Como um método adicional para as etapas de verificação do padrão de metadados e de verificação do conteúdo, é sugerida a comparação entre registros originalmente criados com o Formato MARC 21 e com os instrumentos de descrição e registros obtidos após a conversão. Essa comparação elucidará quais dados não puderam ser obtidos a partir da conversão.

Os resultados das três verificações proveem um *feedback* indicando se há ou não a necessidade de modificações no mapa, nas regras de conversão e/ou na folha de estilo. Uma

vez assegurado que os registros resultantes da conversão utilizando a folha de estilo estão de acordo com o padrão de metadados de destino e com os instrumentos de descrição, não é necessário realizar as verificações a cada utilização da folha de estilo.

4.5 Exportação dos registros e conversão para MARCXML

Entre os processos do modelo, a exportação dos registros é o mais dependente do sistema de gerenciamento de bibliotecas utilizado pela instituição que deseja realizar a conversão. Para a utilização do modelo de conversão 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 e mais comum é por meio de uma interface, seja ela gráfica ou não, provida pelo sistema de gerenciamento de bibliotecas. De modo geral, na interface desses sistemas há indicações das possibilidades de exportação no que diz respeito ao padrão de metadados e a codificação.

Um segundo modo ocorre quando o sistema de gerenciamento de bibliotecas não oferece uma interface de exportação, mas é possível acessar a base de dados por meio do sistema de gerenciamento de banco de dados. Esse modo pode permitir a exportação dos registros codificados com a XML, porém não utilizando qualquer padrão de metadados, apenas os rótulos atribuídos aos campos das tabelas da base de dados.

Considerando esses dois modos, o componente necessário ao processo exportação dos registros é a documentação sobre o sistema de gerenciamento de bibliotecas e/ou sobre o sistema de gerenciamento de banco de dados.

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 (1.0 ou 2.0) 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 fim da transformação, o processador gera um documento XML contendo

os registros convertidos de acordo com o MARCXML. Com esse documento podem ser criados registros codificados com a ISO 2709, como descrito na seção seguinte, ou registros de acordo com outros padrões de metadados, como abordado na seção 3.3.

4.6 Conversão para ISO 2709

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.

De modo geral, as mesmas ferramentas que realizam a conversão de ISO 2709 para MARCXML, entre elas aquelas mencionadas na seção 3.3 (o conjunto de aplicações disponibilizado pela LC e a suíte de aplicativos MarcEdit), também realizam a conversão de MARCXML para ISO 2709.

Neste capítulo foram apresentados o modelo para a conversão de registros proposto nesta pesquisa, seus componentes e seus processos. O modelo, elaborado a partir da literatura sobre conversão de registros, MARC 21, ISO 2709, XML, XSLT e MARCXML, precisa, no entanto, ser validado. Para essa validação, foi elencada como objetivo específico desta pesquisa a aplicação do modelo na conversão dos registros bibliográficos do sistema de gerenciamento de bibliotecas PHL. Essa aplicação do modelo é descrita no capítulo seguinte.

5 Aplicação do modelo na conversão de registros do Formato PHL

Um sistema de gerenciamento de bibliotecas que não possibilita a exportação dos registros nos Formatos MARC 21 é o *Personal Home Library* (PHL), como mencionado anteriormente, utilizado por diversas instituições brasileiras. Com o intuito de converter os registros bibliográficos exportados pelo PHL em registros no Formato MARC 21 para Dados Bibliográficos foi utilizado o modelo para conversão apresentado no capítulo anterior. A aplicação do modelo à conversão dos registros do PHL é descrita neste capítulo.

5.1 *Personal Home Library* (PHL)

Lançado em 2001 e desde então distribuído gratuitamente para o uso em estações monousuárias²², o *Personal Home Library* (PHL) se apresenta como “uma aplicação Web especialmente desenvolvida para administração de coleções e serviços de bibliotecas e centros de informações” (OLIVEIRA, 2002).

O PHL utiliza um banco de dados CDS/ISIS. Como servidor de banco de dados, é utilizado o WWWisis, desenvolvido pela BIREME. As linguagens utilizadas no desenvolvimento do PHL são a *XML-IsisScript*, *Javascript* e HTML. Sendo que os *scripts* que compõem o PHL são distribuídos com o código-fonte aberto, permitindo que sejam personalizados (ESPECIFICAÇÕES, 2009).

O módulo de catalogação do PHL possui planilhas de catalogação customizáveis e possibilita a catalogação de recursos informacionais em diferentes níveis:

- monográfico: para a descrição de um documento no todo, desde que não pertença a uma série periódica, por exemplo, um livro, uma tese, um CD, etc.;
- analítico monográfico: para a descrição de partes de um documento que não pertença a uma série periódica, por exemplo, capítulos de livros, faixas de CDs, etc.;
- série: para a descrição de fascículos de periódicos;
- analítico de série: para a descrição de artigos de periódicos;
- coleção: para a descrição das coleções como um todo (OLIVEIRA, 2011, p. 76).

Como padrão de metadados, o PHL utiliza um padrão próprio, denominado aqui Formato PHL, baseado no padrão LILACS. O Formato PHL é entendido como o padrão de metadados constituído pelos elementos, etiquetas, rótulos, definições e instruções

²² A utilização do PHL em rede, seja intranet ou internet, requer o licenciamento. Mais informações estão disponíveis em: <<http://www.elysio.com.br/site/custos.html>>. Acesso em: 11 ago. 2013.

apresentadas no manual do PHL (OLIVEIRA, 2011). Assim como os campos dos Formatos MARC 21, os campos do Formato PHL possuem etiquetas numéricas, destinadas às aplicações de informática, e rótulos textuais, destinados a agentes humanos. Um fragmento da documentação do Formato PHL é apresentado na Figura 21

Figura 21 – Exemplo de campos, etiquetas e definições do Formato PHL

062 – Editora (obrigatório)

Nome da editora, produtora, gravadora ou empresa responsável pela publicação do documento. Na ausência desta informação, registra-se [s.n.].

064 – Data de Publicação (obrigatório)

Data de publicação, produção ou gravação do documento. Quando a data for desconhecida, registra-se [s.d.].

065 – Data de Publicação Padronizada (obrigatório)

Padronização da data registrada no campo (64 – Data), no padrão ISO (AAAAMMDD) para efeito de busca e ordenação do resultado de uma busca ou geração de relatórios e gráficos estatísticos.

066 – Cidade de Publicação (obrigatório)

Nome da cidade onde foi publicado, produzido ou gravado o documento. Quando a cidade não for conhecida, transcreve-se [s.l.]

067 – País de Publicação (facultativo)

Nome do país de publicação do documento.

Fonte: Adaptado de Oliveira (2011, p. 90-91).

Na Figura 21, o campo do Formato PHL destinado ao registro no nome do publicador do recurso informacional recebe o rótulo “Editora”, a etiqueta “062”, a definição “Nome da editora, produtora, gravadora ou empresa responsável pela publicação do documento” e a instrução “Na ausência desta informação, registra-se [s.n.]”. Cabe notar que, em alguns casos, as definições e instruções são dadas em uma mesma declaração.

Assim como o Formato MARC 21 para Dados Bibliográficos, o Formato PHL dispõe de campos para uso local, de modo que a instituição usuária possa incluir metadados não previstos no padrão, mas necessários às suas necessidades, sejam elas descritivas ou administrativas.

Observa-se que não há no PHL ou em seu manual uma distinção sobre o que são campos, etiquetas, rótulos, definições, instruções ou mesmo sobre o Formato PHL. Essas denominações, inclusive “Formato PHL”, foram estabelecidas nesta pesquisa para o propósito de comparação com o Formato MARC 21 para Dados Bibliográficos.

Um registro exemplificando as etiquetas disponíveis no Formato PHL é apresentado na Figura 22.

Figura 22 – Registro bibliográfico no Formato PHL

001: 001
 002: 211081436231
 003: 0001.42
 103: P281p
 005: Livro
 006: m
 007: ^a000003^b1^c1^d20121222^e1^m2
 009: N
 015: Colaboração de Paulo Angelo Lorandi
 016: Pescuma, Derna
 016: Castilho, Antonio Paulo Ferreira de, 1953-
 018: Projeto de pesquisa
 020: 96
 028: Nacional
 029: S
 030: Método, 3
 040: Português
 061: Segue as mais recentes normas da ABNT
 062: Olho d'Água
 063: 4. ed.
 064: 2007
 065: 20070000
 066: São Paulo
 069: 85-7642-004-X
 085: Outros
 087: Redação acadêmica
 087: Trabalhos científicos^sMetodologia
 087: Relatórios^sRedação
 090: b
 181: o que é? como fazer? : um guia para sua elaboração
 998: 3
 999: ^d20121108^h143623^bsuper
 999: ^d20121108^h143656^bsuper

Fonte: Elaborado pelo autor no PHL 8.2.

A forma de apresentação de um registro no Formato PHL utilizada na Figura 22 destina-se à compreensão dos dados por humanos e não à codificação do registro para processamento por aplicações de informática.

Em suas especificações, o PHL traz que, para o preenchimento do conteúdo dos registros é utilizado o AACR2 (CÓDIGO..., 2004) e que, para a apresentação destes no catálogo, é utilizada a norma “Informação e documentação – Referências – Elaboração” da Associação Brasileira de Normas Técnicas (ABNT) (NBR 6023:2002) (ESPECIFICAÇÕES, 2009).

Observa-se que, se utilizada, a pontuação prescrita no AACR2r para a separação dos elementos e das áreas de descrição entra em conflito com a forma de apresentação do registro, uma vez que, na apresentação, o PHL acrescenta ao final de cada campo a pontuação dada pela norma da ABNT. Assim, entende-se que as bibliotecas usuárias do PHL não utilizam a pontuação prescrita pelo AACR2r para anteceder cada elemento de descrição, sendo essa pontuação utilizada apenas entre elementos presentes em um mesmo campo.

Um registro no Formato PHL, assim como qualquer registro nos Formatos MARC 21, possui três componentes: o conteúdo, a designação do conteúdo e a estrutura. O conteúdo de um registro no Formato PHL pode estar de acordo com as mesmas regras de catalogação, vocabulários e convenções utilizadas na criação de um registro bibliográfico no Formato MARC 21. A designação do conteúdo compreende as etiquetas dos campos do Formato PHL. Para a estrutura do registro, o PHL oferece cinco possíveis codificações, são elas: ISO 2709, CISIS, VLINE, HLINE e XML.

A codificação com a ISO 2709 segue a estrutura genérica descrita na seção 2.2. A CISIS destina-se à utilização do registro na biblioteca de funções desenvolvida pela BIREME para a manipulação de bases de dados ISIS²³. As codificações VLINE e HLINE mostram-se semelhantes à CISIS. Para a codificação com a XML são utilizadas as etiquetas do Formato PHL na composição dos elementos do documento XML.

Uma vez que o modelo para a conversão descrito no Capítulo 4 requer que os registros do padrão de metadados de origem estejam codificados com a XML, tal codificação será utilizada na conversão dos registros bibliográficos no Formato PHL em registros no Formato MARC 21 para Dados Bibliográficos. Assim, para proporcionar um melhor entendimento, a codificação de registros PHL com a XML será aqui denominada PHLXML (em comparação ao MARCXML). O PHLXML é representado na Figura 23.

Figura 23 – Marcação de registros no Formato PHL com o PHLXML

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<db>
  <rec>
    <vXXX> </vXXX>
  </rec>
</db>
```

Fonte: Elaborada pelo autor

²³ Mais informações em: <<http://wiki.bireme.org/pt/index.php/CISIS>>. Acesso em: 11 ago. 2013.

No PHLXML, o elemento *db* compreende todo o conteúdo do documento, exceto a declaração XML. Dentro do elemento *db* há um elemento *rec* para cada registro presente no documento. Em cada elemento *rec* existe um elemento para cada campo do registro. O nome desses elementos é composto pela letra *v* e pela etiqueta do campo no Formato PHL.²⁴ O elemento *v018*, por exemplo, corresponde ao campo título, indicado pela etiqueta 018 no Formato PHL. A Figura 24 exemplifica um registro no Formato PHL com o PHLXML.

Figura 24 – Registro no Formato PHL codificado com o PHLXML

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<db>
  <rec>
    <v001>001</v001>
    <v002>211081436231</v002>
    <v003>001.42</v003>
    <v005>7</v005>
    <v006>m</v006>
    <v007>^a000003^b1^c1^d20121222^e1^m2</v007>
    <v009>N</v009>
    <v015>Colaboração de Paulo Angelo Lorandi</v015>
    <v016>Pescuma, Derna</v016>
    <v016>Castilho, Antonio Paulo Ferreira de, 1953-</v016>
    <v018>Projeto de pesquisa</v018>
    <v020>96</v020>
    <v028>1</v028>
    <v029>S</v029>
    <v030>Método, 3</v030>
    <v040>1</v040>
    <v061>Segue as mais recentes normas da ABNT</v061>
    <v062>Olho d'Água</v062>
    <v063>4. ed.</v063>
    <v064>2007</v064>
    <v065>20070000</v065>
    <v066>São Paulo</v066>
    <v069>85-7642-004-X</v069>
    <v085>9</v085>
    <v087>Redação acadêmica</v087>
    <v087>Trabalhos científicos^sMetodologia</v087>
    <v087>Relatórios^sRedação</v087>
    <v090>b</v090>
    <v103>P281p</v103>
    <v181>o que é? como fazer? : um guia para sua elaboração</v181>
    <v998>3</v998>
    <v999>^d20121108^h143623^bsuper</v999>
    <v999>^d20121108^h143656^bsuper</v999>
  </rec>
```

²⁴ Embora as etiquetas definidas no Formato PHL não contenham a letra *v*, na codificação com a XML tal letra foi acrescentada pelo próprio sistema PHL, uma vez que na XML não é permitido que o primeiro caractere de uma etiqueta seja numérico.

</db>

Fonte: Elaborada pelo autor.

No sistema de gerenciamento de bibliotecas PHL, os dados sobre os itens estão registrados à parte dos registros bibliográficos. Se a instituição possui três exemplares de um mesmo recurso informacional haverá um registro bibliográfico representando o recurso e três registros de itens, cada qual representando um item. Apesar dos dados bibliográficos persistirem separadamente dos dados de item, durante a exportação alguns dados dos itens são condensados e incluídos em campos 007 do registro bibliográfico. Os dados presentes nesse campo são, então, delimitados pelos códigos *^a* (número de tombo), *^b* (número do exemplar), *^c* (código do tipo de aquisição), *^d* (data de aquisição), *^e* (código do status de circulação), *^f* (prazo excepcional) e *^m* (*Master File Number* (MNF) do registro). Esses códigos podem ser observados no elemento v007 da Figura 24.

Após a etapa de compreensão das características gerais do padrão de metadados de origem, segue-se a etapa de mapeamento entre os padrões de metadados de origem (Formato PHL) e de destino (Formato MARC 21 para Dados Bibliográficos). Esse mapeamento é abordado na seção seguinte.

5.2 Mapeamento e regras de conversão

O mapeamento realizado abrangeu as correspondências existentes entre os campos do Formato PHL e os campos, indicadores e subcampos do Formato MARC 21 para Dados Bibliográficos. A partir da necessidade de considerar também os instrumentos de descrição, durante o mapeamento levaram-se em conta as regras de catalogação presentes no *Anglo-American Cataloguing Rules* (AACR2r) (Código de Catalogação Anglo-Americano), um vocabulário local qualquer para a atribuição dos pontos de acesso de assunto e os vocabulários e as convenções presentes no Manual do PHL (OLIVEIRA, 2011). O resultado do mapeamento é apresentado no Quadro 2.

Quadro 2 – Mapa do Formato PHL e do Formato MARC 21 para Dados Bibliográficos

Nível	Formato PHL		Formato MARC 21 para Dados Bibliográficos				
	Tag	Campo	Tag	I1	I2	Sub.	Campo: subcampo
–	001	Código da Biblioteca	–	–	–	–	–
M ²⁵ A ²⁶	002	Identificação do título	001	–	–	–	Número de controle

²⁵ M – Nível monográfico.

Nível	Formato PHL		Formato MARC 21 para Dados Bibliográficos				Campo: subcampo
	Tag	Campo	Tag	I1	I2	Sub.	
M, A	003	Classificação	080	#	#	a	Classificação Decimal Universal: Número de classificação
M, A	003	Classificação	082	0	4	a	Classificação Decimal de Dewey: Número de classificação
M, A	003	Classificação	900	#	#	n	Item: Número de chamada
–	004	Bases de dados	–	–	–	–	–
M, A	005	Tipo de Documento	LDR	–	–	/06	Líder: Tipo de registro
M, A	005	Tipo de Documento* ²⁷	008	–	–	/23	Informações gerais: Forma do item
M, A	006	Nível Bibliográfico*	LDR	–	–	/07	Líder: Nível bibliográfico
M	007	Tombos	900	#	#	a	Item: Número de tombo
M	007	Tombos	900	#	#	b	Item: Exemplar
M	007	Tombos	900	#	#	c	Item: Tipo de aquisição
M	007	Tombos	900	#	#	d	Item: Data da aquisição
M	007	Tombos	900	#	#	e	Item: Status da circulação
M	007	Tombos	900	#	#	f	Item: Prazo excepcional
M, A	008	Meio Eletrônico	856	4	0	u	Localização e acesso eletrônico: <i>Uniform Resource Identifier (URI)</i>
–	009	Separata	–	–	–	–	–
A	010	Autor da parte	100	1	#	a	Ponto de acesso principal - Nome pessoal: Nome pessoal
A	010	Autor da parte	700	1	#	a	Ponto de acesso secundário - Nome pessoal: Nome pessoal
A	011	Autor Coletivo da parte	110	2	#	a	Ponto de acesso principal - Nome corporativo: Nome
A	011	Autor Coletivo da parte	710	2	#	a	Ponto de acesso secundário - Nome corporativo: Nome
A	012	Título e Subtítulo da parte	245	1	0	a	Indicação de título: Título
A	012	Título e Subtítulo da parte	245	1	0	b	Indicação de título: Outras informações sobre o título
A	013	Título traduzido da parte	242	1	0	a	Título traduzido pela agência catalogadora: Título
A	014	Intervalo de Páginas	773	0	8	g	Descrição do todo: Partes relacionadas
M	015	Créditos ou Ementa	508	#	#	a	Nota de créditos de criação/produção: Nota
M	016	Autor	100	1	#	a	Ponto de acesso principal - Nome pessoal: Nome pessoal
M	016	Autor	700	1	#	a	Ponto de acesso secundário - Nome pessoal: Nome pessoal
A	016	Autor	773	0	8	a	Descrição do todo: Ponto de acesso principal
M	017	Autores Coletivos	110	2	#	a	Ponto de acesso principal - Nome corporativo: Nome
M	017	Autores Coletivos	710	2	#	a	Ponto de acesso secundário - Nome corporativo: Nome
A	017	Autores Coletivos	773	0	8	a	Descrição do todo: Ponto de acesso

²⁶ A – Nível analítico.

²⁷ * – Requer mapeamento dos valores.

Nível	Formato PHL		Formato MARC 21 para Dados Bibliográficos				Campo: subcampo
	Tag	Campo	Tag	I1	I2	Sub.	
							principal
M	018	Título	245	1	0	a	Indicação de título: Título
A	018	Título	773	0	8	t	Descrição do todo: Título
M	019	Título traduzido	242	1	0	a	Título traduzido pela agência catalogadora: Título
M	020	Total de Páginas	300	#	#	a	Descrição física: Extensão
M	021	Volume	362	0	#	a	Datas de publicação e/ou designação sequencial: Datas ou designação
A	021	Volume	773	0	8	g	Descrição do todo: Partes relacionadas
M, A	022	Suporte*	008	–	–	/23	Informações gerais: Forma do item
–	026	Tipo de Coleção	–	–	–	–	–
–	027	Total de volumes (nível coleção)	–	–	–	–	–
–	028	Procedência do documento	–	–	–	–	–
–	029	Divulgar	–	–	–	–	–
M	030	Título da Série	490	0	#	a	Indicação de série: Título da série
M	030	Título da Série	245	1	0	a	Indicação de título: Título
A	030	Título da Série	773	0	8	t	Descrição do todo: Título
M	031	Ano e/ou volume	362	0	#	a	Datas de publicação e/ou designação sequencial
A	031	Ano e/ou volume	773	0	8	g	Descrição do todo: Partes relacionadas
M	032	Fascículo	362	0	#	a	Datas de publicação e/ou designação sequencial
A	032	Fascículo	773	0	8	g	Descrição do todo: Partes relacionadas
–	033	N. Kardex	–	–	–	–	–
–	034	Suplemento	–	–	–	–	–
M, A	035	ISSN <i>International Standard Serial Number</i>	022	#	#	a	<i>International Standard Serial Number (ISSN): ISSN</i>
A	035	ISSN International Standard Serial Number	773	0	8	x	Descrição do todo: ISSN
M	038	Informação Descritiva do Suporte	300	#	#	a	Descrição física: Extensão
–	039	Nível de descrição arquivística	–	–	–	–	–
M, A	040	Idiomas do Texto*	008	–	–	/35-37	Informações gerais: Idioma
M, A	040	Idiomas do Texto*	041	#	#	a	Código do idioma: Código do idioma do texto/som ou do título separado
M, A	041	Idioma do Resumo*	008	–	–	/35-37	Informações gerais: Idioma
M, A	041	Idioma do Resumo*	041	#	#	b	Código do idioma: Código do idioma da sumarização ou do resumo
–	043	Povo	–	–	–	–	–
M	050	Instituição da Tese	502	#	#	c	Nota de tese, dissertação: Instituição

Nível	Formato PHL		Formato MARC 21 para Dados Bibliográficos				Campo: subcampo
	Tag	Campo	Tag	I1	I2	Sub.	
M	051	Notas de tese / dissertação	502	#	#	a	Nota de tese, dissertação: Nota
M	052	Instituição patrocinadora do evento	500	#	#	a	Notas gerais: Nota geral
M	053	Nome do evento	111	2	#	a	Ponto de acesso principal - Nome de evento: Nome
M	053	Nome do evento	711	2	#	a	Ponto de acesso secundário - Nome de evento: Nome
A	053	Nome do evento	773	0	8	a	Descrição do todo: Ponto de acesso principal
M	054	Data do Evento	111	2	#	d	Ponto de acesso principal - Nome de evento: Data do evento
M	054	Data do Evento	711	2	#	d	Ponto de acesso secundário - Nome de evento: Data do evento
M	055	Data padronizada do evento	111	2	#	d	Ponto de acesso principal - Nome de evento: Data do evento
M	055	Data padronizada do evento	711	2	#	d	Ponto de acesso secundário - Nome de evento: Data do evento
M	056	Local do Evento	111	2	#	c	Ponto de acesso principal - Nome de evento: Local do evento
M	056	Local do Evento	711	2	#	c	Ponto de acesso secundário - Nome de evento: Local do evento
-	057	País do evento	-	-	-	-	-
M, A	058	Patrocinadores do Projeto	500	#	#	a	Notas gerais: Nota geral
M, A	059	Nome do Projeto de Pesquisa	500	#	#	a	Notas gerais: Nota geral
M, A	060	Número do Projeto de Pesquisa	500	#	#	a	Notas gerais: Nota geral
M, A	061	Notas Gerais	500	#	#	a	Notas gerais: Nota geral
M	062	Editora	260	#	#	b	Publicação, distribuição, etc.: Nome do publicador, distribuidor, etc.
A	062	Editora	773	0	8	d	Descrição do todo: Local de publicação, publicador e data de publicação
M	063	Edição	250	#	#	a	Indicação de edição: Indicação de edição
A	063	Edição	773	0	8	b	Descrição do todo: Edição
M, A	064	Data de Publicação	260	#	#	c	Publicação, distribuição, etc.: Data de publicação, distribuição, etc.
M, A	064	Data de Publicação	008	-	-	/07-10	Informações gerais: Data 1
A	064	Data de Publicação	773	0	8	d	Descrição do todo: Local de publicação, publicador e data de publicação
M, A	065	Data de Publicação Padronizada	260	#	#	c	Publicação, distribuição, etc.: Data de publicação, distribuição, etc.
M, A	065	Data de Publicação Padronizada	008	-	-	/07-10	Informações gerais: Data 1
A	065	Data de Publicação Padronizada	773	0	8	d	Descrição do todo: Local de publicação, publicador e data de publicação
M, A	066	Cidade de Publicação	260	#	#	a	Publicação, distribuição, etc.: Local

Nível	Formato PHL		Formato MARC 21 para Dados Bibliográficos				Campo: subcampo
	Tag	Campo	Tag	I1	I2	Sub.	
							de publicação, distribuição, etc.
A	066	Cidade de Publicação	773	0	8	d	Descrição do todo: Local de publicação, publicador e data de publicação
M, A	067	País de Publicação	260	#	#	a	Publicação, distribuição, etc.: Local de publicação, distribuição, etc.
M, A	067	País de Publicação	008	-	-	/15-17	Informações gerais: Local de publicação, produção ou execução
M	068	Símbolo	500	#	#	a	Notas gerais: Nota geral
M, A	069	ISBN <i>International Standard Book Number</i>	020	#	#	a	<i>International Standard Book Number (ISBN): ISBN</i>
A	069	ISBN International Standard Book Number	773	0	8	z	Descrição do todo: ISBN
M, A	070	Imagem do Objeto	856	4	2	u	Localização e acesso eletrônico: Uniform Resource Identifier (URI)
M, A	071	Tipo de Conteúdo*	008	-	-	/24-27	Informações gerais: Natureza do conteúdo
M, A	071	Tipo de Conteúdo	008	-	-	/29	Informações gerais: Publicação de conferência
M, A	071	Tipo de Conteúdo	008	-	-	/33	Informações gerais: Forma literária
M, A	071	Tipo de Conteúdo	008	-	-	/34	Informações gerais: Biografia
M, A	071	Tipo de Conteúdo	500	#	#	a	Notas gerais: Nota geral
M, A	072	Total de Referências	504	#	#	b	Nota de bibliografia, etc.: Número de referências
M, A	074	Alcance Temporal Desde	648	#	4	a	Ponto de acesso de assunto: Termo cronológico: Termo cronológico
M, A	075	Alcance Temporal Até	648	#	4	a	Ponto de acesso de assunto: Termo cronológico: Termo cronológico
M, A	076	Descriptor Pré codificado	650	#	4	a	Pontos de acesso de assunto - Termo tópico: Termo tópico
M, A	078	Indivíduo como Tema	600	1	4	a	Pontos de acesso de assunto - Nome pessoal: Nome pessoal
M, A	079	Bibliografia de cursos*	526	8	#	a	Nota sobre programa de estudo: Nome do programa
M, A	082	Outras Localidades	651	#	4	a	Pontos de acesso de assunto - Nome geográfico: Nome
M, A	083	Resumo	520	3	#	a	Sumário, etc.: Sumário
M, A	084	Código HTML	887	#	#	a	Campo de informação não MARC: Conteúdo do campo não MARC
M, A	085	Área do Conhecimento	500	#	#	a	Notas gerais: Nota geral
M, A	086	Notas de conteúdo	505	8	0	t	Nota de conteúdos formatados: Nota
M, A	087	Descritores de conteúdo	650	#	4	a	Pontos de acesso de assunto - Termo tópico: Termo tópico
M, A	087	Descritores de conteúdo	650	#	4	x	Pontos de acesso de assunto - Termo tópico: Subdivisão geral
M, A	088	Descritores Secundários	650	#	4	a	Pontos de acesso de assunto - Termo tópico: Termo tópico
M, A	089	Observações do Bibliotecário	590	#	#	a	Nota local: Nota

Nível	Formato PHL		Formato MARC 21 para Dados Bibliográficos				Campo: subcampo
	Tag	Campo	Tag	I1	I2	Sub.	
–	090	Disponibilidade	–	–	–	–	–
–	091	Nível da restrição	–	–	–	–	–
M, A	096	Código HTML restrito	887	#	#	a	Campo de informação não MARC: Conteúdo do campo não MARC
M, A	103	Cutter / PHA	090	#	#	a	Notação de autor: Notação
M, A	103	Cutter / PHA	900	#	#	n	Item: Número de chamada
–	105	Gênero documental	–	–	–	–	–
A	121	Créditos ou Ementa (nível analítico)	508	#	#	a	Nota de créditos de criação/produção: Nota
A	141	Informação Descritiva da parte	300	#	#	a	Descrição física: Extensão
M	181	Subtítulo	245	1	0	b	Indicação de título: Outras informações sobre o título
A	181	Subtítulo	773	0	8	t	Descrição do todo: Título
M	182	Subtítulo da Série	490	0	#	a	Indicação de série: Título da série
M	182	Subtítulo da Série	245	1	0	b	Indicação de título: Outras informações sobre o título
A	182	Subtítulo da Série	773	0	8	t	Descrição do todo: Título
M, A	186	Comentários WIKI Folksonomia	590	#	#	a	Nota local: Nota
A	996	Identificação do Todo	773	0	8	w	Descrição do todo: Número de controle do registro
–	998	Cópia de segurança do MFN do registro	–	–	–	–	–
M, A	999	Datas de Controle	005	–	–	–	Data e hora da última transação
M, A	999	Datas de Controle	008	–	–	/00-05	Informações gerais: Data de criação do registro
M, A	999	Datas de Controle	998	#	#	a	Histórico de modificações do registro: Data da modificação
M, A	999	Datas de Controle	998	#	#	b	Histórico de modificações do registro: Hora da modificação
M, A	999	Datas de Controle	998	#	#	c	Histórico de modificações do registro: Usuário

Fonte: Elaborado pelo autor.

No Quadro 2 são observadas as seguintes situações:

- **correspondência exata** – campos/subcampos com o mesmo significado nos dois padrões: título do recurso (PHL 018 – MARC 21 245\$a), nome do publicador (PHL 062 – MARC 21 260\$b), etc.;
- **correspondência um-para-muitos** – campo do PHL correspondente a mais de um subcampo no MARC 21: pessoa responsável (PHL 016 – MARC 21 100\$a e 700\$a), etc.;

- **correspondência muitos-para-um** – mais de um campo do PHL correspondente a um mesmo subcampo do MARC 21: Número de páginas e Informação descritiva do suporte (PHL 020 e 038 – MARC 21 300\$a), etc.;
- **correspondência provável** – campo do PHL provavelmente correspondente a um subcampo no MARC 21: Número de classificação (PHL 003 – MARC 21 080\$a ou 082\$a), etc.;
- **não correspondência** – campo do PHL sem subcampo correspondente no MARC 21:
 - mapeado para o subcampo de Nota geral (MARC 21 500\$a): Nome do projeto de pesquisa (PHL 059), Área do conhecimento (PHL 085), etc.
 - mapeado para um subcampo local (MARC 21 9xx), quando não pôde ser considerado como nota geral e nem descartado: Tombos (PHL 007 – MARC 21 900\$a\$b\$c\$d\$e\$f); Datas de controle (PHL 999 – MARC 21 998\$a\$b\$c), etc.
 - descartado, quando pôde: Cópia de segurança do MFN do registro (PHL 998), etc.

Cabe observar que, em alguns casos, mais de uma dessas situações ocorreram a um mesmo campo do Formato PHL.

As situações que não possibilitaram definir correspondências e, por isso, os dados dos registros PHL foram descartados, incluem, em sua maior parte, os campos do Formato PHL utilizados para a gestão do recurso informacional ou do registro bibliográfico, além de campos destinados à descrição de coleções e de tipos de recursos informacionais convencionalmente não presentes no ambiente das bibliotecas.

As situações de correspondência um-para-muitos, muitos-para-um e de correspondência provável foram resolvidas por meio de regras de conversão. Na redação das regras de conversão para algumas correspondências muitos-para-um, foi dada a preferência aos dados provenientes de um campo, sendo os demais campos do Formato PHL utilizados apenas nos casos em que o campo cuja prioridade foi dada não estivesse presente no registro. Os campos Data de publicação (PHL 064) e Data de publicação padronizada (PHL 065), por exemplo, foram mapeados para o subcampo \$c do campo Publicação, distribuição, etc. (MARC 21 260). No entanto, foi dada a preferência ao campo 064, uma vez que, de acordo com o Manual do PHL (OLIVEIRA, 2011, p. 89), seu conteúdo mais se assemelha ao conteúdo indicado para o subcampo \$c do campo 260 (LIBRARY OF CONGRESS, 2012a).

Para fins de discussão, nesta seção são realizados apontamentos sobre o mapeamento e o mapa dele resultante.

Em um registro no Formato MARC 21 para Dados Bibliográfico, como descrito na seção 2.1, alguns dados sobre o recurso informacional representado e sobre o próprio registro estão presentes no líder e nos campos de controle (campos 00x). No mapeamento observou-se que nem todos os valores desses campos puderam ser extraídos a partir dos campos de um registro no Formato PHL. Situação semelhante ocorreu aos valores dos indicadores. Em ambos os casos optou-se por (1) utilizar os valores mais prováveis²⁸ ou os mais genéricos como valores padrão (*default*) para os dados obrigatórios e, se possível, para os não obrigatórios, e (2) deixar em branco os espaços destinados aos dados não obrigatórios quando um valor padrão não pôde ser definido.

Em relação às especificidades dos padrões de metadados, tem-se como exemplo as orientações sobre o campo Autor (PHL 016) e aos campos Ponto de acesso principal – Nome pessoal (MARC 21 100) e Ponto de acesso secundário – Nome pessoal (MARC 21 700). Em um registro PHL, o campo 016 é repetível, assim, se o recurso catalogado possui dois autores, o registro possui dois campos 016, como o exemplo da Figura 25.

Figura 25 – Fragmento de um registro no Formato PHL de um recurso com dois autores

016 Pescuma, Derna
016 Castilho, Antonio Paulo Ferreira de, 1953-

Fonte: Elaborado pelo autor.

O campo 016 do Formato PHL não possibilita a individualização dos diferentes dados que podem compor o ponto de acesso que representa uma pessoa. A forma completa do nome, as datas de nascimento e de morte são alguns dos dados frequentemente acrescentados aos pontos de acesso de nomes pessoais para distinguir pessoas com o mesmo nome. Na segunda ocorrência do campo 016 da Figura 25, o ano de nascimento do autor segue seu nome sem que haja um indicativo de separação, ou seja, uma indicação provida pelo padrão de metadados que possibilite a individualização do ano e seu processamento por aplicações de informática.

Um registro bibliográfico no Formato MARC 21 pode conter apenas um campo 100, ao passo que não há limitação da quantidade de campos 700. Se o recurso informacional

²⁸ A definição de quais valores seriam mais prováveis ocorreu a partir da consideração de que, de modo geral, as bibliotecas possuem em seus acervos publicações monográficas textuais impressas.

catalogado possui dois autores, em geral, o autor que for mencionado na fonte de informação principal primeiro terá seu ponto de acesso registrado no campo 100, enquanto que os demais autores serão registrados em campos 700, como exemplifica a Figura 26.

Figura 26 – Fragmento de um registro no Formato MARC 21 para Dados Bibliográficos de um recurso com dois autores

```
100 1 # $aPescuma, Derna
700 1 # $aCastilho, Antonio Paulo Ferreira de,$d1953-
```

Fonte: Elaborado pelo autor.

Diferentemente do campo 016 do Formato PHL, os campos 100 e 700 do Formato MARC 21 para Dados Bibliográficos possuem partes (subcampos) destinadas aos dados que podem ser acrescentados aos pontos de acesso de nome pessoal. No campo 700 da Figura 26, o ano de nascimento do autor está contido no subcampo \$d, separado de seu nome (subcampo \$a), podendo ser identificado e processado por aplicações de informática individualmente.

As orientações sobre essas especificidades do Formato MARC 21 para Dados Bibliográficos podem ser transpostas nas seguintes regras de conversão, que auxiliarão o profissional da Ciência da Computação na elaboração da folha de estilo XSLT:

- para a primeira ocorrência do 016, crie um 100; crie um \$a; inclua o conteúdo do primeiro 016;
- para cada ocorrência do 016 que não seja a primeira, crie um 700; crie um \$a; inclua o conteúdo do 016.

Uma vez que o campo 016 do Formato PHL não possibilita a individualização das datas que estão associadas a uma pessoa, tais datas não podem ser processadas e incluídas nos subcampos adequados de um registro no Formato MARC 21 para Dados Bibliográficos.²⁹ A impossibilidade desse processamento resulta em campos como o mostrado na Figura 27.

Figura 27 – Fragmento de um registro no Formato MARC 21 para Dados Bibliográficos obtido a partir da conversão de um registro no Formato PHL

```
100 1 # $aPescuma, Derna
700 1 # $aCastilho, Antonio Paulo Ferreira de, 1953-
```

²⁹ A identificação das datas dentro do campo 016 é possível por meio da identificação de padrões de caracteres no conteúdo de tal campo. Essa identificação exige, no entanto, complexas regras de transformação, sendo assim, está condicionada aos conhecimentos da XSLT do responsável pela elaboração da folha de estilo de transformação.

Fonte: Elaborado pelo autor.

Os valores de alguns metadados do Formato PHL, embora sejam apresentados ao catalogador e ao usuário final como rótulos textuais, são armazenados e exportados como códigos numéricos, como é o caso dos valores registrados no campo Idioma (PHL 040).

Alguns campos do Formato MARC 21 para Dados Bibliográficos também possuem seus valores codificados para serem processados por agentes não humanos. Um exemplo é o campo 008, que possui, entre seus 40 caracteres, três caracteres (posições 35-37) destinados a representar o idioma do recurso informacional descrito no registro. No entanto, apesar de os dois padrões de metadados codificarem um mesmo tipo de dado, eles os fazem utilizando vocabulários diferentes: o Formato PHL utiliza o código atribuído ao idioma em sua base de dados e o Formato MARC 21 para Dados Bibliográficos utiliza a *MARC Code List for Languages*³⁰ (Lista de códigos MARC para idiomas), provida pela LC.

Assim, além do mapeamento entre os campos e subcampos, é necessário o mapeamento dos valores definidos nos diferentes vocabulários. O Quadro 3 apresenta o mapa resultante do mapeamento dos vocabulários utilizados para representar o idioma nos padrões de metadados de origem e de destino.

Quadro 3 – Mapa dos esquemas de codificação utilizados para representar idiomas

<i>Formato PHL: códigos atribuídos na base de dados³¹</i>		<i>Formato MARC 21 para Dados Bibliográficos: Lista de códigos MARC para idiomas</i>	
Código	Idioma	Código	Idioma
1	Português	por	Portuguese
2	Espanhol	spa	Spanish
3	Inglês	eng	English
4	Italiano	ita	Italian
5	Francês	fre	French
6	Alemão	ger	German
7	Japonês	jpn	Japanese
8	Latim	lat	Latin
9	Outro	###	No information provided

Fonte: Elaborado pelo autor.

Uma vez que o Formato PHL utiliza os códigos atribuídos na base de dados e não uma lista padronizada externa a ela, o mapeamento apresentado no Quadro 3 precisa ser revisto pela instituição na conversão dos registros.

³⁰ Disponível em: <<http://www.loc.gov/marc/languages>>. Acesso em: 18 fev. 2013.

³¹ Os idiomas indicados no Quadro são aqueles que, por padrão, são inseridos na base de dados durante a instalação do PHL.

Além do idioma, outros dados são codificados de forma processável por agentes não humanos nos dois padrões de metadados. Os campos que requerem o mapeamento entre vocabulários estão identificados no Quadro 2 com um asterisco (*).

Algumas instruções providas na documentação do Formato PHL se sobrepõem às providas pelo AACR2r. Um exemplo dessa sobreposição é a instrução presente no campo Título da série (PHL 030) que traz “utiliza-se também este campo para registro do título de uma série monográfica ou coleção, seguido do número do volume correspondente, separado por vírgula” (OLIVEIRA, 2011, p. 83), enquanto que, segundo o AACR2r, a indicação de volume deve ser precedida por espaço, ponto e vírgula e espaço (CÓDIGO..., 2004, regra 1.6A.1). Situações como essa devem ser expressas nas regras de conversão. Um exemplo de regra de conversão para essa situação é: *se presente o 030, crie um 490; crie um \$a; inclua o conteúdo do 030 presente antes da vírgula; inclua um espaço e um ponto e vírgula; crie um \$v; inclua o conteúdo do 030 presente depois da vírgula.*

5.3 Elaboração da folha de estilo XSLT

Com base no mapa apresentado no Quadro 2 e nas regras de conversão foi elaborada uma folha de estilo XSLT para a transformação de documentos XML contendo registros no Formato PHL em documentos XML contendo os registros no Formato MARC 21 para Dados Bibliográficos.

Para a elaboração da folha de estilo foi utilizada a versão 2.0 da XSLT e o *Oxygen XML Editor*. Os testes de conversão foram realizados no *Oxygen XML Editor* e no *MarcEdit*.

O primeiro passo para a elaboração da folha de estilo foi definir como os dados do documento XML de origem (PHLXML) seriam organizados nos elementos do documento XML de destino (MARCXML). Para isso foi realizada uma comparação entre a organização dos elementos nesses dois documentos.

No PHLXML, o elemento do documento (elemento que compreende todo o conteúdo do documento) é o *db*. Dentro de um *db* há um elemento *rec* para cada registro. No MARCXML, o elemento do documento é o *collection*, dentro do qual há um elemento *record* para cada registro. A partir dessa comparação inicial, foi indicado na folha de estilo que, para cada elemento *rec* do documento de origem, seria criado um elemento *record* no documento de destino. Essa indicação foi realizada com as regras de transformação apresentadas na Figura 28.

Figura 28 – Código inicial da folha de estilo

```

<xsl:template match="/">
  <marc:collection
    xsi:schemaLocation="http://www.loc.gov/MARC21/slim
    http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd"
    xmlns:marc="http://www.loc.gov/MARC21/slim"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" >
    <xsl:for-each select="db/rec">
      <marc:record>
        <xsl:call-template name="bibliographicRecord"/>
      </marc:record>
    </xsl:for-each>
  </marc:collection>
</xsl:template>

```

Fonte: Elaborada pelo autor.

As regras de transformação da Figura 28 indicam ao processador de transformação: inicie a transformação pelo nó raiz do documento; crie um elemento *marc:collection* (o namespace *marc* representa o MARCXML); dentro do *collection*, para cada elemento *rec* que estiver dentro do elemento *db*, crie um elemento *marc:record*; para compor o conteúdo do *marc:record*, utilize o *template bibliographicRecord*.

O *template bibliographicRecord* compreende um conjunto de regras de transformação criado a partir do mapa e das regras de conversão. O início desse *template* e algumas de suas regras de transformação são indicados na folha de estilo pelo código apresentado na Figura 29.

Figura 29 – O *template bibliographicRecord* e algumas regras de transformação

```

<xsl:template name="bibliographicRecord">
  <!-- Indicação de edição (MARC 250) – Edição (PHL 063) -->
  <xsl:if test="$bibliographicLevel = 'm' and v063">
    <marc:datafield tag="250" ind1=" " ind2=" ">
      <marc:subfield code="a">
        <xsl:value-of select="v063"/>
      </marc:subfield>
    </marc:datafield>
  </xsl:if>
  <!-- Pontos de acesso secundários - Nome pessoal (MARC 700) – Autor (PHL 016) e Autor da
  parte (PHL 010) -->
  <xsl:choose>

```

```

<xsl:when test="$bibliographicLevel = 'm'">
  <xsl:for-each select="v016[position()>1]">
    <marc:datafield tag="700" ind1="1" ind2=" ">
      <marc:subfield code="a">
        <xsl:value-of select="."/>
      </marc:subfield>
    </marc:datafield>
  </xsl:for-each>
</xsl:when>

<xsl:when test="$bibliographicLevel = 'a'">
  <xsl:for-each select="v010[position()>1]">
    <marc:datafield tag="700" ind1="1" ind2=" ">
      <marc:subfield code="a">
        <xsl:value-of select="."/>
      </marc:subfield>
    </marc:datafield>
  </xsl:for-each>
</xsl:when>

</xsl:choose>

</xsl:template>

```

Fonte: Elaborada pelo autor.

A primeira regra de transformação da Figura 29 indica ao processador: se o nível bibliográfico do registro PHL é monográfico (*\$bibliographicLevel = 'm'*) e existir o campo 063, crie o campo 250 do MARC 21 com os indicadores em branco; crie dentro desse campo o subcampo \$a e insira nele o conteúdo do campo 063 do PHL.

A segunda regra indica a partir de quais elementos do documento de origem devem ser criados os pontos de acesso secundários de nome pessoal:

- quando o nível é monográfico, para cada campo 016 do registro PHL que não seja o primeiro (ou seja, a segunda, a terceira, a quarta, etc. ocorrência do 016), crie um campo 700 com o primeiro indicador com valor “1” e o segundo indicador com valor em branco; crie o subcampo \$a e insira nele conteúdo do 016;
- quando o nível é analítico, para cada campo 010 do registro PHL que não seja o primeiro, crie um campo 700 com o primeiro indicador com valor “1” e o segundo indicador com valor em branco; crie o subcampo \$a e insira nele conteúdo do 010.

Essas são duas regras de transformação simples da folha de estilo. Outras regras, principalmente aquelas destinadas à criação de campos com múltiplos dados, por exemplo, o campo 008, requerem a utilização de outros procedimentos de transformação, tais como a

inclusão, a junção, a separação, a edição e/ou a criação de dados com base em condições e em diferentes campos dos registros PHL.

5.4 Verificação da folha de estilo

Durante e após a elaboração da folha de estilo foram realizadas verificações para assegurar a adequação dos registros convertidos ao MARCXML, ao Formato MARC 21 para Dados Bibliográficos e aos instrumentos de descrição (regras de catalogação, vocabulários e convenções).

A verificação sintática, que detém-se à adequação dos registros ao MARCXML, poderia ser realizada por qualquer aplicação de informática que comparasse o documento XML contendo os registros com o esquema MARCXML e identificasse se tal documento está de acordo com o esquema em questão. Desse modo, a verificação sintática foi realizada pelo *Oxygen XML Editor* imediatamente após a conversão.

Uma vez realizada essa verificação sintática e constatado que os registros convertidos estavam de acordo com o MARCXML, deu-se prosseguimento à verificação do padrão de metadados e a verificação do conteúdo, ambas realizadas por um agente humano.

Para a verificação do padrão de metadados foi consultada a documentação oficial do Formato MARC 21 para Dados Bibliográficos (LIBRARY OF CONGRESS, 2012a) e realizado o seguinte questionamento: “as etiquetas dos campos, os valores dos indicadores, os códigos dos subcampos e as posições de caracteres estão de acordo com o Formato?”.

Para a verificação do conteúdo foi realizado o mesmo questionamento, no entanto, voltado principalmente às regras do AACR2r: “as áreas de descrição, os elementos de descrição e os sinais de pontuação estão de acordo com o AACR2r?”. A verificação do padrão de metadados e a verificação do conteúdo foram realizadas diversas vezes até a obtenção dos resultados esperados (registros de acordo com o mapa, com o padrão de metadados, com as regras de catalogação, com os vocabulários e com as convenções).

5.5 Exportação e conversão dos registros

As instruções para a exportação dos registros do PHL e para sua conversão estão presentes em um manual elaborado durante esta pesquisa (ASSUMPCÃO, 2013). Esta seção apresenta uma síntese das instruções do manual e algumas questões relacionadas à exportação e à conversão dos registros no Formato PHL.

Primeiramente, devido a uma falha observada no PHL, o documento contendo os registros exportados, em alguns casos, pode não ser um documento XML bem formado. Como mencionado na seção 3.1, o caractere & não pode aparecer no texto de um documento XML, pois seu uso é reservado à codificação, sendo que, se utilizado, entrará em conflito com os delimitadores da codificação, tornando o documento mal formado e, portanto, não processável por um analisador XML. No entanto, quando tal caractere é necessário dentro de um texto, há uma forma de indicá-lo sem que ocorra o conflito com a codificação: substituindo o & por & (RAY, 2003). A falha observada no PHL é que, no documento XML resultante da exportação, o caractere & não é indicado da forma correta: é utilizado o & ao invés de &, o que torna o documento mal formado. No contexto dos dados catalográficos, o & é utilizado em alguns casos nos nomes de publicadores, como é o caso da editora L&PM. Assim, antes de submeter os registros exportados pelo PHL à conversão, é necessário verificar se há alguma ocorrência do & e, se houver, substituir tal caractere por &.

Após essa verificação, a conversão dos registros pode ser realizada por qualquer processador de transformação que aceite folhas de estilo elaboradas com a XSLT 2.0. Diante das possibilidades, optou-se pela utilização do MarcEdit, que dispõem, além de ferramentas para edição de registros nos Formatos MARC, de um processador de transformação que satisfaz os requisitos necessários à conversão em questão. A partir dessa escolha, os seguintes passos foram delineados e descritos no manual:

- exportação dos registros do PHL;
- instalação do MarcEdit;
- inserção da folha de estilo XSLT no MarcEdit;
- conversão de PHL para MARC 21; e
- conversão de MARCXML para ISO 2709 (opcional).

O manual que descreve os passos para a conversão utilizando a folha de estilo está disponível no Apêndice A e em <<http://docs.fabricioassumpcao.com/conversao-de-phl-para-marc-21-v-1.pdf>>. A folha de estilo, elaborada como descrito na seção 5.3, verificada conforme a seção 5.4 e utilizada em conjunto com o manual, está disponível em <<http://docs.fabricioassumpcao.com/phl2marc21-1-0.xsl>>. Tanto o manual quanto a folha de estilo estão disponíveis gratuitamente.

A folha de estilo reflete as correspondências entre o Formato PHL e o Formato MARC 21 para Dados Bibliográficos indicadas no mapa do Quadro 2, elaborado a partir do

mapeamento para uma situação geral e não para as necessidades de uma biblioteca específica. Desse modo, as bibliotecas que optarem pela utilização dessa folha de estilo devem realizar as adequações necessárias para que os registros resultantes da conversão estejam de acordo com suas próprias convenções, decisões expressas em políticas e em manuais de catalogação.

Um exemplo dessas adequações diz respeito ao sistema de classificação bibliográfica utilizado pela biblioteca. O campo Classificação (PHL 003) corresponde aos campos Classificação Decimal Universal (MARC 21 080) e Classificação Decimal de Dewey (MARC 21 082), entre outros. Na folha de estilo estão presentes duas regras de transformação: a primeira direciona o conteúdo do 003 para o 080 e a segunda direciona o 003 para o 082. A primeira regra, no entanto, está desativada, de modo que os registros MARC 21 resultantes da conversão não contenham o campo 080. Assim, se a biblioteca utiliza a Classificação Decimal Universal, deve ativar a primeira regra de transformação e desativar a segunda, fazendo com que o campo 080 esteja presente nos registros convertidos.

Para a desativação de uma regra é necessário colocá-la entre as *tags* `<!--` e `-->`. Essas *tags* são utilizadas em um documento XML para indicar o início e o término de um comentário, portanto, o conteúdo entre elas não é interpretado pelo processador de transformação. Para a ativação de uma regra, basta remover as *tags* que indicam o início e o término do comentário.

Após a apresentação do modelo para a conversão de registros, realizada no Capítulo 4, neste capítulo foi descrita sua aplicação na conversão dos registros bibliográficos do sistema de gerenciamento de bibliotecas PHL. Desta aplicação, realizada para a validação do modelo, resultaram dois instrumentos: uma folha de estilo e um manual para sua utilização, que estão disponíveis para as bibliotecas usuárias do PHL que desejam converter seus registros bibliográficos. No capítulo seguinte são apresentadas considerações finais sobre o modelo, sobre a utilização de folhas de estilo na conversão de registros e sobre os resultados da aplicação do modelo na conversão dos registros bibliográficos do PHL.

6 Considerações finais

A automação de bibliotecas, iniciada principalmente a partir da década de 1960, trouxe a necessidade de disponibilizar os dados resultantes da catalogação, até então armazenados e processados em ambientes analógicos, em ambientes digitais. Para disponibilizar os dados nos novos ambientes foram necessárias estruturas que tornassem os dados processáveis por aplicações de informática. Surgiram, então, os Formatos MARC, os formatos deles derivados (MARC II, MARC 21, UNIMARC, etc.) e as codificações.

As possibilidades de codificação de registros nos Formatos MARC 21, acompanhando os avanços da Ciência da Computação, evoluíram desde uma estrutura sequencial, destinada ao armazenamento dos dados em fitas magnéticas, até a XML, um conjunto de regras para a marcação de documentos desenvolvido com foco na Web. Do encontro dos Formatos MARC 21 com a XML surgiu o MARCXML, uma linguagem de marcação que, apesar das críticas, tem possibilitado a codificação de registros de modo mais condizente com as tecnologias de informática da atualidade.

Em posse de padrões de metadados como os Formatos MARC 21, a necessidade de converter os registros analógicos, em sua maior parte em fichas catalográficas, em registros digitais deu origem aos projetos de conversão retrospectiva, para os quais diversas tecnologias foram e têm sido utilizadas.

Com a utilização cada vez mais crescente de aplicações de informática para a representação, a recuperação e a disseminação de recursos informacionais, surge a necessidade de converter registros digitais criados com diferentes padrões de metadados, visando, entre outros, o intercâmbio de dados entre instituições, por exemplo, em programas de catalogação cooperativa, e a migração dos dados entre aplicações de informática. Originou-se, assim, o que a literatura denomina reutilização de metadados.

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 bibliotecas usuárias desses sistemas de gerenciamento, 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, os objetivos desta pesquisa sobre reutilização de metadados surgiram a partir do problema “como converter os registros exportados por esses sistemas de gerenciamento de bibliotecas em registros nos Formatos MARC 21?”.

O primeiro e o segundo objetivos específicos, “realizar estudo teórico sobre a conversão de registros, os Formatos MARC 21, a codificação com a ISO 2709, a XML, a XSLT e o MARCXML” e “destacar os principais aspectos da utilização da XML na codificação de documentos e as possibilidades oferecidas pela XSLT na transformação de documentos XML”, foram alcançados nos capítulos 2 e 3.

O terceiro objetivo específico, “elaborar um modelo para a conversão de registros, definindo suas etapas e seus componentes”, foi alcançado com a apresentação do modelo na seção 4.1 e de suas etapas e componentes nas seções de 4.2 a 4.6. Sobre o modelo proposto e sobre sua utilização, são traçadas as seguintes considerações:

- o modelo proposto busca apresentar uma generalização e uma formalização dos processos e dos componentes necessários à conversão de registros 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 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;
- 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 tal folha;

- 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; e
- 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 bibliotecas em seus planos para conversão.

Cabe ressaltar que esta pesquisa 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 podem ser úteis às suas atividades.

O quarto objetivo específico, “aplicar o modelo na conversão de registros bibliográficos exportados pelo sistema de gerenciamento de bibliotecas *Personal Home Library* (PHL)”, foi alcançado no capítulo 5. Sobre essa aplicação do modelo, são realizadas as seguintes considerações:

- os resultados obtidos com a aplicação do modelo foram registros no Formato MARC 21 para Dados Bibliográficos em conformidade com as especificações desse padrão e com as regras de catalogação do AACR2r;
- a folha de estilo XSLT e as instruções para sua utilização, instrumentos resultantes da aplicação do modelo, oferecem às bibliotecas usuárias do PHL uma alternativa viável e gratuita³² para a conversão de seus registros bibliográficos;
- sem esses instrumentos ou quaisquer outros que desempenhassem sua função, a conversão estaria limitada e condicionada às tarefas de digitação ou de cópia dos dados para planilhas MARC 21, tarefas essas que, dependendo do número de registros a ser convertido e da disponibilidade de profissionais qualificados, demandariam muito tempo;
- as bibliotecas que desejarem converter seus registros do Formato PHL utilizando a folha de estilo devem realizar modificações em tal folha de modo que os registros resultantes da conversão melhor se adequem as suas necessidades locais.

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

³² A folha de estilo e as instruções para sua utilização estão disponíveis gratuitamente em <<http://fabricioassumpcao.com/conversao-de-phl-para-marc-21>>. Acesso em: 09 ago. 2013.

entre sistemas de gerenciamento de bibliotecas que não utilizam padrões de metadados internacionalmente aceitos.

Partindo dessa proposição, 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 de metadados 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 por sistemas de gerenciamento de bibliotecas 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 continuidade desta pesquisa e da aplicação do modelo em outras comunidades além da comunidade de bibliotecas.

Nota-se também que a XML é uma das codificações que podem ser utilizadas com os registros no domínio bibliográfico, sendo necessários estudos sobre outras codificações e sobre as possibilidades para a conversão trazidas por elas. Outras possibilidades de estudo também estendem-se em direção à conversão de registros MARC 21 visando o uso de tecnologias da Web Semântica para a representação, o intercâmbio e o acesso aos dados sobre recursos informacionais.

Por fim, destaca-se nessas considerações finais que, independentemente da tecnologia utilizada, a conversão de registros no contexto das bibliotecas significa evitar o retrabalho, ou seja, evitar que um recurso informacional já catalogado precise ser catalogado novamente. Evitar o retrabalho, por conseguinte, implica redução dos custos com a catalogação e do tempo gasto na disponibilização dos registros aos usuários, que, por sua vez, têm o catálogo

como um dos principais meios de acesso à informação contida nos recursos informacionais que integram as coleções das bibliotecas.

Referências

ALMEIDA, M. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 5-13, maio/ago. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12903>>. Acesso em: 29 jan. 2013.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

ALVES, R. C. V.; SANTOS, P. L. V. A. da C. Metadados em ciência da informação: considerações preliminares sobre padrões para a construção normalizada de representações. In: CONGRESSO DE LA CIBERSOCIEDAD: Crisis analógica, futuro digital, 4., 2009. **Anais...** [S.l.: s.n.], 2009. Disponível em: <<http://www.cibersociedad.net/congres2009/es/coms/metadados-em-cincia-da-informacao-consideracoes-preliminares-sobre-padroes-para-a-construcao-normalizada-de-representacoes/994/>>. Acesso em: 14 out. 2012.

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 (ENACAT), 1., 2012, Rio de Janeiro; ENCONTRO DE ESTUDOS E PESQUISAS EM CATALOGAÇÃO (EEPC), 3., 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: 18 fev. 2013.

ASENSI ARTIGA, V.; RODRÍGUEZ MUÑOZ, J. V. El proceso de catalogación automatizada. In: PINTO MOLINA, María (Org.). **Catalogación de documentos: teoría y práctica**. 2. ed. rev. e atual. Madrid: Síntesis, 2001. p. 105-143.

ASSUMPÇÃO, F. S. **Conversão de registros bibliográficos do PHL para registros em MARC 21**. Versão 1.0. Marília, 2013. Disponível em: <<http://docs.fabricioassumpcao.com/conversao-de-phl-para-marc-21-v-1.pdf>>. Acesso em: 31 mar. 2013.

ASSUMPÇÃO, F. S. et al. **Catalogação descritiva e metadados no Brasil: uma análise bibliométrica (1997-2011)**. Marília, 2013. Não publicado.

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: 18 fev. 2013.

AVRAM, H. D. Machine-Readable Cataloging (MARC) Program. In: ENCYCLOPEDIA of Library and Information Science. New York: Marcel Dekker, 2003. p. 1712-1730. Disponível em: <http://polaris.gseis.ucla.edu/gleazer/260_readings/Avram.pdf>. Acesso em: 14 out. 2012.

AVRAM, H. D. **The MARC Pilot Project**: Final Report on a Project Sponsored by The Council on Library Resources, Inc. Washington, D.C.: Library of Congress, 1968. Disponível em: <<http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED029663>>. Acesso em: 21 abr. 2012.

BALBY, C. N. Formatos de intercâmbio de registros bibliográficos: conceitos básicos. **Cadernos da F.F.C.**, Marília, v. 4, n. 1, p. 29-35, 1995.

BARBOSA, A. P. **Novos rumos da catalogação**. Rio de Janeiro: BNG/BRASILART, 1978.

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: 14 fev. 2013.

BORKO, H. Information Science: what is it? **American Documentation**, Washington, D.C., 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: 25 fev. 2013.

BRADLEY, N. **The XML Companion**. 3rd ed. Boston: Addison-Wesley Professional, 2001.

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.

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: 18 fev. 2013.

CÓDIGO de catalogação anglo-americano. 2. ed., rev. 2002. São Paulo: FEBAB, 2004.

CÔRTE, A. R. e et al. Automação de bibliotecas e centros de documentação: o processo de avaliação e seleção de softwares. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 241-256, set./dez. 1999.

COYLE, K. Future considerations: the functional library systems record. **Library Hi Tech**, n. 22, v. 2, p. 166-174, 2004.

DANSKIN, A. Mature consideration: developing bibliographic standards and maintaining values. **New Library World**, n. 105, v. 3/4, p. 113-117, 2004.

DARKO-AMPEM, K. Retrospective conversion of serials and card catalogue records: A case study of project management in academic libraries. **Library Management**, v. 27, n. 3, p. 121-134, 2006. Disponível em: <<http://dx.doi.org/10.1108/01435120610652879>>. Acesso em: 28 set. 2012.

DEMO, P. **Metodologia científica em ciências sociais**. 3. ed. rev. e aum. São Paulo: Atlas, 1995.

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.

ESPECIFICAÇÕES. Gurupi: InfoArte, [2009?]. Disponível em: <http://www.elysio.com.br/site/especificacoes.html>>. Acesso em: 05 jan. 2013.

ETUBI, M. Retrospective Conversion of Arabic Collections at a Nigerian University Library. **African Journal of Library, Archives and Information Science**, v. 19, n. 2, p. 177-181, 2009.

FLAMINO, A. N. **MARCXML**: um padrão de descrição para recursos informacionais em Open Archives. 2006. 164 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista - UNESP, Marília. 2006. Disponível em: http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/flamino_an_me_mar.pdf>. Acesso em: 18 fev. 2013.

FUSCO, E. **Aplicação dos FRBR na modelagem de catálogos bibliográficos digitais**. São Paulo: Cultura Acadêmica, 2011.

GARRIDO ARILLA, M. R. **Teoría e historia de la catalogación de documentos**. Madrid: Síntesis, 1996.

GORMAN, M. What is the future of cataloguing and cataloguers? In: IFLA GENERAL CONFERENCE, 63., 1997, Copenhagen. **Conference Programme and Proceedings**. [S.l.]: IFLA, 1997.

HILLMANN, D. et al. RDA Vocabularies: Process, Outcome, Use. **D-Lib Magazine**, v. 16, n. 1/2, Jan./Feb. 2010. Disponível em: <http://dlib.org/dlib/january10/hillmann/01hillmann.html>>. Acesso em: 08 nov. 2012.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Statement of International Cataloguing Principles**. [S.l.]: IFLA, 2009.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 2709**. [S.l.], [2011?]. Norma ISO-2709 em Português. Disponível em: <http://wiki.bireme.org/pt/index.php/Arquivo:2709BR.pdf>>. Acesso em: 24 out. 2012.

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: 24 out. 2012.

JAKŠIĆ, M. Mapping of bibliographical standards into XML. **Software: Practice and Experience**, v. 34, n. 11, p. 1051-1064, 2004. Disponível em: <<http://dx.doi.org/10.1002/spe.603>>. Acesso em: 11 set. 2012.

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: 18 fev. 2013.

KEITH, C. Using XSLT to manipulate MARC metadata. **Library Hi Tech**, v. 22, n. 2, p. 122-130, 2004.

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.

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

LIBRARY OF CONGRESS. **Bibliographic Framework Initiative General Plan**. Washington, D.C., 2011. Disponível em: <<http://www.loc.gov/marc/transition/pdf/bibframework-10312011.pdf>>. Acesso em: 29 set. 2012.

LIBRARY OF CONGRESS. **Frequently Asked Questions (FAQ)**. Washington, D.C., 2006. Disponível em: <<http://www.loc.gov/marc/faq.html#definition>>. Acesso em: 29 set. 2012.

LIBRARY OF CONGRESS. **MARC 21 Format for Bibliographic Data**. Washington, D.C., 2012a. 1999 Edition. Update No. 1 (October 2001) through Update No. 15 (September 2012). Disponível em: <<http://www.loc.gov/marc/bibliographic/ecbdhome.html>>. Acesso em: 1 dez. 2012.

LIBRARY OF CONGRESS. **MARC 21 XML Schema**: official web site. Washington, D.C., 2011b. Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 1 dez. 2011.

LIBRARY OF CONGRESS. **MARC 21**: Harmonized USMARC and CAN/MARC. Washington, D.C., 1998. Disponível em: <<http://www.loc.gov/marc/annmarc21.html>>. Acesso em: 29 set. 2012.

LIBRARY OF CONGRESS. **MARC XML Design Considerations**. Washington, D.C., 2004. Disponível em: <<http://www.loc.gov/standards/marcxml/marcxml-design.html>>. Acesso em: 14 fev. 2013.

LIBRARY OF CONGRESS. **Metadata Object Description Schema**. Washington, 2013. Disponível em: <<http://www.loc.gov/standards/mods/>>. Acesso em: 16 abr. 2013.

LIBRARY OF CONGRESS. **RDA in MARC**. Washington, 2012b. Disponível em: <<http://www.loc.gov/marc/RDAinMARC.html>>. Acesso em: 25 set. 2013.

LIBRARY OF CONGRESS. **Schema MARC21slim.xsd**. Washington, D.C., 2009. Disponível em: <<http://www.loc.gov/standards/marcxml/xml/spy/spy.html>>. Acesso em: 30 dez. 2012.

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: 14 out. 2012.

MACEDO, N. D. **Iniciação à pesquisa bibliográfica**: guia do estudante para a fundamentação do trabalho de pesquisa. 2. ed. rev. São Paulo: Edições Loyola, 1996.

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

MEY, E. S. A. **Introdução à catalogação**. Brasília: Briquet de Lemos, 1995.

MILLER, D. R.; CLARKE, K. S. **Putting XML to work in the library**: tools for improving access and management. Chicago: American Library Association, 2004.

MÖNCH, C.; AALBERG, T. Automatic Conversion from MARC to FRBR. **Lecture Notes in Computer Science**, v. 2769, p. 405-411, 2003. Disponível em: <http://dx.doi.org/10.1007/978-3-540-45175-4_37>. Acesso em: 14 out. 2012.

MORENO, F. P.; BRASCHER, M. MARC, MARCXML e FRBR: relações encontradas na literatura. **Informação & Sociedade**: estudos, João Pessoa, v. 17, n. 3, p. 13-25, set./dez. 2007. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/834>>. Acesso em: 14 out. 2012.

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: 14 fev. 2013.

NOSSOS clientes. Gurupi: InfoArte, 2013. Disponível em: <<http://www.elysio.com.br>>. Acesso em: 18 fev. 2013.

OLIVEIRA NETTO, A. A. **Metodologia da pesquisa científica**: guia prático para a apresentação de trabalhos acadêmicos. 3. rev. atual. Florianópolis: Visual Books, 2008.

OLIVEIRA, E. M. S. de. **Manual do PHL 8.2**. Revisão de 27/09/2011. Gurupi: InfoArte, 2011.

OLIVEIRA, E. M. S. de. **Sobre o PHL@Elysio**. Gurupi: InfoArte, [2002?]. Disponível em: <<http://www.elysio.com.br/site/phl.html>>. Acesso em: 13 dez. 2012.

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**, Belo Horizonte, v. 3, n. 1, p. 41-46, jan./jun. 1998.

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**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, n. esp., p. 14-32, 2º sem. 2004.

PEREIRA, A. M.; SANTOS, P. L. V. A. da C. O uso estratégico das tecnologias em catalogação. **Cadernos da F.F.C.**, Marília, 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: 18 fev. 2013.

PHL.NETOPAC. Gurupi: InfoArte, 2013. Disponível em: <http://www.elysio.com.br/site/phlnet_index.html>. Acesso em: 18 fev. 2013.

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: 14 out. 2012.

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**, Brasília, v. 13, n. 1, p. 21-26, jan./jun. 1985.

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

RESOURCE Description and Access (RDA). Chicago: American Library Association, 2013.

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: 05 dez. 2012.

SANTOS, P. L. V. A. da C.; ALVES, R. C. V. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. **DataGramZero: Revista de Ciência da Informação**, Rio de Janeiro, v. 10, n. 6, dez. 2009. Disponível em: <http://www.dgz.org.br/dez09/Art_04.htm>. Acesso em: 29 jan. 2013.

SANTOS, P. L. V. A. da C.; CORRÊA, R. M. R. **Catalogação: trajetória para um código internacional**. Niterói: Intertexto, 2009.

SANTOS, P. L. V. A. da C.; FLAMINO, A. N. MARC 21 e XML como ferramentas para a consolidação da catalogação cooperativa automatizada: uma revisão de literatura. In: VIDOTTI, S. A. B. G. (Org.). **Tecnologia e conteúdos informacionais: abordagens teóricas e práticas**. São Paulo: Polis, 2004. p. 113-138.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, 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: 21 jan. 2013.

SCHMIDT, N.; PATEL, A. MARC record conversion: a generalised approach. **Computer Standards & Interfaces**, v. 21, p. 287-297, 1999.

SIQUEIRA, M. A. **XML na Ciência da Informação**: uma análise do MARC 21. 2003. 133 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003. Disponível em: <http://www.athena.biblioteca.unesp.br/exlibris/bd/bma/33004110043P4/2003/siqueira_ma_me_mar.pdf>. Acesso em: 14 out. 2012.

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: 19 fev. 2013.

TAYLOR, A. G.; JOUDERY, D. N. **The organization of information**. 3rd ed. Westport: Libraries Unlimited, 2009.

TAYLOR, M.; DICKMEISS, A. Delivering MARC/XML records from the Library of Congress catalogue using the open protocols SRW/U and Z39.50. In: **WORLD LIBRARY AND INFORMATION CONGRESS**, 71th, 2005, Oslo. **Proceedings...** Oslo: IFLA, 2005.

TENNANT, R. A bibliographic metadata infrastructure for the twenty-first century. **Library Hi Tech**, v. 22, n. 2, p. 175-181, 2004.

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

W3C. **XML Path Language (XPath) 2.0 (Second Edition)**: W3C Recommendation 14 December 2010. Cambridge, 2010. Disponível em: <<http://www.w3.org/TR/xpath20>>. Acesso em: 20 jan. 2013.

W3C. **XML Schema**. Revision 1.153, date: 2012/06/22. Cambridge, 2012a. Disponível em: <<http://www.w3.org/XML/Schema>>. Acesso em: 09 out. 2012.

W3C. **XML**. Last modified: 2012/01/24. Cambridge, 2012b. Disponível em: <<http://www.w3.org/XML>>. Acesso em: 03 dez. 2012.

W3C. **XSL Transformations (XSLT) Version 1.0**: W3C Recommendation 16 November 1999. Cambridge, 1999. Disponível em: <<http://www.w3.org/TR/1999/REC-xslt-19991116>>. Acesso em: 20 jan. 2013.

W3C. **XSL Transformations (XSLT) Version 2.0**: W3C Recommendation 23 January 2007. Cambridge, 2007. Disponível em: <<http://www.w3.org/TR/xslt20>>. Acesso em: 25 set. 2012.

W3C. **XSL Transformations (XSLT) Version 3.0**: W3C Working Draft 10 July 2012. Cambridge, 2012. Disponível em: <<http://www.w3.org/TR/xslt-30>>. Acesso em: 20 jan. 2013.

WALMSLEY, P. **FunctX XSLT Functions**. Traverse City: Datypic, 2007. Disponível em: <<http://www.xsltfunctions.com>>. Acesso em: 20 jan. 2013.

WOODLEY, M. S. Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information. In: BACA, Murtha (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: 18 fev. 2013.

YEE, M. M. New perspectives on the shared cataloging environment and a MARC 21 shopping list. **Library Resources & Technical Services**, v. 48, n. 3, p. 165-178, 2004.

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://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Zafalon,%20Z.R. doutorado C.I. 2012.pdf](http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Zafalon,%20Z.R.%20doutorado%20C.I.%202012.pdf)>. Acesso em: 18 fev. 2013.

**APÊNDICE A – Manual para a conversão dos registros
bibliográficos do PHL**

Fabício Silva Assumpção

Conversão de registros bibliográficos do PHL para
registros no Formato MARC 21 para Dados
Bibliográficos

Versão 1.1
1 de agosto de 2013

Marília, 2013

Fabrício Silva Assumpção

assumpcao.f(at)gmail.com

fabricioassumpcao.com



Esta obra foi licenciado sob uma Licença [Creative Commons Atribuição 3.0 Não Adaptada](https://creativecommons.org/licenses/by/3.0/).

Este manual está disponível em: <http://docs.fabricioassumpcao.com/conversao-de-phl-para-marc-21-v-1-1.pdf>

Apresentação

Este manual descreve os passos necessários para a conversão de registros bibliográficos exportados pelo *Personal Home Library* (PHL) em registros no Formato MARC 21 para Dados Bibliográficos.

Este manual e a folha de estilo XSLT nele mencionada foram elaborados como parte da dissertação **“Conversão de registros em XML para em MARC 21: um modelo baseado em folhas de estilo XSLT”**, escrita por Fabrício Silva Assumpção sob a orientação da Prof.^a Dr.^a Plácida Leopoldina Ventura Amorim da Costa Santos, com o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a sob os auspícios do Programa de Pós-Graduação em Ciência da Informação da Unesp, Campus de Marília.

1 Exportação dos registros do PHL

- a) Acesse o PHL com uma conta de usuário que possua privilégios de administrador.
- b) Clique em Exportação.
- c) No campo Base de dados selecione Catálogo.
- d) No campo Padrão selecione XML.
- e) Clique em Confirmar.

PHL © Elysis - Nome x

127.0.0.1:2185/cgi-bin/wxis.exe?IsisScript=phl82.x

Buscas | Serviços / Renovações / Reservas | Gráficos | Manual | phl.NET | phl.NEWS

Biblioteca 24h aberta

Menu Principal
Nome da Biblioteca
001 - super - 20:33

- ▶ Buscas & Reservas
- ▶ Circulação
- ▶ Entrada
- ▶ Edição
- ▶ Exportação
- ▶ Importação
- ▶ Relatórios
- ▶ Diagnóstico
- ▶ Alterar Senha
- ▶ Configurações (Tabelas)
- ▶ Configurações (Arquivos)

Sair

Exportação

Base de dados: Catálogo

Padrão: XML

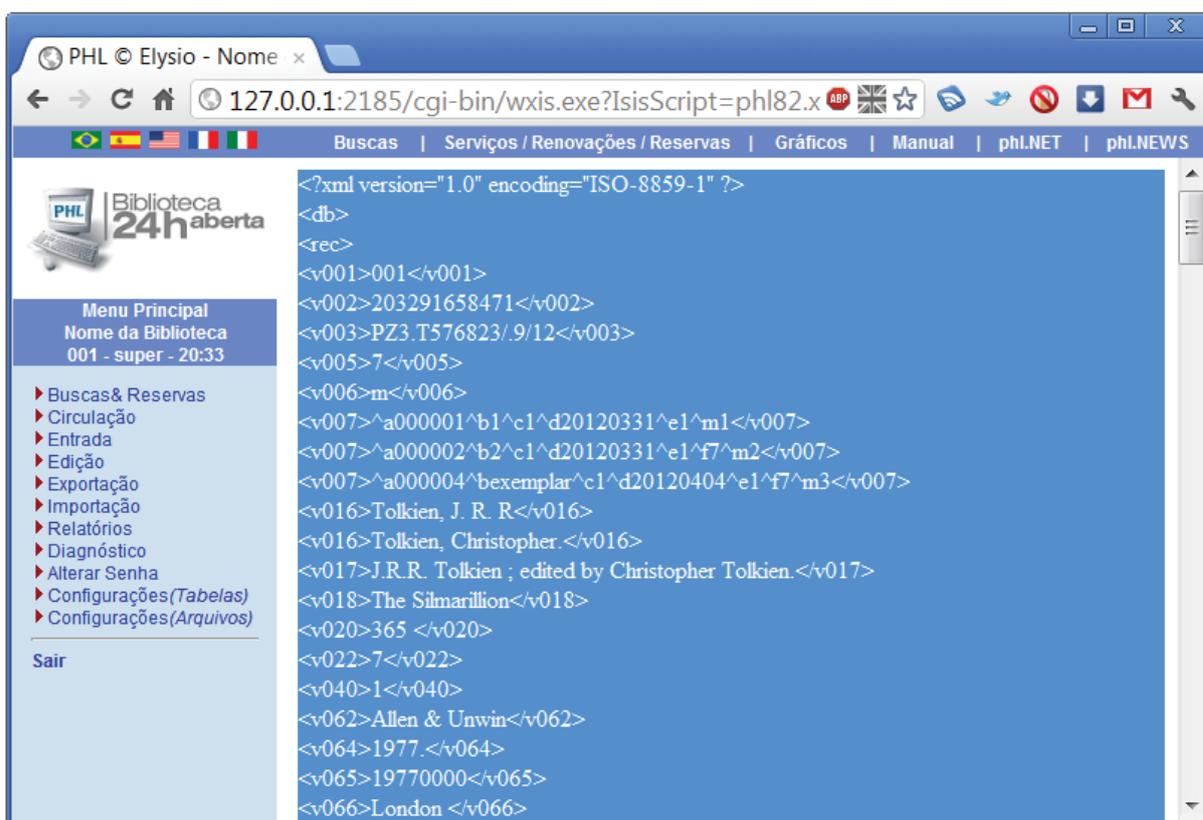
Do MFN: 1

Até: 999999

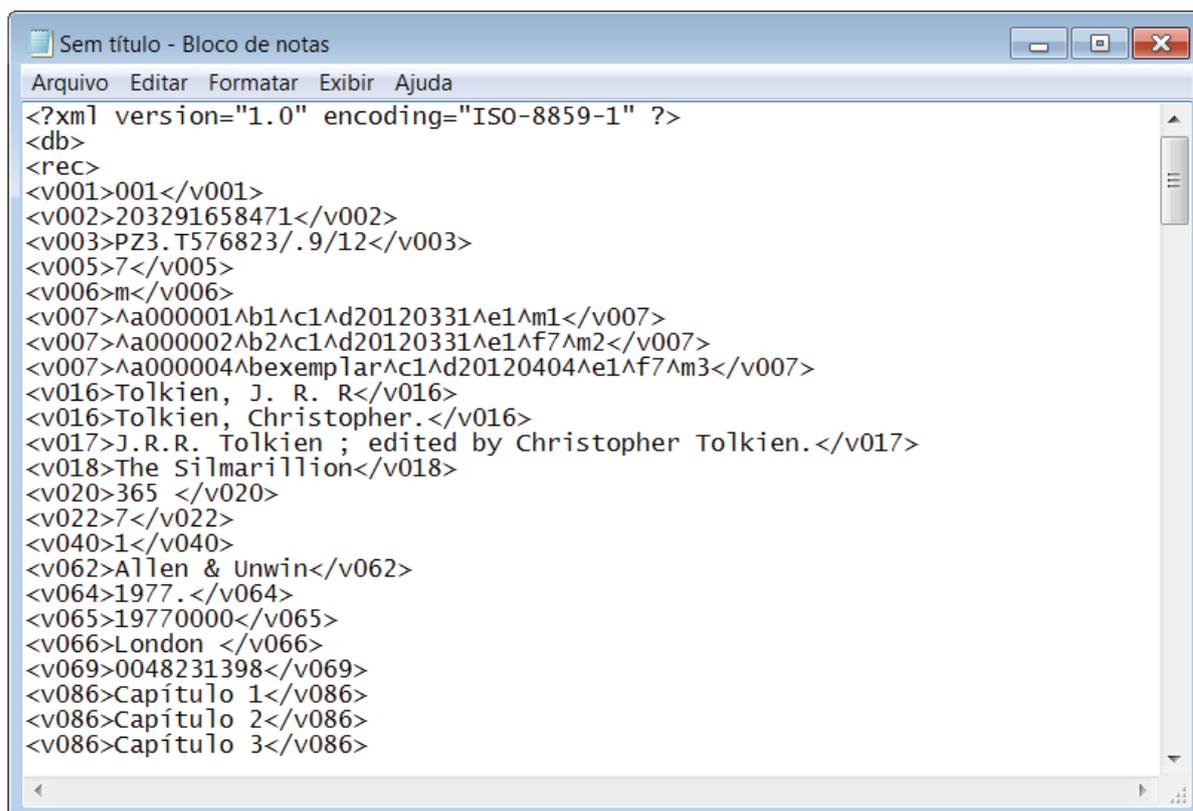
Expressão:

Confirmar

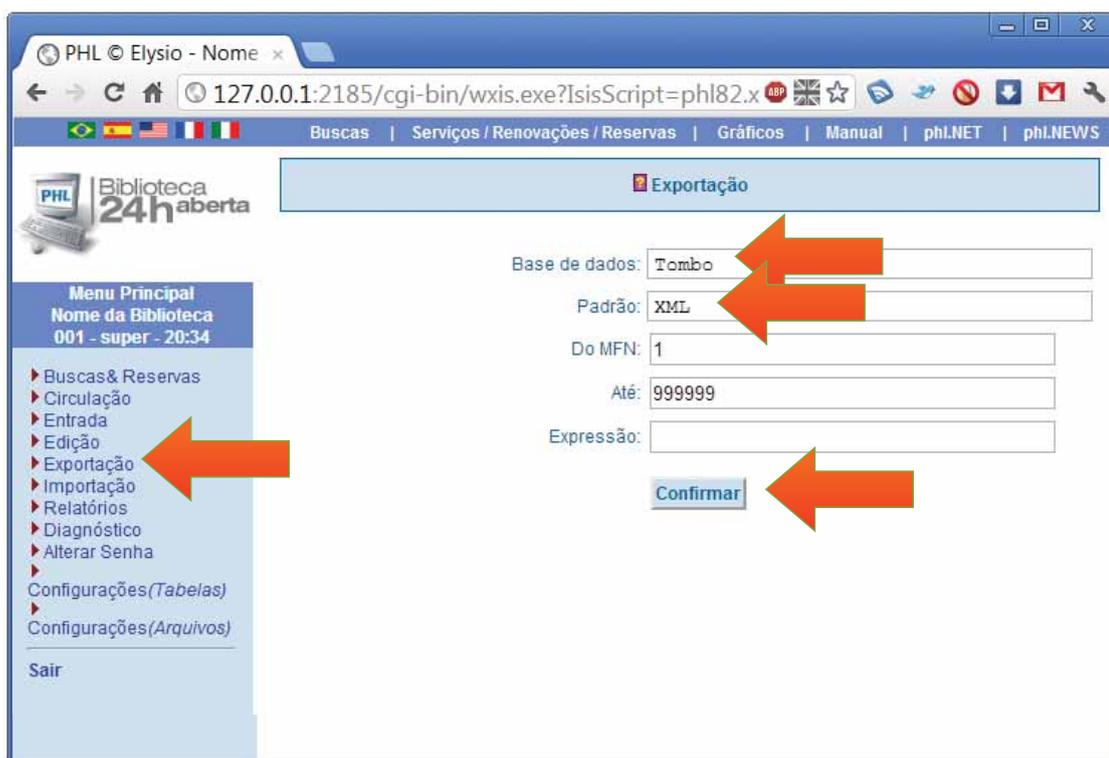
- f) Selecione todo o conteúdo, desde `<?xml version="1.0" encoding="ISO-8859-1" ?>` até `</db>`.
- g) Copie o conteúdo (Ctrl+C).



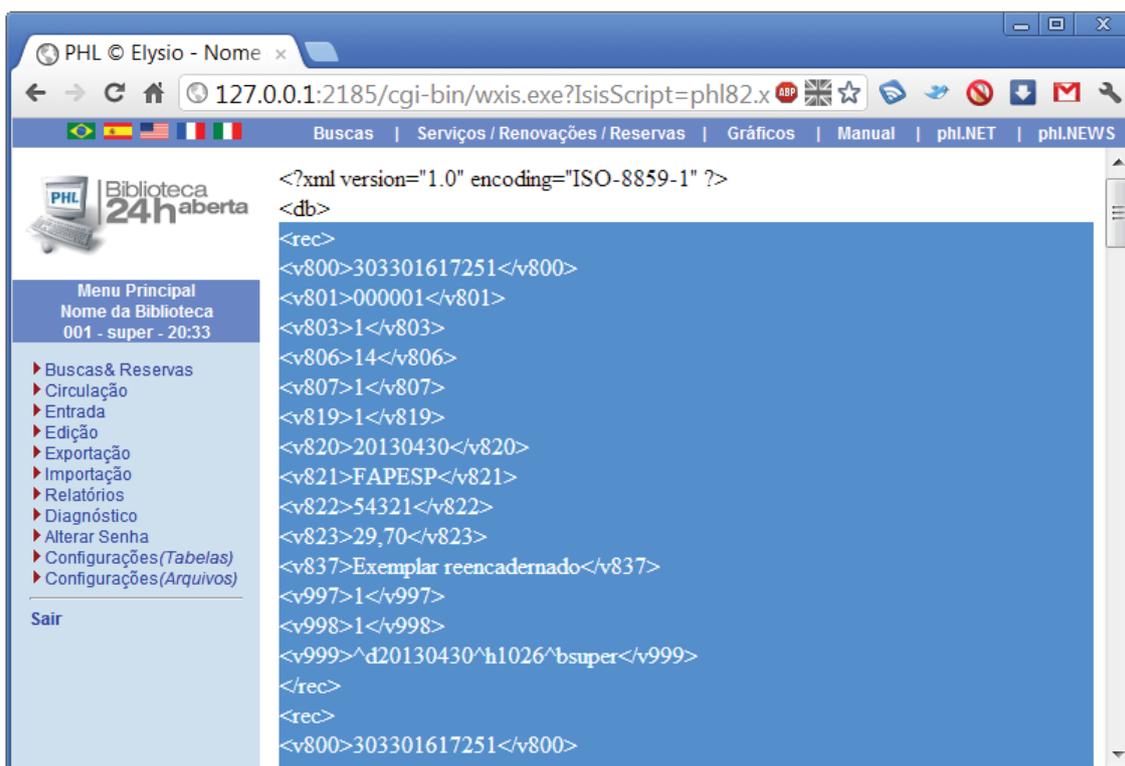
- h) Abra o Bloco de notas ou um editor de texto similar, por exemplo, o [Notepad++](#).
- i) Cole o conteúdo copiado (Ctrl+V).



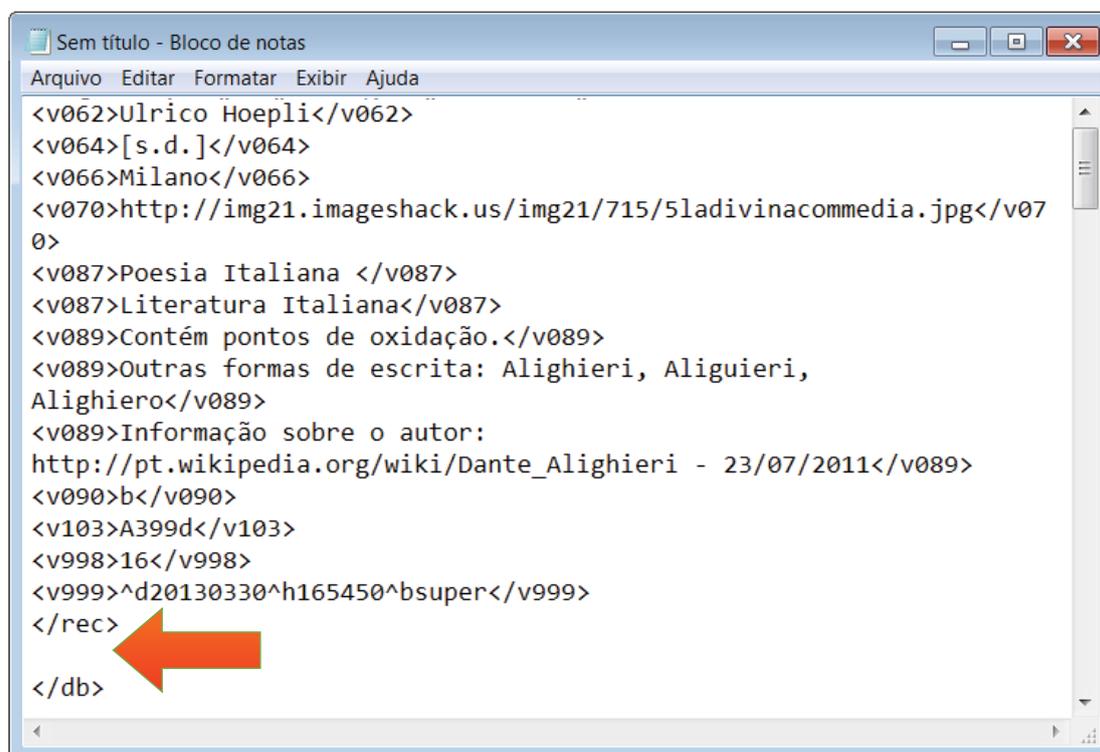
- j) Cole o conteúdo copiado (Ctrl+V).
- k) Volte ao PHL.
- l) Clique em Exportação.
- m) No campo Base de dados selecione Tombo.
- n) No campo Padrão selecione XML.
- o) Clique em Confirmar.



- p) Selecione todo o conteúdo que está **entre** <db> até </db>.
- q) **Atenção!** Não selecione <db> e </db>.
- r) Copie o conteúdo (Ctrl+C).

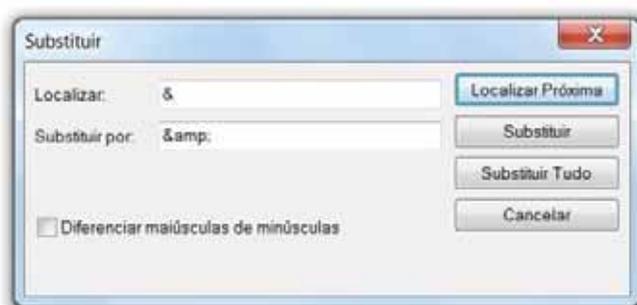


- s) Volte ao bloco de notas.
- t) Vá até o final do arquivo e clique antes de </db> (última linha do arquivo).
- u) Cole o conteúdo copiado (Ctrl+V) antes do </db>.

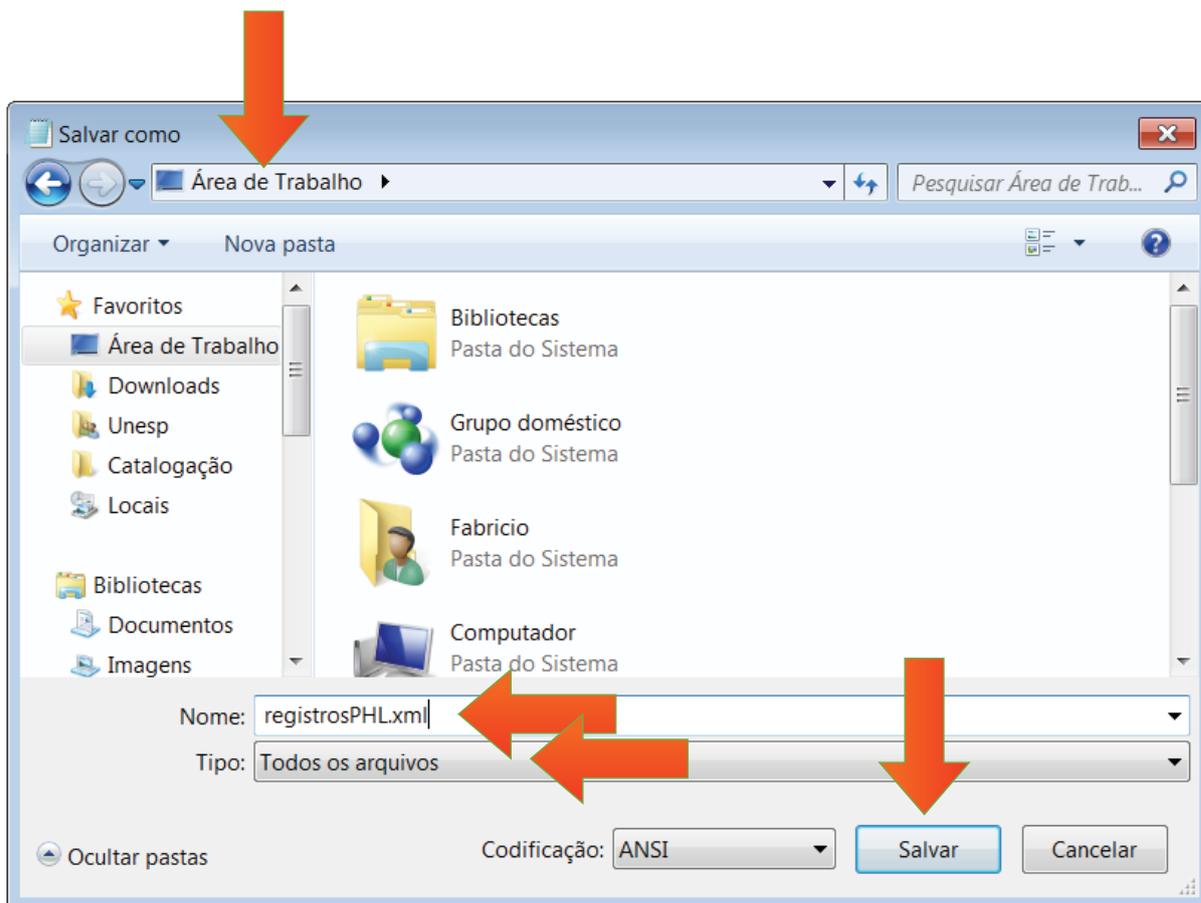


v) O caractere & é reservado para situações específicas da XML. Assim, para ser utilizado no conteúdo de um elemento, é necessário indicá-lo de forma especial utilizando o código &. Esse caractere é utilizado em alguns casos nos nomes de editoras (por exemplo, L&PM). Após colar o conteúdo copiado no Bloco de notas, é necessário verificar se o & aparece alguma vez em algum dos registros exportados e, se aparecer, substituí-lo. Para isso:

- I. Clique no menu Editar
- II. Clique em Substituir
- III. Preencha o campo Localizar com &
- IV. Preencha o campo Substituir por com &
- V. Clique em Substituir tudo.



- w) Clique no menu Arquivo.
- x) Clique em Salvar.
- y) Selecione o local em que deseja salvar o arquivo.
- z) No campo Tipo selecione Todos os arquivos.
- aa) No campo Nome digite registrosPHL.xml.
- bb) Clique em Salvar.



cc) O arquivo registrosPHL.xml será criado no local selecionado.

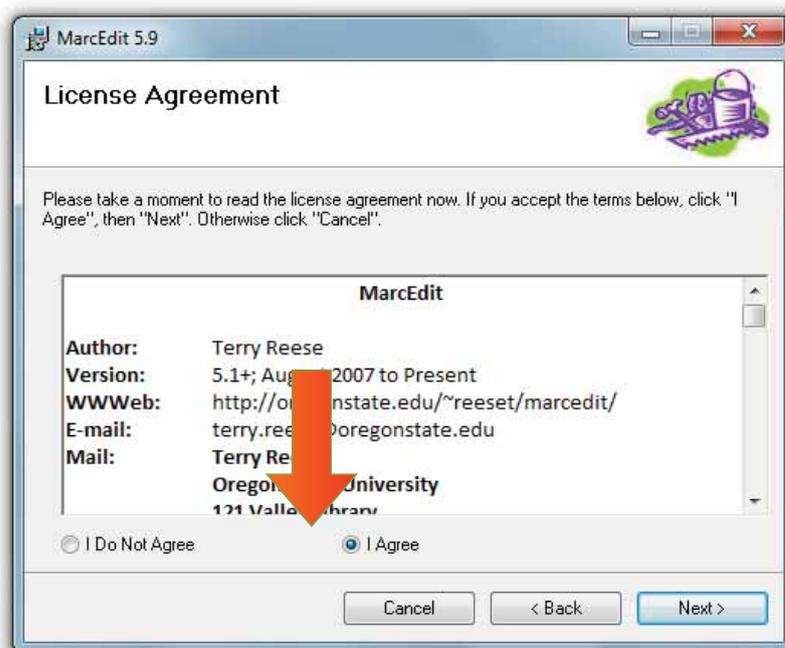
dd) Siga para o próximo capítulo.

2 Instalação do MarcEdit

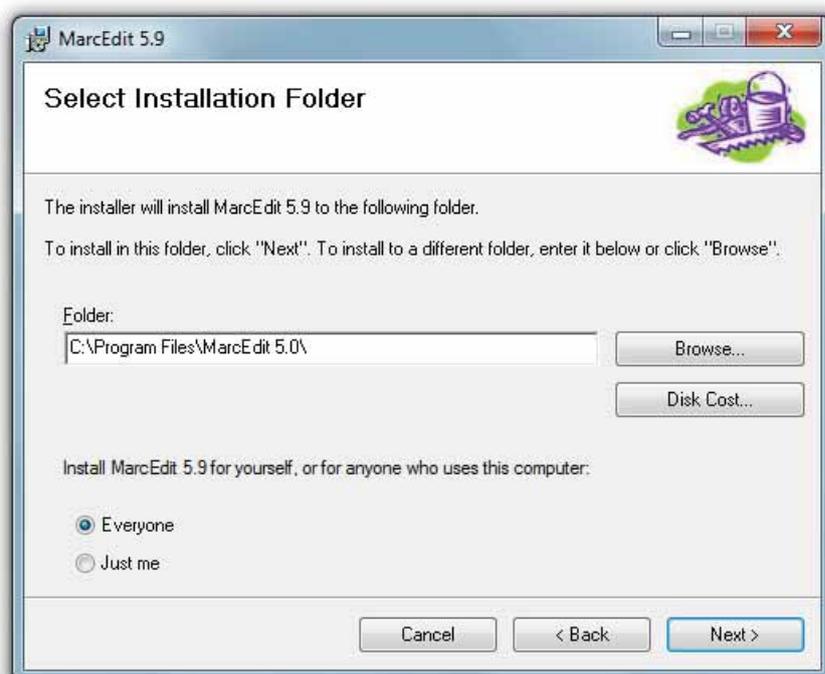
- Acesse people.oregonstate.edu/~reese/marcedit e faça download do MarcEdit.
- Clique duas vezes sobre o arquivo baixado para iniciar a instalação.
- Clique em Next.



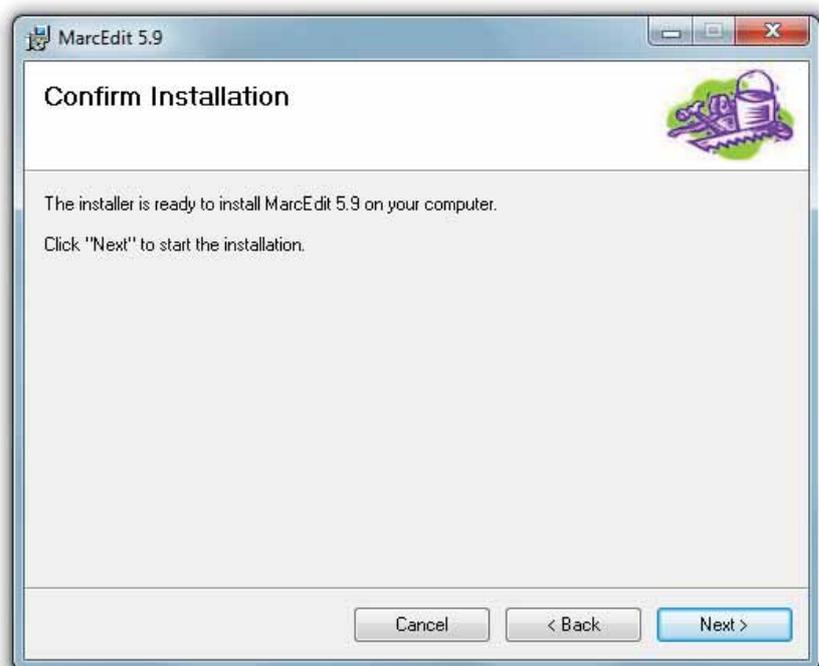
- Selecione I Agree.
- Clique em Next.



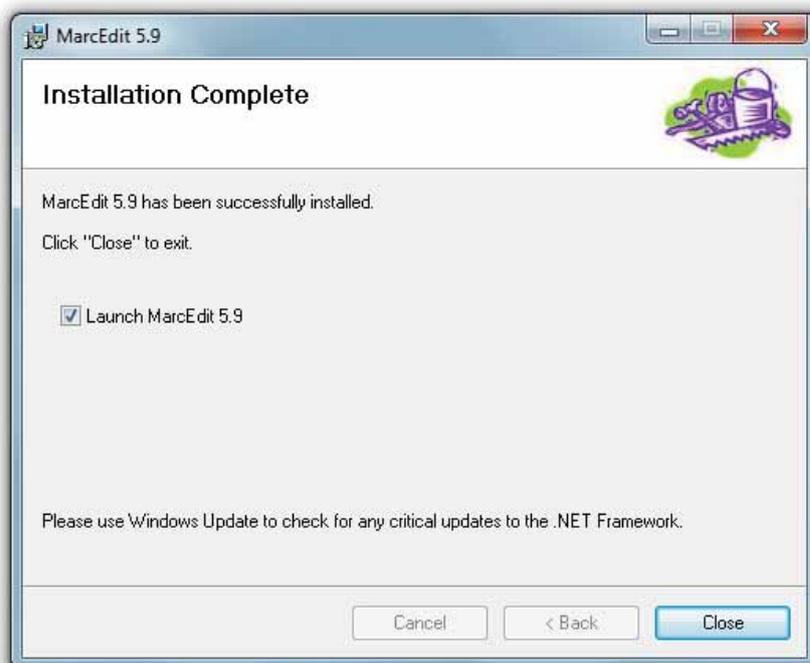
f) Clique em Next.



g) Clique em Next.



h) Clique em Close.

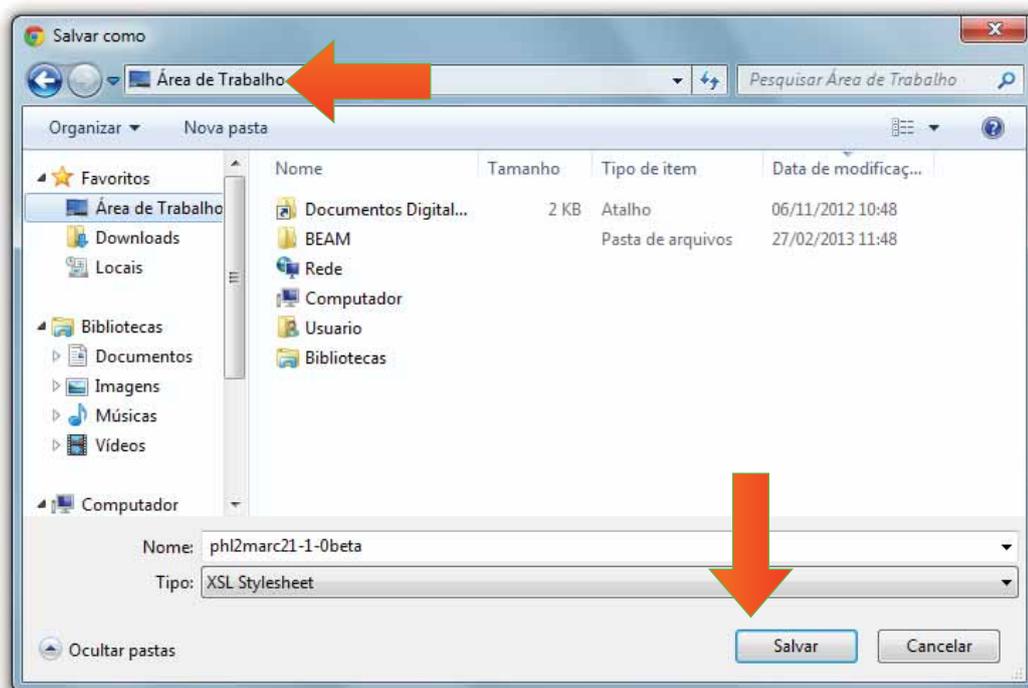


i) Opcionalmente, instale a tradução para o português seguindo as instruções dadas em: [youtube.com/watch?v=rRtZxM3zhKY](https://www.youtube.com/watch?v=rRtZxM3zhKY).

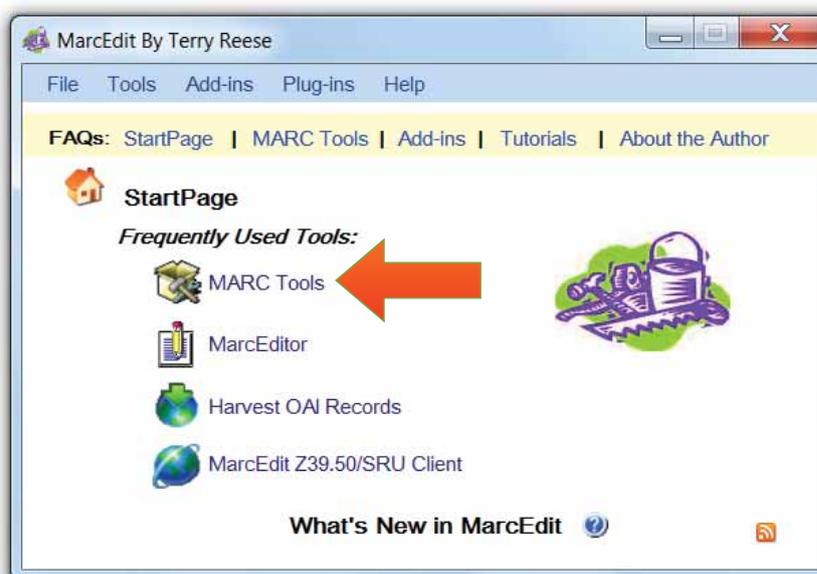
j) Siga para o próximo passo.

3 Inserção da folha de estilo no MarcEdit

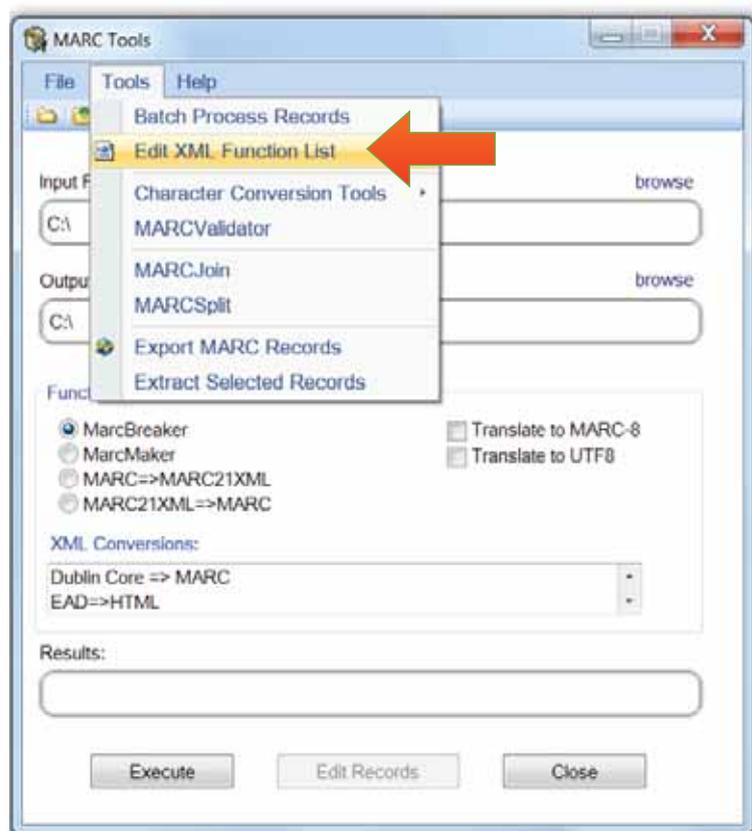
- Acesse fabricioassumpcao.com/conversao-de-phl-para-marc-21.
- Clique com o botão direito sobre a versão mais recente da folha de estilo PHL2MARC21, clique em Salvar link como ou Salvar destino como.
- Escolha o local em que deseja salvar o arquivo da folha de estilo.
- Clique em Salvar.



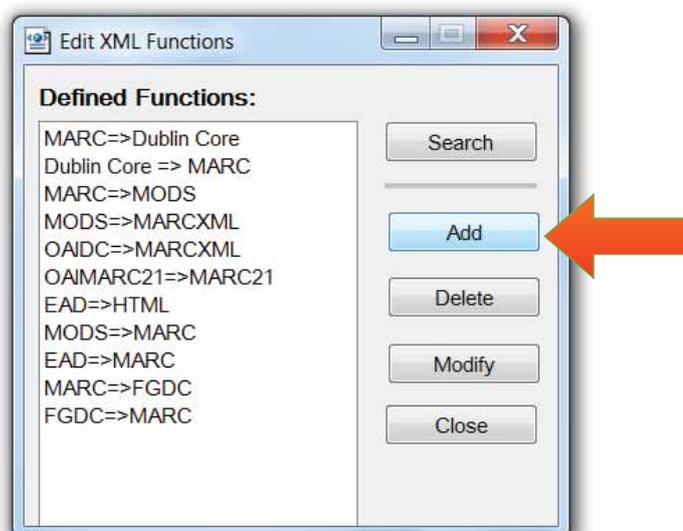
- Abra o MarcEdit.
- Clique em MARC Tools.



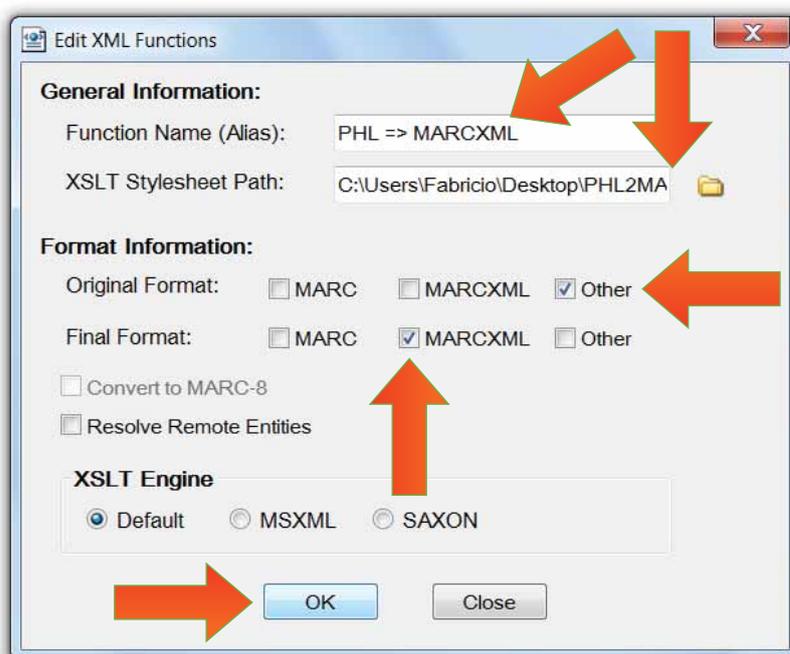
- g) Clique no menu Tools.
- h) Clique em Edit XML Function List.



- i) Clique em Add.



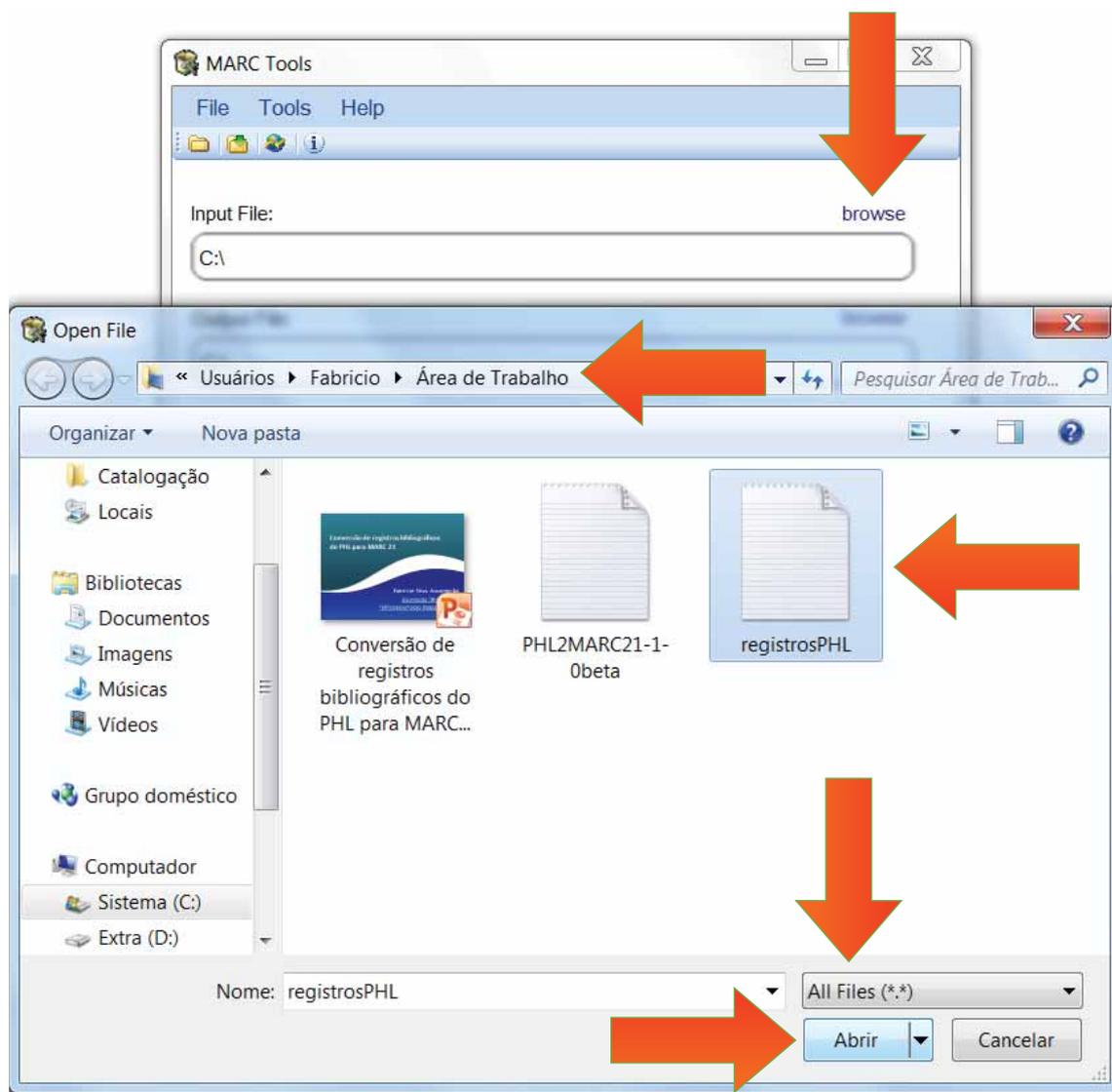
- j) Preencha o campo Function Name (Alias) com PHL => MARCXML.
- k) Clique no ícone ao final do campo XSLT Stylesheet Path.
- l) Escolha o local em que a folha de estilo PHL2MARC21.xsl foi salva.
- m) Selecione a folha de estilo.
- n) Clique em Abrir.
- o) Em Original Format marque a opção Other.
- p) Em Final Format marque a opção MARCXML.
- q) Clique em OK.



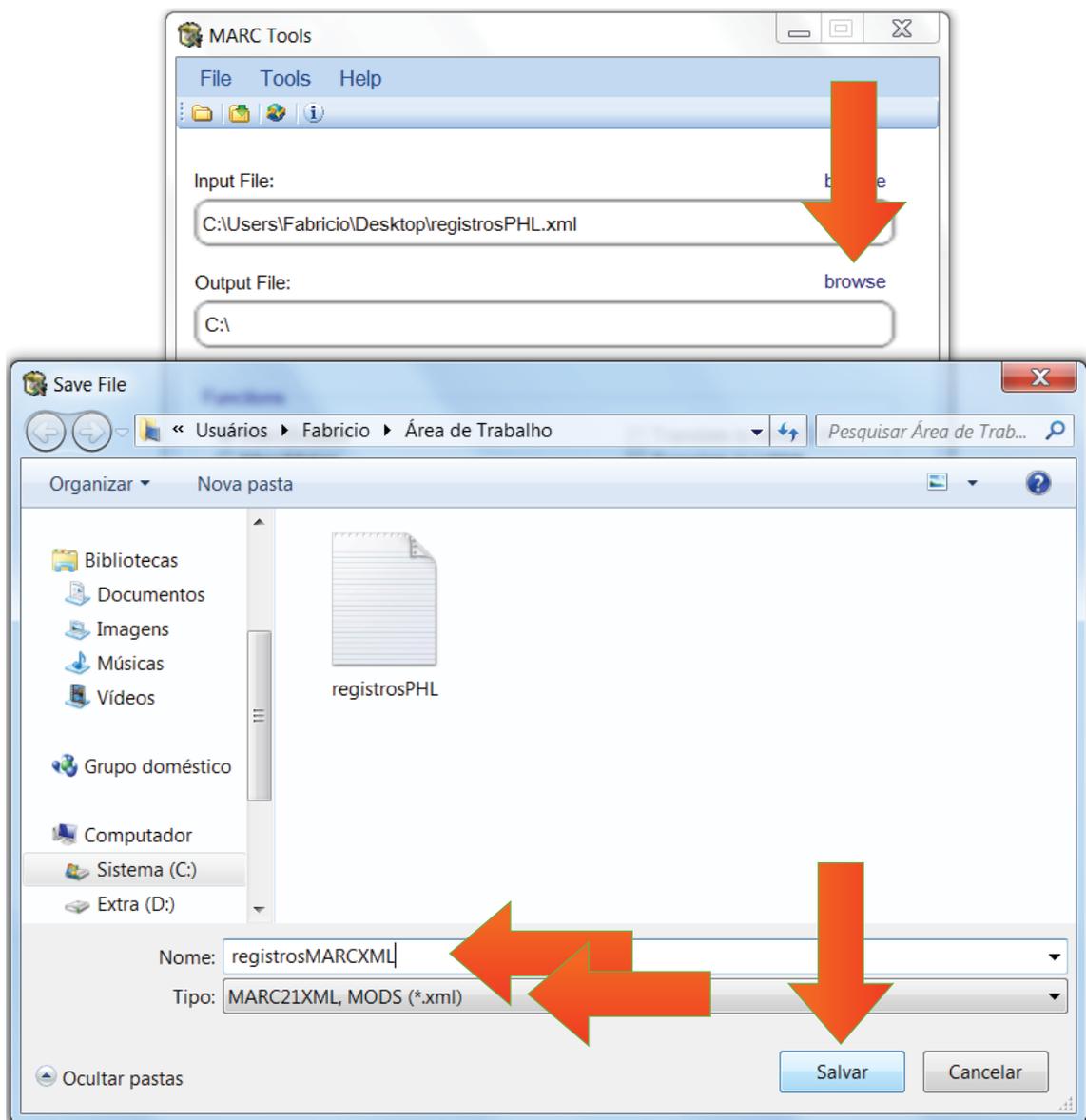
- r) Clique em Close.
- s) A janela MARC Tools será exibida.
- t) Siga para o próximo passo.

4 Conversão: PHL para MARC 21

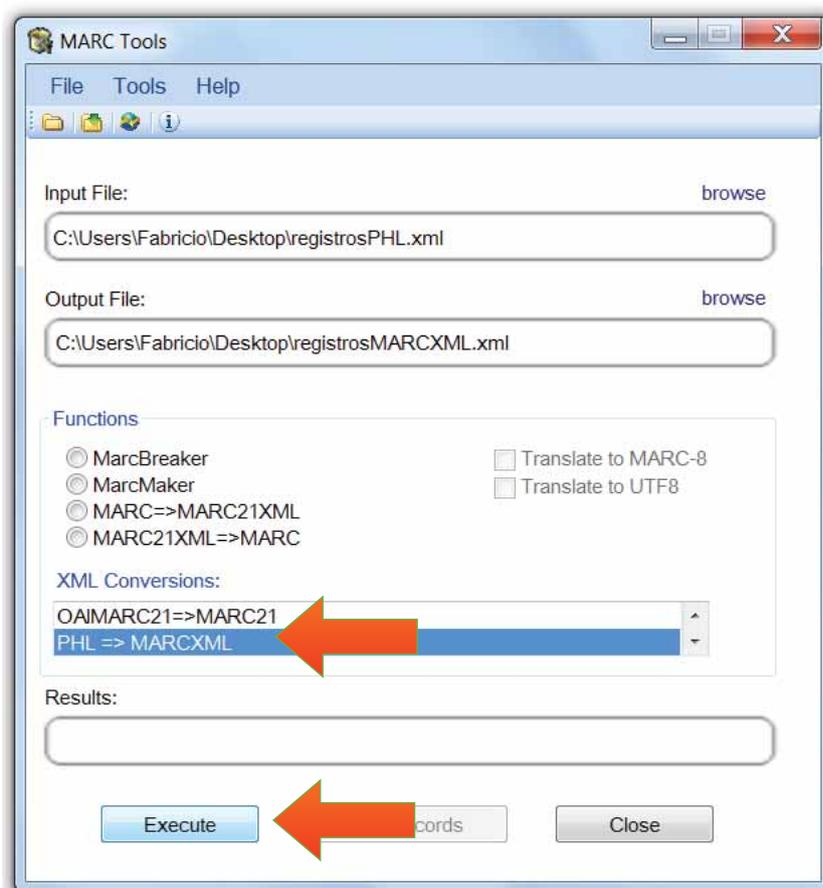
- a) Em Input File clique em Browse.
- b) Escolha o local em que o arquivo registrosPHL foi salvo.
- c) Selecione All Files (*.*)).
- d) Selecione o arquivo registrosPHL.
- e) Clique em Abrir.



- f) Em Output File clique em Browse.
- g) No campo Tipo selecione MARC21XML, MODS (*.xml).
- h) No campo Nome digite registrosMARCXML.
- i) Clique em Salvar.



- j) Em XML Conversions selecione PHL => MARCXML.
- k) Clique em Execute.

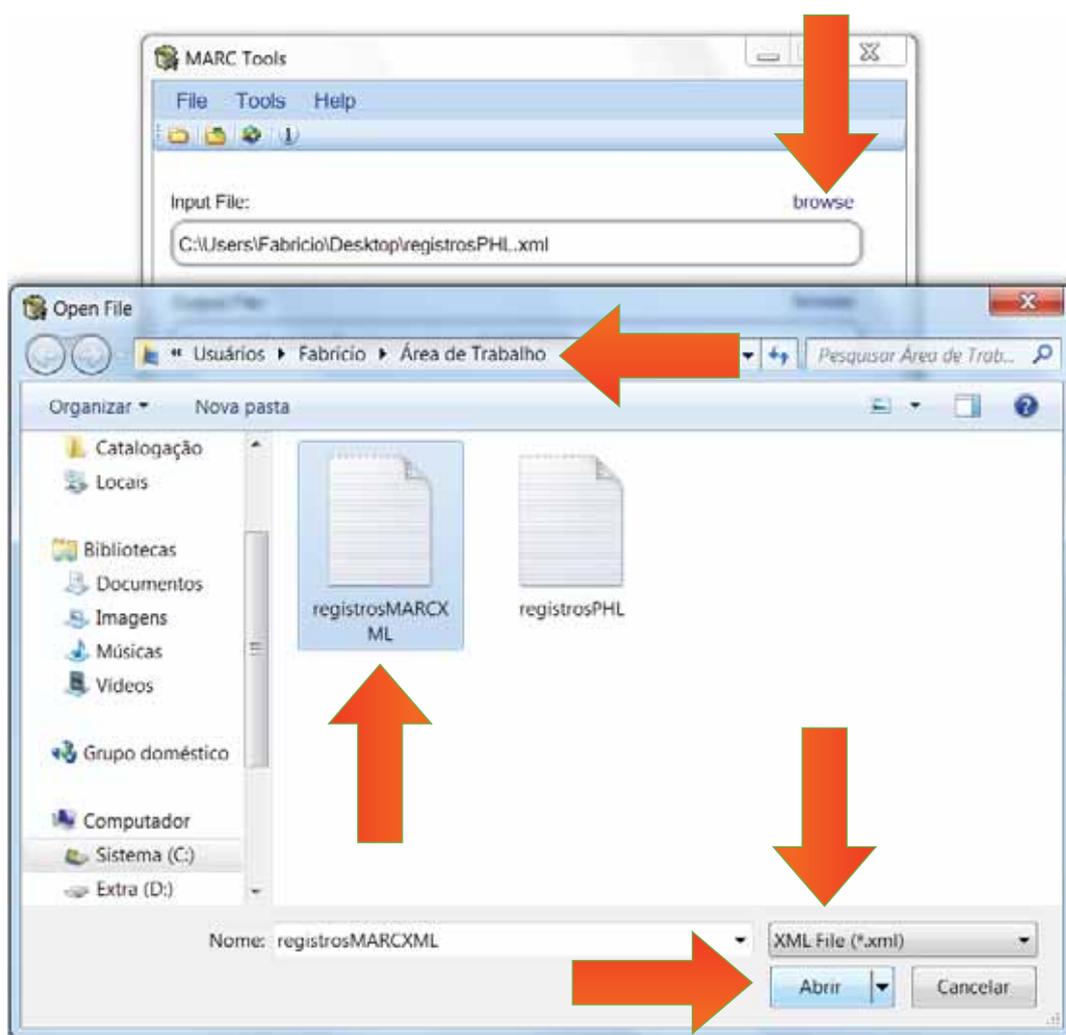


- l) O arquivo registrosMARCXML.xml será criado no local selecionado.
- m) Siga para o próximo passo.

5 Conversão: MARCXML para ISO 2709 (opcional)

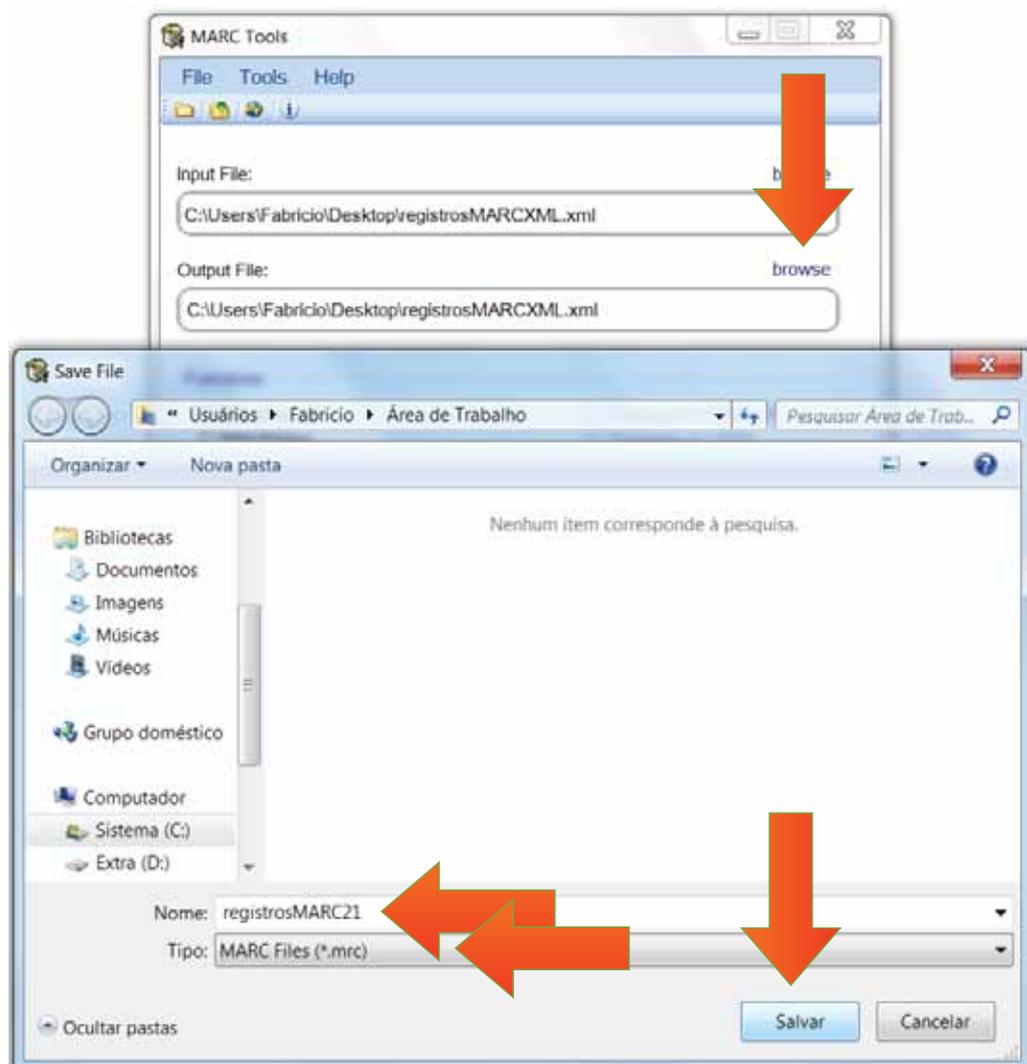
Ao final do capítulo anterior foi criado um documento XML contendo os registros no Formato MARC 21 para Dados Bibliográficos. Os registros deste documento estão de acordo com o MARCXML. Em alguns casos, no entanto, é necessário que os registros MARC 21 estejam de acordo com a ISO 2709 (arquivo .mrc) para que possam ser utilizados por um determinado software. Nesses casos, os registros MARCXML devem ser convertidos para ISO 2709. Os procedimentos para essa conversão estão descritos neste capítulo.

- a) Volte à janela MARC Tools.
- b) Em Input File clique em Browse.
- c) Selecione o local em que o arquivo registrosMARCXML.xml foi salvo.
- d) Selecione XML File (*.xml).
- e) Selecione o arquivo registrosMARCXML.xml.
- f) Clique em Abrir.

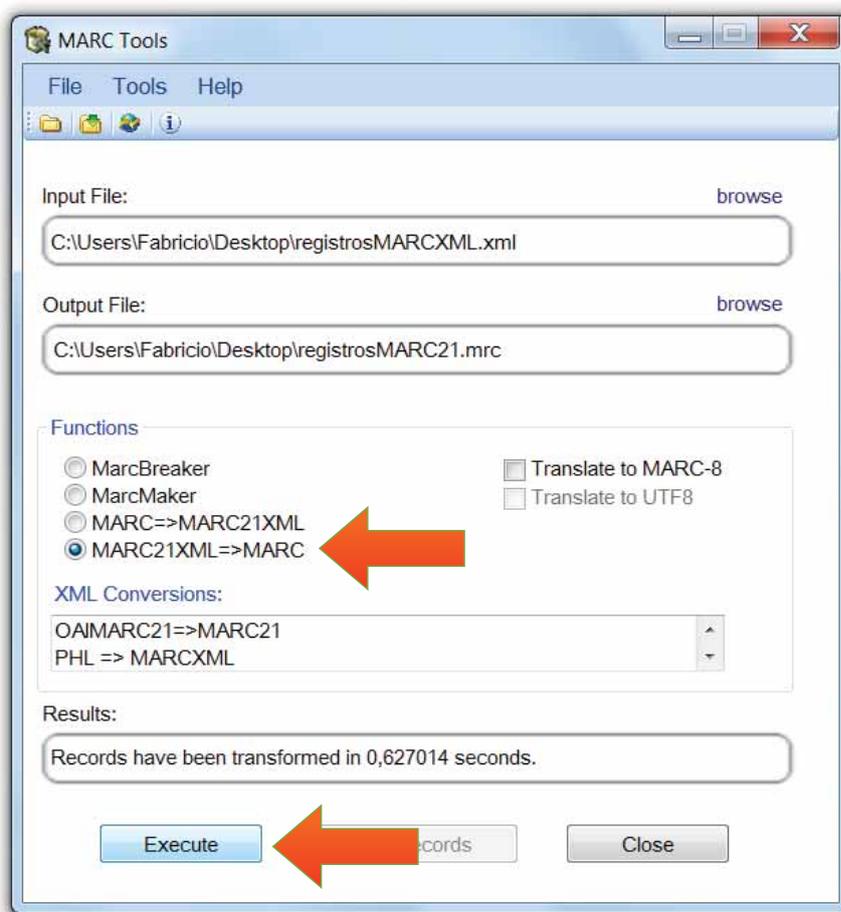


- g) Em Output File clique em Browse.

- h) No campo Tipo selecione MARC Files (*.mrc).
- i) No campo Nome digite registrosMARC21.
- j) Clique em Salvar.



- k) Em Functions marque a opção MARC21XML => MARC.
- l) Clique em Execute.



m) O arquivo registrosMARC21 será criado no local indicado.