



XVII Encontro Nacional de Pesquisa em Ciência da Informação – XVII ENANCIB

GT 8 – INFORMAÇÃO E TECNOLOGIA

COLETA AUTOMÁTICA PARA POVOAMENTO DE REPOSITÓRIOS DIGITAIS: CONVERSÃO DE REGISTROS UTILIZANDO XSLT

Silvana Ap. Borsetti Gregorio Vidotti, Flavia Maria Bastos, Ana Paula Grisoto, Felipe Augusto Arakaki, Juliano Benedito Ferreira

Modalidade da apresentação: Comunicação oral

Resumo: A criação de repositórios digitais vem ao encontro das propostas de divulgação e aumento da visibilidade das Instituições de Pesquisa e Universidades que tem como objetivo ampliar a divulgação das atividades de pesquisa desenvolvidas e da sua produção junto à comunidade. O Repositório Institucional UNESP lançado em 2013, insere-se no contexto global de Universidades que tem como objetivo à disseminação e preservação de sua produção científica, acadêmica, artística, técnica e administrativa. Como forma de solucionar o problema de povoamento de Repositórios Digitais a partir da coleta de registros das diversas fontes, este trabalho tem por objetivo apresentar os processos utilizados para a conversão de registros e reuso dos metadados de diferentes fontes de coletas de dados com base no Perfil de Aplicação de metadados adotado para inclusão de registros no Repositório Institucional UNESP. Nesse sentido, esse estudo parte do método hipotético-dedutivo que ao identificar o problema, opta por uma solução (coleta automática e uso das folhas de estilo XSLT para conversão de registros) e realizam-se testes com conjuntos de registros de diferentes fontes de dados. Por fim, foram apresentados os processos utilizados e os resultados para coleta automática no Repositório Institucional UNESP. Considera-se que a inserção de dados a partir da coleta automática, apresentou-se como uma solução para o povoamento em repositórios digitais. Concomitantemente, o uso das folhas de estilo para conversão de registros foi ratificado pelo processo apresentado resultando como um método eficiente para alimentação de dados no repositório.

Palavras-chave: Reuso de metadados; Repositório Institucional UNESP; Folhas de estilo XSLT; Conversão automática.

Abstract: *The creation of digital repositories meets the proposed disclosure and increased visibility of the institutions and universities that aims to broaden the dissemination of developed research activities and production in the community. The UNESP Institutional Repository began in 2013, is part of the global context of universities that aims to spread and preserve their scientific, academic, artistic, technical and administrative. As a way to solve the problem of collecting records from various sources, this paper aims to present the processes used for converting records and reuse of metadata from different sources of data collection based on the application profile adopted for inclusion in the UNESP Institutional Repository. Thus, this study is the hypothetical-deductive method to identify the problem, chooses a solution (automatic collection and use of XSLT stylesheets for converting records) and are tested with sets of records from different data sources. Finally, the processes used have been presented and their results for automatically collecting the UNESP Institutional Repository. Entering data from the automatic collection, it was presented as a solution to the settlement in digital*

repositories. Concurrently, the use of style sheet for records conversion have been ratified by the process presented resulting as an efficient method for data feed in the repository.

Keywords: *Reuse of metadata; UNESP Institutional Repository; XSLT stylesheets; Automatic conversion.*

1 INTRODUÇÃO

A criação de repositórios digitais vem ao encontro das propostas de divulgação e aumento da visibilidade das Instituições de Pesquisa e Universidades que tem como objetivo ampliar a divulgação das atividades de pesquisa desenvolvidas e da sua produção junto à comunidade. Segundo Leite et al. (2012) os repositórios digitais são considerados como um serviço de informação científica em ambiente digital que estão focados no gerenciamento da produção científica e/ou acadêmica de uma instituição proporcionando a interoperabilidade com outros sistemas de informação. Nesse contexto, cada vez mais Universidades têm adotado a implementação de repositórios por ser um meio eficiente para divulgar, disponibilizar e preservar a produção desenvolvida.

O Repositório Institucional UNESP foi lançado em outubro de 2013 e se insere no contexto global de Universidades que tem como objetivo à disseminação e preservação de sua produção científica, acadêmica, artística, técnica e administrativa. E ainda, a partir da demanda de criação do Repositório de Produção Científica do CRUESP (Conselho de Reitores das Universidades Estaduais Paulistas), cujo objetivo é reunir em um ambiente a produção da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), da Universidade de São Paulo (USP) e da Universidade Estadual de Campinas (UNICAMP). Em conjunto com a demanda da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que solicitou às Universidades Estaduais Paulistas que criassem repositórios institucionais com o objetivo de divulgar os trabalhos financiados por ela.

Ao iniciar o planejamento da implantação do Repositório Institucional UNESP, observou-se que dentre as formas comumente utilizadas para inclusão de registros, uma das possibilidades é por meio do autoarquivamento que consiste em "[...] depósito efetuado pelos próprios pesquisadores de suas respectivas produções científicas em repositórios digitais de acesso aberto." (FERRARI; PIRES, 2014, p. 24). Porém, a Unesp decidiu envidar esforços para criar uma forma de povoamento automático, que se deu por meio da inserção de registros em lote a partir da coleta automática dos metadados de publicação científica (objetos digitais) em diferentes bases indexadoras e fontes de informação (plataforma de busca e portais).

A decisão de incluir objetos a partir de coletas automáticas de registros de diferentes fontes foi tomada, pois, a implementação do autoarquivamento depende que o pesquisador inclua sua produção no repositório e isso poderia demandar muito tempo, uma vez que seria necessário criar campanhas de divulgação junto à comunidade, e ainda não seria garantido que o depósito fosse realizado pelo pesquisador. Outro possível problema decorrente do autoarquivamento é o tempo que demandaria o controle de qualidade das informações fornecidas pelos pesquisadores e a verificação dos direitos autorais dos documentos e que devido a emergência da ocasião, necessitava de um método mais eficaz.

Para atender a demanda da criação do CRUESP, inaugurado em outubro de 2013, a Unesp precisava povoar seu repositório com o maior número de registros da produção científica de pesquisadores vinculados à Universidade. Como meta inicial, foram definidas algumas bases de dados para a coleta de registros: *Web of Science*, *Scopus* e publicações dos periódicos da *Scientific Electronic Library Online* (SciELO), posteriormente as fontes de coletas foram ampliadas para a *PubMed*, Currículos da Plataforma Lattes e registros MARC 21 importados da Biblioteca Digital de Teses e Dissertações (C@THEDRA) e também da Biblioteca Digital de Trabalhos de Conclusão de Curso (C@PELO) da Unesp.

Entretanto, ao serem realizadas as primeiras coletas de registros, deparou-se com estruturas de metadados complementemente diferentes que necessitavam de estudos e a realização de mapeamentos dos metadados para inclusão no repositório. Como forma de solucionar o problema da coleta de registros das diversas fontes com estruturas de metadados distintos, este trabalho tem por objetivo apresentar os processos utilizados para a conversão de registros e reuso dos metadados de diferentes fontes de coletas de dados como *Web of Science*, *Scopus*, *SciELO*, *PubMed*, C@thedra - Registros MARC 21 de dissertações e teses, C@pelo - Registros MARC 21 de trabalho de conclusão de curso, Currículo Lattes (Plataforma Lattes - CNPq) com base no Perfil de Aplicação adotado para inclusão no Repositório Institucional UNESP (Perfil de Aplicação RI UNESP) e no formato suportado pelo DSpace, *software* utilizado pelo Repositório.

Nesse sentido, esse estudo parte do método hipotético-dedutivo proposto por Popper e relatado por Marconi e Lakatos (2006) que consiste em ao identificar um problema, optar por uma solução (conjectura) e realizam-se testes de falseamento com o intuito de refutar ou comprovar a hipótese. Nesse contexto, o problema caracteriza-se em como reutilizar dados vindos de diferentes fontes, a solução foi pautada nas coletas automáticas por meio das folhas de estilo *eXtensible Stylesheet Language Transformation* (XSLT), os testes foram aplicados em conjuntos de registros de diferentes fontes de dados e por fim, foram apresentados os

processos utilizados e seus resultados para coleta automática no Repositório Institucional UNESP.

2 REFERENCIAL TEÓRICO

Com o uso de diferentes esquemas de representação para descrever um recurso informacional é necessário adotar alternativas para o reuso de dados no intuito de minimizar o retrabalho da descrição. Entre uma das técnicas para o reuso dos dados está contido na utilização das folhas de estilo *eXtensible Stylesheet Language Transformation* (XSLT). Segundo Mendes, Borbinha e Manguinhas (2012, p. 90, tradução nossa) a "XSLT é usada para coletar, filtrar e associar dados vindos de documentos XML." Segundo Almeida (2002, p. 7) a *eXtensible Markup Language* (XML)

[...] possibilita ao autor especificar a forma dos dados no documento, além de permitir definições semânticas. Um arquivo eletrônico XML pode conter, simultaneamente, dados e a descrição da estrutura do documento, através do DTD-*Data Type Definitions* (gramáticas que conferem estrutura ao documento XML).

Siqueira (2003, p. 71) complementa que a “[...] XML é uma linguagem para criar padrões de comunicação entre sistemas de computadores, o que permitirá a integração tanto da base de dados como de arquiteturas, hardwares e métodos de programação usados, favorecendo a interoperabilidade.” Santarem Segundo (2004, p. 57) destaca ainda a importância da XML para

[...] Internet e grandes intranets, porque provê a capacidade de interoperação dos computadores por ter um padrão flexível, aberto e independente de dispositivo. As aplicações podem ser construídas e atualizadas mais rapidamente e também permitem múltiplas formas de visualização dos dados estruturados.

Nesse contexto, segundo Jovanovic e Gasevic (2005, p. 540, tradução nossa) "Uma folha de estilo XSLT consiste em sequência de regras, na terminologia XSLT chamado templates, que coincide com certos elementos do documento XML de entrada e transformá-los para o formato XML-alvo."

As folhas de estilo XSLT são documentos contendo conjuntos de regras escritas com a linguagem XSLT que, de modo geral, indicam ao *software* responsável pela transformação o que ele deve fazer com os dados de um arquivo XML para transformá-lo em outro arquivo. (ASSUMPCÃO et al., 2014, p. 6).

Ilik, Storlien e Olivarez (2014, p. 189, tradução nossa) explicam que "Cada folha de estilo XSLT descreve um conjunto de documentos XML (os documentos de origem) e que

devem ser convertidos em outros documentos (os documentos de resultados) [...]” em qualquer formato como em *Hyper Text Markup Language* (HTML)¹ ou em *Comma-Separated Values* (CSV)². Assumpção (2013, p. 44) complementa que

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.

Keith (2004) esclarece que a conversão de registros que estão em XML são mais facilitadas se utilizados padrões de metadados como *Dublin Core* e *Metadata Object Description Standard* (MODS). Segundo Arakaki, Santos e Alves (2015) o *Dublin Core* surgiu em um *Workshop* promovido pela *Online Computer Library Center* (OCLC) e pela *National Center for Supercomputing Applications* (NCSA) no ano de 1995 e atualmente é administrado pela *Dublin Core Metadata Initiative* (DCMI). O MODS “[...] é um padrão administrado pela LC [Library of Congress] para conjuntos de elementos bibliográficos que podem ser usados para uma variedade de fins, em particular para aplicações de biblioteca.” (LIBRARY OF CONGRESS, 2016a, não paginado, tradução nossa).

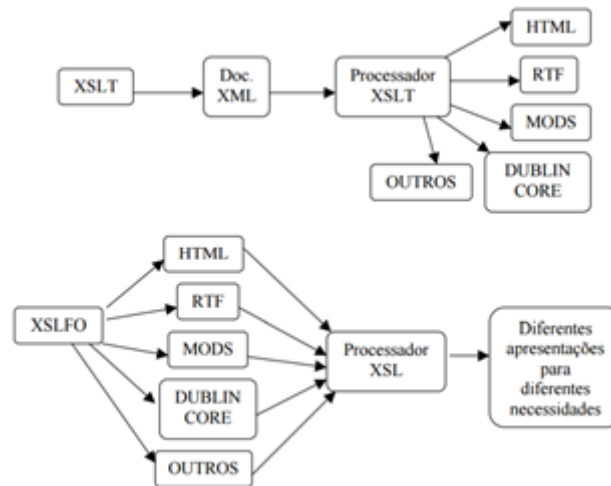
A utilização de padrões de metadados internacionalmente reconhecidos dispensa o estudo da estrutura do registro em XML e facilita o processo de mapeamento dos dados. Segundo o *World Wide Web Consortium*³ (2007, não paginado, tradução nossa) "Uma transformação na linguagem XSLT é expresso na forma de uma folha de estilo, cuja sintaxe é bem formada em XML e em conformidade com os espaços de nomes na recomendação XML." Na figura 1, Flaminio (2006) explica como o processo é realizado com a utilização das folhas de estilo XSLT.

¹ “O HTML define um tipo de documento simples, com marcações fixas projetadas para uma classe de relatórios técnicos de uso comum em escritórios, como, por exemplo, cabeçalhos, parágrafos, listas, ilustrações e algumas possibilidades para hipertexto e multimídia. É o padrão atualmente em uso na Internet.” (ALMEIDA, 2002, p.7).

² O CSV é “[...] utilizado para a troca e conversão de dados entre os vários programas de planilhas em um curto espaço de tempo.” (SHAFRANOVICH, 2005, não paginado). Ou seja, um arquivo de texto no formato de uma planilha: cada linha do arquivo é uma linha da planilha; dentro de cada linha, o conteúdo de cada célula está separado por uma vírgula e entre aspas.

³ O *World Wide Web Consortium* (W3C) é um consórcio internacional e a principal organização de padronização da *World Wide Web*.

Figura 1 - Vários estilos de apresentação a partir de um mesmo documento XML



Fonte: Flamino (2006, p. 112)

Conforme apresentado na figura 1, Flamino (2006) complementa que "A XSLT foi uma importante ferramenta para a representação e conversão de registros MARC para outros formatos de representação como o MODS e outros formatos baseados na XML." Dessa forma, considera-se que uma folha de estilo XSLT é um documento contendo regras escritas na linguagem XSLT. Essas regras instruem o *software* (processador de transformação) sobre como transformar os documentos XML. Na possibilidade do reuso dos dados, alguns estudos similares utilizaram a folha de estilo XSLT para converter registros no formato de acordo com suas necessidades informacionais, conforme exposto a seguir.

Keith (2004) utilizou as folhas de estilo XSLT para manipular registros que estavam no formato *Machine Readable Catalogin* (MARC)⁴ e seus diversos formatos como MARC21, MARCXML além do MODS e do *Dublin Core*. Segundo Ferreira (2002, p. iii) "Os formatos MARC21 são padrões amplamente usados para representação e exportação de dados bibliográficos, de autoridade, classificação, informação de comunidade e dados de coleção, em formato legível por máquina." Já o MARCXML é uma estrutura que permite trabalhar com os dados do MARC no formato XML. (LIBRARY OF CONGRESS, 2016b).

Nesse mesmo contexto, Ramos de Carvalho et al. (2004) e Dimić, Milosavljević e Surla (2010) relataram a conversão registros do formato MARC21 e *Universal MARC format*

⁴ Segundo Flamino (2006), o MARC foi criado pela *Library of Congress* dos Estados Unidos (LC) com intuito de minimizar os esforços e reduzir os custos dos processos técnicos e do compartilhamento de informações. A partir da estrutura do MARC surgiram outros formatos como o MARCXML e MARC21.

(UNIMARC) para o formato XML. O formato UNIMARC é outra versão do MARC que envolve três elementos do registro bibliográfico: estrutura de registro; designação de conteúdo; conteúdo de dados e é administrado pela *International Federation of Library Associations and Institutes* (IFLA). (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTES, 1999).

Rudic e Surla (2009) utilizaram a folha de estilo XSLT para converter registros do formato YUMARC utilizado pelo sistema *BISIS library software* que era adotado por 35 bibliotecas da Sérvia e que necessitavam a conversão de seus registros para o formato MARC21.

Assumpção (2013) e Assumpção e Santos (2014) utilizaram as folhas de estilo para converter registros do *software* de bibliotecas *Personal Home Library* (PHL) que utiliza padrão próprio para o formato MARC21.

Walsh (2011) apresentou o caso da conversão de registros no formato MARC da biblioteca da *Ohio State University's* para seu repositório institucional que utiliza o padrão *Dublin Core*. O autor destacou ainda outros exemplos de reuso de metadados como o caso relatado por Averkampe e Lee (2009) que utilizaram a folha de estilo XSLT para conversão de registros de dissertações da *ProQuest UMI Dissertation Publishing* para o repositório institucional da *University of Iowa*.

Masar (2014) expôs os procedimentos, as dificuldades e as soluções da biblioteca da *Tomas Bata University in Zlín* na República Tcheca para inclusão de dados no repositório a partir de registros da *Scopus* e da *Web of Science*.

Diante da possibilidade de converter registros em XML para o Perfil de Aplicação de metadados adotado pelo Repositório Institucional UNESP pensou-se na possibilidade da conversão de registros coletados de diversas fontes para povoar o Repositório Institucional UNESP.

3 MÉTODO DE TRABALHO

Cada uma das fontes de dados utilizadas para coleta de registros para o repositório possui uma estrutura de metadados diferente, desse modo, foi necessária a criação de uma folha de estilo XSLT para cada uma delas.

Segundo Assumpção e Santos (2014) para conversão de registros utilizando folhas de estilo é necessário alguns requisitos para que seja possível sua elaboração. O primeiro é escolher um editor XML, processador de transformação e a versão da XSLT que se quer

trabalhar, o segundo é conhecer as especificações das linguagens de marcação dos padrões de metadados tanto de origem, quanto de destino.

A especificação da linguagem de marcação utilizada com o padrão de metadados de origem nem sempre está disponível em uma *Document Type Definition* (DTD) ou em esquema XML, ou mesmo acessível ao profissional usuário do sistema de gerenciamento de bibliotecas. Nesses casos, uma possibilidade é apreender a estrutura definida pela linguagem de marcação a partir dos documentos XML que estão de acordo com ela, ou seja, a partir da análise dos registros exportados em XML. (ASSUMPÇÃO; SANTOS, 2014, p. 3813-3814).

Desse modo, as folhas de estilo criadas contêm regras que instruem o processador de transformação para que transforme os registros de cada padrão em registros no padrão *Dublin Core*, de acordo com o Perfil de Aplicação de metadados e com a linguagem de marcação genérica do DSpace (DSpace/XML) com base nas especificações e análise dos registros exportados.

No Repositório Institucional UNESP os registros coletados passam por duas conversões, a primeira delas converte os arquivos coletados XML em DSpace/XML, a segunda converte os registros em DSpace/XML para um arquivo em CSV (DSpace/CSV) e posterior importação. Dessa forma, as etapas comuns a todas as fontes de coleta de registros são:

- 1) Coleta dos registros (via interface ou em lote): A coleta de registros geralmente é feita por *software* desenvolvido pelo analista do Repositório, pelo *software* *MARC Edit*, ou outra ferramenta disponibilizada pela base de dados.
- 2) Estudo da estrutura do XML: A cada nova base de dados é estudada a estrutura do registro, mapeada de acordo com o Perfil de Aplicação do RI UNESP e transcritas as regras para conversão dos registros. Nesse momento há possibilidade de padronizar algumas informações como a geração automática da referência, padronização de alguns nomes que são recorrentes nos registros como nomes de universidades, entre outras informações.
- 3) Conversão dos registros XML em DSpace-XML: Após a criação das regras de conversão, a folha de estilo é aplicada ao conjunto de registros, que irá transformar o arquivo XML de acordo com o Perfil de Aplicação do Repositório. As folhas de estilo utilizadas na conversão dos registros coletados em registros DSpace/XML são: *Web of Science* → DSpace/XML (*wos_xml-dspace_xml.xml*); *Scopus* → DSpace/XML (*scopus_xml-dspace_xml.xml*); *SciELO* → DSpace/XML (*scielo_xml-dspace_xml*); *PubMed* → DSpace/XML (*pubmed_xml-dspace_xml*); Registros MARC 21 de dissertações e teses → DSpace/XML (*cathedra-marxml-dspacexml*) e trabalhos de conclusão de curso (*capelo-*

marcxml-dspacexml); Currículo Lattes/XML → DSpace/XML (*lattes_xml-dspace_xml.xsl*).

- 4) Conversão dos registros DSpace-XML para DSpace-CSV: Após os registros padronizados no formato DSpace-XML é aplicado uma nova folha de estilo XSLT elaborada pela equipe do Repositório, que transforma o arquivo DSpace-XML para DSpace-CSV (*dspace_xml-dspace_csv.xsl*);
- 5) Etapa de deduplicação dos registros: Uma vez que foram coletados registros de diversas fontes de dados, pode ocorrer que um registro esteja em mais de uma base de dados, assim é necessário verificar se os registros que já estão no repositório, então é realizada por meio de um programa elaborado pela equipe a deduplicação dos registros;
- 6) Verificação manual de cada registro: Nessa etapa são incluídas e/ou corrigidas informações nos registros. Quando é verificado que o autor não pertence à Universidade o registros é excluído. Outro caso de exclusão de registro ocorre quando por algum motivo o processo de deduplicação não excluiu todos os registros que já estavam no repositório. Foi observado que nesses casos, geralmente o título do registro que ainda não foi importado, estava com grafia errada ou com ano diferente, dificultando o processo de deduplicação.
- 7) Coleta do PDF: É verificado ainda, a política de acesso dos objetos digitais e se é permitida a disponibilização do arquivo em formato *Portable Document Format* (PDF). Para conferência da disponibilização é consultada a página do periódico sobre os direitos autorais e/ou as seguintes bases:
 - O SHERPA/ROMEO: O SHERPA/ROMEO é um banco de dados que contém políticas editoriais de periódicos internacionais. (UNIVERSITY OF NOTTINGHAM, 2016).
 - Diadrom: É um serviço do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICTI) que contém informações sobre autorizações para armazenamento de artigos de periódicos brasileiros em repositório digitais. (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2016?).
 - DULCINEA: O projeto DULCINEA possui informações sobre as políticas de direitos e licenças de publicações de periódicos espanhóis. (DULCINEA, 2008).
- 8) Mapeamento dos registros com as coleções: Após a planilha de verificação ser concluída, os registros são mapeados para as coleções do Repositório. Para realizar esse mapeamento é utilizado uma lista com as coleções que os departamentos e programas de Pós-Graduação pertencem, então é comparado o metadado de afiliação Unesp com a lista de coleções. Os registros que não foram mapeados são comparados utilizando o metadado autor com os nomes dos autores dos artigos, com um arquivo contendo todos os docentes da Unesp.

9) Importação dos registros no Repositório: Por fim, os registros são importado para o repositório em CSV.

3.1 COLETA DE REGISTROS EM DIFERENTES FONTES

A primeira coleta de registros foi realizada na base de dados *Web of Science*, utilizando o *web service*. O *web service* permite consultar a *Web of Science* e baixar os registros de forma automática. Os registros foram coletados pelos analistas do repositório e enviados para a equipe de bibliotecários para que realizassem a conversão e tratamento dos metadados.

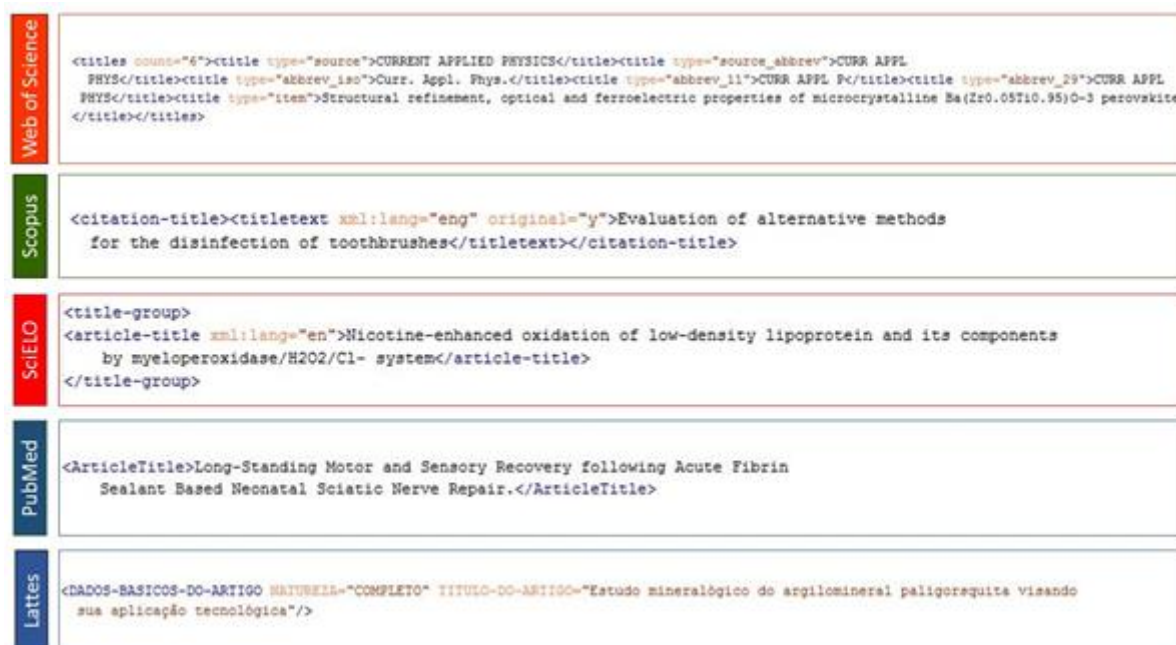
Os registros em XML da *Scopus* foram adquiridos por meio de compra o que reduziu o tempo gasto com as coletas, uma vez que nesse período não havia condições de elaborar estratégias para coletas automáticas. Seguindo os outros procedimentos já realizados para coleta de metadados, os registros da base *PubMed* foram coletados via interface da própria base pelos bibliotecários do Repositório.

Já os registros da SciELO foram coletados pelo analista pela própria interface da SciELO e cada registro vem separado em um arquivo XML. Assim, para facilitar a conversão foi necessário juntá-los em um único arquivo utilizado uma folha de estilo nomeada *join-scielo-files.xsl*. Após os registros estarem em um único arquivo é realizada a conversão dos registros.

Os registros do currículo Lattes são coletados a partir da seleção dos IDs Lattes dos docentes da Universidade e baixados por um programa desenvolvido pelo analista do repositório. Nesse processo, todos os currículos são baixados um a um em arquivos XML. Para facilitar a conversão dos dados, foi necessário juntá-los em um único arquivo. Após a reunião dos arquivos, a folha de estilo *lattes_xml-dspace_xml.xsl* é utilizada para realizar a conversão em DSpace-XML.

A figura 2 apresenta um extrato de registros de cada fonte de dados coletado em XML em um único arquivo.

Figura 2 - Título de registros em XML coletados



Fonte: Adaptado de Assumpção (2014)

Já os registros de teses, dissertações e trabalhos de conclusão de curso são coletados pelo analista a partir do *software* Aleph, utilizado pela Rede de Bibliotecas da Unesp. Antes da conversão, é necessário transformar os registros MARC 21/ISO 7909 (arquivo .mrc) em MARC 21/MARCXML (arquivo .xml); para essa conversão é utilizado o *software Marc Edit*. O formato de saída dos registros após a conversão com o *Marc Edit* é o formato DSpace-XML.

3.2 CONVERSÃO DOS REGISTROS XML PARA DSPACE-XML

Após a coleta dos registros em XML eles são convertidos em DSpace-XML, esse processo consiste em converter os registros na estrutura e no padrão de acordo com o Perfil de Aplicação de metadados adotado pelo Repositório. Os metadados utilizados pelo Repositório são apresentados no quadro 1.

Quadro 1 - Metadados utilizados no Repositório Institucional UNESP⁵.

Metadados	Termos utilizados	Obrigatório	Repetível	Documentos	Idioma
-----------	-------------------	-------------	-----------	------------	--------

⁵**Legenda:** O - Obrigatório; A - Obrigatório se aplicável ao documento que está sendo descrito; F - Facultativo; R - Repetível; NR - Não repetível; G - Todos os documentos; ART- Artigos (inclui: resenhas, editoriais, cartas, notas, trabalhos e resumos apresentados em eventos); DM - Dissertação de mestrado; TD - Tese de doutorado; TL - Tese de livre-docência; TCC - Trabalho de conclusão de curso; L - Livros; S - Sim, indicar o idioma; N - Não indicar o idioma; PT - Indicar sempre o idioma “pt” (português).

Orientador	dc.contributor.advisor	A	R	DM, TD, TCC	N
Autor	dc.contributor.author	O	R	G	N
Instituição	dc.contributor.institution	O	R	G	N
Data de publicação	dc.date.issued	O	NR	G	N
Nota	dc.description	F	R	G	N
Resumo	dc.description.abstract	O	R	G	S
Afiliação	dc.description.affiliation	A	R	ART	N
Afiliação do autor UNESP	dc.description.affiliationUnesp	A	R	ART	N
Agência de fomento/financiamento	dc.description.sponsorship	A	R	G	N
Número do processo/financiamento	dc.description.sponsorshipId	A	R	G	N
Extensão do recurso	dc.format.extent	F	NR	G	N
URL do recurso informacional	dc.identifier	A	R	G	N
Identificador do Aleph (SYS)	dc.identifier.aleph	A	NR	G	N
Identificador da CAPES para Programas de Pós-Graduação	dc.identifier.capes	A	R	G	N
Referência do documento	dc.identifier.citation	A	NR	G	N
DOI	dc.identifier.doi	A	R	G	N
Nome do arquivo ou o caminho do arquivo	dc.identifier.file	A	R	G	N
ISSN	dc.identifier.issn	A	R	ART	N
Identificador do Currículo Lattes	dc.identifier.lattes	O	R	ART	N
Identificador PubMed	dc.identifier.pubmed	A	NR	ART	N
Identificador PubMed Central	dc.identifier.pmc	A	NR	ART	N
Identificador SciELO	dc.identifier.scielo	A	NR	ART	N
Identificador Scopus	dc.identifier.scopus	A	R	ART	N
Identificador do registro no Repositório	dc.identifier.uri	O	NR	G	N
Identificador Web of Science	dc.identifier.wos	A	R	ART	N
Idioma	dc.language.iso	O	R	G	N
Publicador	dc.publisher	A	R	G	N
Fonte	dc.relation.ispartof	A	NR	ART	N
Fonte do registro	dc.source	A	NR	G	N
Palavra-chave	dc.subject	O	R	G	S
Título	dc.title	O	NR	G	S
Título alternativo	dc.title.alternative	A	R	G	S
Tipo	dc.type	O	NR	G	N
Condições de acesso	dc.rights.accessRights*	O	NR	G	N
Licença de uso	dcterms.license	A	R	G	N
Detentor dos direitos	dcterms.rightsHolder	A	R	G	N
Unidade	unesp.campus	O	R	G	PT
Departamento	unesp.department	A	R	G	PT
Programa de Pós-Graduação	unesp.graduateProgram	A	R	G	PT
Área de concentração	unesp.knowledgeArea	A	R	G	PT
Linha de pesquisa	unesp.researchArea	A	R	G	PT
Data para a disponibilização do conteúdo integral	unesp.embargo	A	NR	G	PT

Fonte: Elaborado pelos autores

O Perfil de Aplicação RI UNESP foi desenvolvido para atender os artigos 2º e 3º da Portaria UNESP nº 88 e guiou-se pela necessidade de prover metadados que possibilitassem a localização, a identificação, o acesso, o armazenamento e a preservação da produção científica, acadêmica e administrativa da Universidade, proporcionando ainda, a criação de indicadores para essa produção e assegurando a capacidade de interoperar com sistemas nacionais e internacionais.

Nesse contexto, os dados são mapeados e convertidos para o Perfil de Aplicação RI UNESP. O resultado da conversão dos registros XML para DSpace-XML é apresentado na figura 3, um fragmento dos registros apresentando o título de um registro de cada uma das fontes utilizadas para coleta.

Figura 3 - Registro convertido para DSpace-XML

Web of Science	<pre><dcvalue element="title" language="en">Structural refinement, optical and ferroelectric properties of microcrystalline Ba(Zr0.05Ti0.95)O-3 perovskite</dcvalue></pre>
Scopus	<pre><dcvalue element="title" language="en">Evaluation of alternative methods for the disinfection of toothbrushes</dcvalue></pre>
SciELO	<pre><dcvalue element="title" language="en">Nicotine-enhanced oxidation of low-density lipoprotein and its components by myeloperoxidase/H2O2/Cl- system</dcvalue></pre>
PubMed	<pre><dcvalue element="title" language="en">Long-Standing Motor and Sensory Recovery following Acute Fibrin Sealant Based Neonatal Sciatic Nerve Repair</dcvalue></pre>
Lattes	<pre><dcvalue element="title" language="pt">Estudo mineralógico do argilomineral paligorsquita visando sua aplicação tecnológica</dcvalue></pre>
MARC21	<pre><dcvalue element="title" language="pt">Estudo do potencial terapêutico da maitenina em células de câncer de colo uterino</dcvalue></pre>

Fonte: Elaborado pelos autores

Após a conversão em DSpace-XML os registros já estão na estrutura adotada pelo Repositório, a próxima etapa do processo é converter os registros em arquivo CSV para que as demais atividades de tratamento das informações possam ser executadas.

3.3 CONVERSÃO DE REGISTROS EM FORMATO CSV

Para a conversão dos registros em CSV é utilizado a folha de estilo *dspace_xml-dspace_csv.xslt*. Os arquivos CSV podem ser melhor visualizados se forem abertos no *software* livre *Libre Office*⁶. A figura 4 apresenta arquivos no formato CSV.

No arquivo CSV, cada coluna corresponde a um metadado, exceto as colunas “id” e “collection”, que correspondem, respectivamente, ao ID do item no DSpace e ao handle da coleção da qual o item faz parte. Conforme apresentado na figura 4, todos os metadados devem estar acompanhados de [], seja vazio (apenas []) ou com o código do idioma: [pt], [en], [es], etc.

Figura 4 - Registros em CSV

	A	B	C	D	E	F	G	H	I	J
1	id	collection	dc.contributor	dc.contributor	dc.contributor	dc.date.issued	dc.description	dc.description	dc.description	dc.description
2	+	11449/77075	Sarti, Flávia	Oliveira, Kari	Universidade	2013-12-16	This work aims to identify a	Este trabalho	Coordenação	
3	+	11449/77075	Souza Neto,	Recco, Kethy	Universidade	2014-09-01	This study deals with teachi	Este estudo	Coordenação	
4	+	11449/11029	Brandão, Clá	Kemper, Ber	Universidade	2015-01-08	The objective of this study	O objetivo do presente estud		
5	+	11449/77162	Damasceno,	Dallaqua, Bru	Universidade	2015-02-06	The purpose of this study was to investiga	Fundação de		
6	+	11449/77090	Castiglia, Ya	Sudre, Eliana	Universidade	2015-02-13				
7	+	11449/77191	Fernandes Jô	Andrade, Bru	Universidade	2015-02-23	Natural products such as e	Produtos naturais, como os		
8	+	11449/77215	Oliveira, Clau	Rossini, Brun	Universidade	2015-02-24	The Characidae family is th	A família Cha	Coordenação	
9	+	11449/77090	Modolo, Norm	Cabral, Luca	Universidade	2015-02-26	Background: it has been su	Introdução: a	Coordenação	
10	+	11449/77222	Henry, Raoul	Silva, Carolin	Universidade	2015-02-26			Coordenação	
11	+	11449/77215	Marino, Celso	Revolti, Paol	Universidade	2015-02-27	Eucalyptus has become on	O Eucalyptus passou a ser i		
12	+	11449/77222	Silva, Reinald	Wunderlich,	Universidade	2015-02-27			Coordenação	
13	+	11449/11029	Alves, Ana Li	Yamada, Ana	Universidade	2015-03-27	Several methods of cell ther	As diferentes	Fundação de	
14	+	11449/77215	Souza Neto,	Dreyer, Carin	Universidade	2015-04-06	Dengue is the arboviro	Dengue é a	Fundação de	
15	+	11449/77168	Silva, Márcia	Ramos, Brun	Universidade	2015-04-24	Introduction: A genetic pred	Introdução: O Trabalho de P		
16	+	11449/77185	Chiacchio, Si	Quevedo, Da	Universidade	2015-06-03	Stress can interfere with an	O estresse pode interferir nc		
17	+	11449/11029	Meira, Cezin	Silva, Elisa S	Universidade	2015-06-22	To investigate the effects o	Com o objetiv	Fundação de	
18	+	11449/76985	Pietro, Rosen	Freitas, Jolin	Universidade	2015-07-02	Piper umbellatum L. Miq. is	Piper umbella	Fundação de	
19	+	11449/11029	Rahal, Sheila	Mesquita, Lu	Universidade	2015-07-03	The aim of this study was to	O objetivo do	Coordenação	
20	+	11449/77185	Modolo, Rafa	Ferreira, Bru	Universidade	2015-07-20	Responsible for significant	Responsável por perdas ecc		
21	+	11449/77215	Foresti, Faus	Jaime, Cristi	Universidade	2015-07-29	In the present study, six sp	No presente	Comisión Nac	

Fonte: Elaborado pelos autores

Após esse processo, o arquivo está pronto para próxima etapa que consiste na remoção dos registros que já estão no Repositório por meio de um *software* criado pelo analista. Posteriormente, é realizada a verificação individual de cada registro e coleta dos objetos digitais.

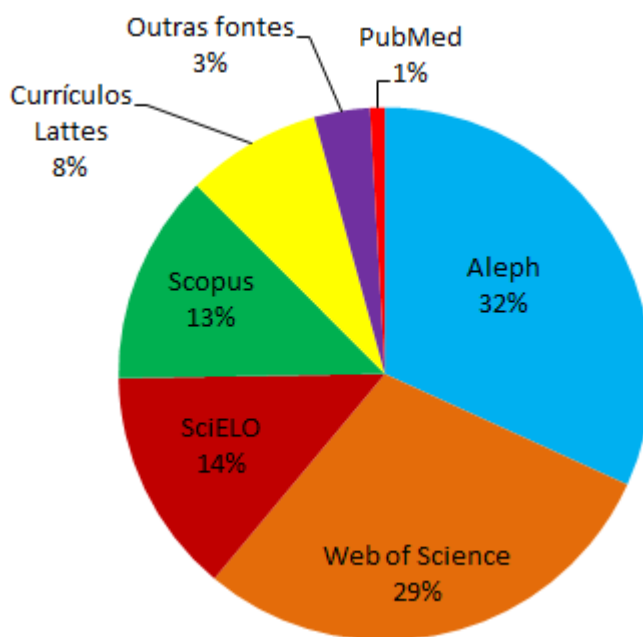
O processo de verificação dos registros provenientes do Currículo *Lattes* demanda mais tempo por serem registros que contém informações preenchidas pelos autores, muitas vezes são descritas de forma sucinta, o que gera a necessidade de complementar os registros.

⁶ Disponível em: <<https://pt-br.libreoffice.org/>>. Acesso em: 12 set. 2016.

4 RESULTADOS DA IMPORTAÇÃO

A importação é o processo que permite a inclusão de itens de forma automática e pode ser realizada de duas formas, a primeira é via interface, o arquivo CSV contendo os registros pode ser importado com um limite de 100 registros por arquivo. A segunda forma de importação é por linha de comando, realizada pelo analista e cada arquivo CSV pode conter até 1.000 registros. Dentre as formas de importação disponíveis no DSpace, no repositório é utilizada principalmente a importação através de arquivos CSV por linha de comando, por ser um processo mais rápido. A figura 5 apresenta as origens das fontes dos registros que constam no Repositório.

Figura 5 - Origem das fontes dos registros



Fonte: Elaborado pelos autores

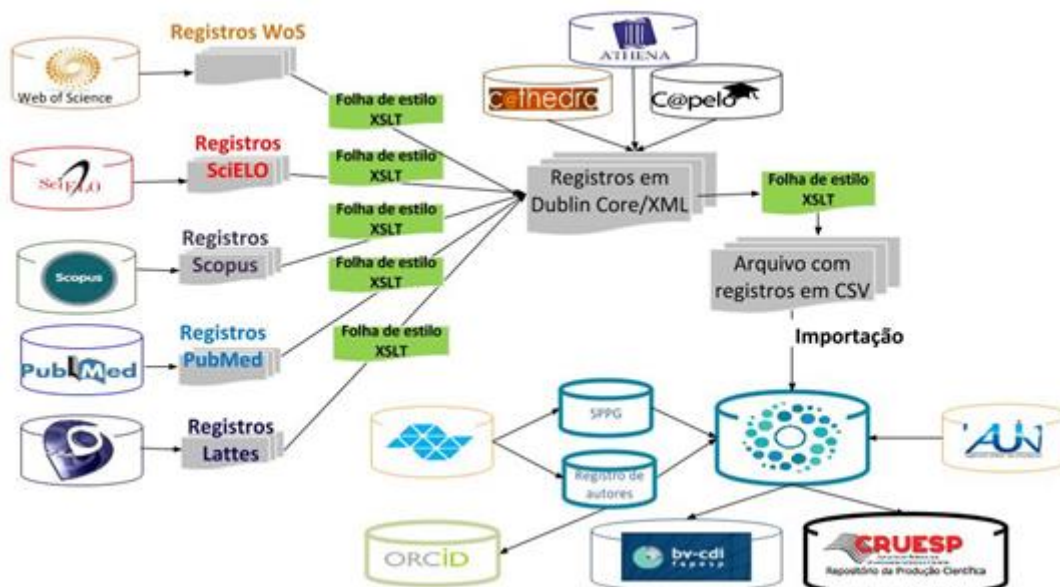
A base do catálogo Aleph corresponde com a maior parcela com 32.471 mil registros (32%) que foram convertidos do formato MARC21 e que representa ainda os dados das bases *C@pelo* e *C@thedra*. A *Web of Science* é a segunda maior fonte de dados representando 29.847 mil registros (29%). A *SciELO* aparece em terceiro lugar com 13.962 mil registros (14%). A *Scopus* ficou em quarto lugar com 13.037 mil registros (13%). Em quinto lugar estão os dados vindos dos Currículos *Lattes* dos docentes da universidade, totalizando 8.376 mil registros (8%). Em sexto lugar, com 3.447 mil registros (3%) estão outras fontes como o

autoarquivamento de teses e dissertações adotado em dezembro de 2015, entre outros dados inseridos manualmente. Por fim, os dados da *PubMed* representam 917 registros (1%).

Durante as coletas observou-se que muitos registros estavam indexados em mais de uma fonte de dados, o que justifica a maior quantidade de registros da *Web of Science* que foi a primeira fonte de coleta de dados e assim sucessivamente.

A figura 6 apresenta de forma geral, as coletas e conversão dos registros até a etapa de importação para o Repositório Institucional UNESP. Uma vez que os dados são inseridos no repositório, há dados que são direcionados ao Repositório do CRUESP e dados referentes aos auxílios da FAPESP são enviados por meio de uma *Application Programming Interface* (API). O Repositório firmou parceria ainda com a Agência Unesp de Inovação (AUIN) que insere os dados de patentes. A Unesp fornece dados para que os alunos de Pós-Graduação possam realizar o autoarquivamento de dissertações e teses conforme a Portaria UNESP 396, de 10 de setembro de 2015 e entrou em vigor em dezembro de 2015. Os estudos de reuso de metadados continuam e um dos próximos passos é o desenvolvimento da base de registros de autores e a integração do repositório com o *Open Researcher and Contributor ID* (ORCID), que é um currículo internacionalmente reconhecido e tem sido utilizado por diversas bases de dados.

Figura 6 - Coleta, conversão e importação de registros do Repositório Institucional UNESP



Fonte: Adaptado de Bastos e Vidotti (2015)

O desenvolvimento do método utilizado para o reuso dos dados tem tomado reconhecimento nacional e muitas instituições solicitaram auxílio e transferência dos

procedimentos adotados, como a Fundação Oswaldo Cruz (FIOCRUZ); o Instituto de Pesquisas Energéticas e Nucleares (IPEN); o Instituto de Engenharia Nuclear (IEN); a Universidade Estadual de Campinas (UNICAMP); a Universidade Federal de São Carlos (UFSCAR); a Universidade Federal de São Paulo (UNIFESP); a Universidade Federal de Uberlândia (UFU); a Universidade Federal do Rio de Janeiro (UFRJ); e a Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

5 CONCLUSÃO

O trabalho apresentou o estudo de coleta de registros e reuso de metadados de diferentes fontes. Constatou-se que o domínio da estrutura do XML foi importante para o desenvolvimento das atividades, pois representa peça-chave para conversão dos registros com o uso de folhas de estilo XSLT. A inserção de dados a partir da coleta automática apresentou-se como solução viável para o povoamento em Repositórios Digitais. Concomitantemente, o uso das folhas de estilo para conversão de registros foi ratificado pelo processo apresentado, resultando como um método eficiente para alimentação de dados no Repositório.

O papel do bibliotecário foi fundamental no processo de estudo na elaboração das folhas de estilo, no processo de mapeamento dos metadados de acordo com o Perfil de Aplicação RI UNESP estabelecido pelo Repositório e na conversão dos registros. Cabe destacar que o conhecimento dos metadados que foi substancial para realização dos mapeamentos, controle de qualidade e auxílio nas coletas dos dados. Uma vez que, compreendida a estrutura do XML da fonte de origem é necessário o mapeamento dos dados para o Repositório Institucional UNESP. Outro ponto a ser destacado é o trabalho desenvolvido em conjunto com profissionais de outras áreas como a Ciência da Computação, no desenvolvimento de programas para coleta, vinculação de coleções e deduplicação de registros.

A coleta automática de registros favorece o povoamento de repositórios poupando esforços que demandariam muito tempo para inserção de grandes quantidades de registros, tanto por parte da equipe, quanto dos pesquisadores. Com a coleta automática de registros, o Repositório Institucional UNESP têm conseguido ampliar suas fontes de coleta e desse modo, contemplar a produção do maior número de áreas.

Entretanto, cabe destacar que nas coletas automáticas e conversão de registros pode haver perdas de dados, principalmente no mapeamento de um padrão de metadados para outro. Outro problema identificado foi a falta de padronização dos dados e a necessidade de

inclusão novas informações nos registros para adequação do Perfil de Aplicação RI UNESP. A coleta dos arquivos em PDF e verificação das políticas editoriais dos artigos apresentaram-se como um dos processos mais demorados, por necessitarem de maior intervenção por parte da equipe do Repositório.

Considerando os resultados satisfatórios da coleta automática de registros em diferentes fontes e conversão por meio do uso de folhas de estilo para o povoamento do Repositório Institucional UNESP, espera-se que este trabalho continue contribuindo com as instituições que queiram realizar o povoamento de seus repositórios com coleta automática de registros.

Como trabalho futuro, o Repositório Institucional UNESP pretende identificar as principais fontes de referência para cada área, para ampliar o número de fontes para a coleta de registros e desse modo, espera-se que o maior número possível de publicações produzidas pela Universidade seja contemplada no repositório.

AGRADECIMENTOS

Os autores agradecem ao apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com os editais CNPq/PQ - Processo: 312544/2013-8, CNPq/MCTI CHSSA – Processo: 472058/2014-2, da Fundação para o Desenvolvimento da UNESP (FUNDUNESP), e todos envolvidos no desenvolvimento das atividades do Repositório Institucional UNESP em especial ao Vitor, Oberdan, Fabrício, Renata, Jaider, Bruna, Luiz, Monique e Milena.

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, 2002. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200001&lng=en&nrm=iso&tlng=pt>. Acesso em: 27 jul. 2016.

ARAKAKI, F. A.; SANTOS, P. L. V. A. da C.; ALVES, R. C. V. Panorama das pesquisas sobre o padrão de metadados Dublin Core no Brasil. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 20, n. 1, p. 86-97, jan./abr., 2015. Disponível em: <<https://revista.acbsc.org.br/racb/article/view/983>>. Acesso em: 27 jul. 2016.

ASSUMPCÃO, F. S. **Conversão de registros em XML para MARC 21: um modelo baseado em folhas de estilo XSLT**. 2013. 135 f. Dissertação (mestrado)-Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Filosofia e Ciências de Marília, 2013. Disponível em: <<http://hdl.handle.net/11449/93658>>. Acesso em: 27 jul. 2016.

ASSUMPÇÃO, F. S.; SANTOS, P. L. V. A. da C. Conversão de registros em XML para MARC 21: um modelo baseado em XSLT. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (ENANCIB), 15., 2014, Belo Horizonte. **Anais do XV Encontro Nacional de Pesquisa em Ciência da Informação...** Belo Horizonte: ECI/UFMG, 2014. p. 3802-3821. Disponível em: <<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>>. Acesso em: 26 jul. 2016.

ASSUMPÇÃO, F. S. et al. A conversão de registros na implantação de repositórios institucionais: o caso do Repositório Institucional UNESP. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 18., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014. p. 1-16. Disponível em: <<http://hdl.handle.net/11449/123645>>. Acesso em: 26 jul. 2016.

AVERKAMP, S.; LEE, J. Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository. **Code4Lib Journal**, Issue 7, 26 de jun. de 2009. Acesso em: <<http://journal.code4lib.org/articles/1647>>. Acesso em: 27 jul. 2016.

BASTOS, F. M. VIDOTTI, S. A. B. G. **Evolução do Repositório Institucional UNESP**. 2015. 8 Slides, Slide apresentado no evento Comemoração dos 6 anos da Biblioteca Digital e 2 anos do Repositório Institucional UNESP, São Paulo, dezembro de 2015.

DIMIĆ, B.; MILOSAVLJEVIĆ, B.; SURLA, D. XML schema for UNIMARC and MARC 21. **The Electronic Library**, v. 28, n. 2, p. 245-262, 2010. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/02640471011033611>>. Acesso em: 27 jul. 2016. DOI: 10.1108/02640471011033611

DULCINEA: derechos de explotación y permisos para el auto-archivo de revistas científicas españolas. 2008. Disponível em: <<http://www.accesoabierto.net/dulcinea/default.php?contenido=acerca>>. Acesso em: 27 jul. 2016.

FERRARI, R. D. D.; PIRES, G. de L. Auto-arquivamento e acesso aberto: deveres e direitos digitais na sociedade em rede. **RDBCI**, Campinas, SP, v. 12, n. 1, p. 22-38, fev. 2014. ISSN 1678-765X. Disponível em: <<http://www.sbu.unicamp.br/seer/ojs/index.php/rbci/article/view/3874>>. Acesso em: 29 mar. 2016.

FERREIRA, M. M. **MARC21**: formato condensado para dados bibliográficos. 2. ed. Marília: UNESP-Marília Publicações, 2002.

FLAMINO, A. N. **MARCXML**: um padrão de descrição para recursos informacionais em Open Archives. 2006. 164 p. Dissertação (mestrado)-Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2006. Disponível em: <<http://hdl.handle.net/11449/93694>>. Acesso em: 27 jul. 2016.

ILIK, V.; STORLIEN, J.; OLIVAREZ, J. Notes on operations metadata make over: transforming MARC records using XSLT. **Library Resources & Technical Services**. v. 58, n. 3, July 2014. p. 187. Disponível em: <<https://www.questia.com/library/journal/1G1-376680887/notes-on-operations-metadata-makeover-transforming>>. Acesso em: 26 jul. 2016.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA.

Diadorim: sistema de classificação das revistas. IBICT, 2016? Disponível em: <<http://diadorim.ibict.br/teste.jsp>>. Acesso em: 27 jul. 2016.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTES. **Universal Bibliographic Control and International MARC Core Programme**: UNIMARC: an introduction. EUA: IFLA, 1999. Disponível em: <http://www.ifla.org/files/assets/uca/publications/universal-bibliographic-control-and-international-marc-core-programme_1999.pdf>. Acesso em: 27 jul. 2016.

JOVANOVIĆ, J.; GASEVIĆ, D. Achieving knowledge interoperability: An XML/XSLT approach. **Expert Systems with Applications**. v. 29, 2005, p. 535–553. Disponível em: <<http://www.sciencedirect.com.ez87.periodicos.capes.gov.br/science/article/pii/S0957417405000710>>. Acesso em: 26 jul. 2016. DOI: 10.1016/j.eswa.2005.04.024

KEITH, C. Using XSLT to manipulate MARC metadata. **Library Hi Tech**, v. 22, n. 2, p. 122-130, 2004. Disponível em: <<http://dx.doi.org/10.1108/07378830410524549>>. Acesso em: 26 jul. 2016.

LEITE, F. et al. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: IBICT, 2012. Disponível em: <<http://livroaberto.ibict.br/handle/1/703>>. Acesso em: 16 jul. 2016.

LIBRARY OF CONGRESS. **MODS**: Metadata Object Description Schema. EUA: LC, 2016a. Disponível em: <<http://www.loc.gov/standards/mods/>>. Acesso em: 27 jul. 2016.

LIBRARY OF CONGRESS. **MARXML**: MARC 21 XML Schema. EUA: LC, 2016b. Disponível em: <<http://www.loc.gov/standards/marxml/>>. Acesso em: 27 jul. 2016.

MARCONI, M. A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 6. ed. São Paulo: Atlas, 2006.

MASAR, I. Bibliographic import from Scopus and WoS. [slides] In: OPEN REPOSITORIES 2014, Helsinki, Finland, June 9-13, 2014. **Presentation...** Disponível em: <<http://www.doria.fi/handle/10024/97747>>. Acesso em: 28 jul. 2016.

MENDES, R.; BORBINHA, J.; MANGUINHAS, H. Extracting Output Schemas from XSLT Stylesheets and Their Possible Applications. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 12., Malaysia, 2012. **Anais...** DCMI, EUA, 2012. Disponível em: <<http://dcpapers.dublincore.org/pubs/article/view/3666>>. Acesso em: 26 jul. 2016.

RAMOS DE CARVALHO, J. et al. Meta-information about MARC: an XML framework for validation, explanation and help systems. **Library Hi Tech**, v. 22, n. 2, p. 131-137, 2004. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/07378830410524558>>. Acesso em: 27 jul. 2016. DOI: 10.1108/07378830410524558

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

2016. DOI: 10.1108/02640470911004057

SANTAREM SEGUNDO, J. E. **Recursos tecno-metodológicos para descrição e recuperação de informações na Web**. 2004. 157 f. Dissertação (Mestrado em Ciência da Informação)–Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília. 2004. Disponível em: <<http://hdl.handle.net/11449/93618>>. Acesso em: 27 jul. 2016.

SHAFRANOVICH, Y. **Common Format and MIME Type for Comma-Separated Values (CSV) Files** [RFC4180]. October 2005. Informational. Disponível em: <<https://tools.ietf.org/html/rfc4180>>. Acesso em: 26 jul. 2016.

SIQUEIRA, M. A. **XML na Ciência da Informação: uma análise do MARC21**. 109 f. 2003. Dissertação (Mestrado em Ciência da Informação)-Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003. Disponível em: <<http://hdl.handle.net/11449/93720>>. Acesso em: 27 jul. 2016.

UNIVERSITY OF NOTTINGHAM. **About RoMEO**. UK, 2016. Disponível em: <<http://www.sherpa.ac.uk/romeo/about.php?la=en&fidnum=&mode=simple>>. Acesso em: 27 jul. 2016.

WALSH, M. P. Repurposing MARC Metadata for an Institutional Repository: Working with Special Collections and University Press Monographs. **Library Resources & Technical Services**, v. 55, n. 1 p. 33-44, 2011. Disponível em: <<http://hdl.handle.net/1811/47564>>. Acesso em: 26 jul. 2016.

WORLD WIDE WEB CONSORTIUM. **XSL Transformations (XSLT) Version 2.0: W3C Recommendation** 23 January 2007. Cambridge, 2007. Disponível em: <<http://www.w3.org/TR/xslt20>>. Acesso em: 26 jul. 2016.