



SNBU 2014
Belo Horizonte - MG

XVIII Seminário Nacional de
Bibliotecas Universitárias
16 a 21 de novembro

XVIII Seminário Nacional de Bibliotecas Universitárias

SNBU 2014

A CONVERSÃO DE REGISTROS NA IMPLANTAÇÃO DE REPOSITÓRIOS INSTITUCIONAIS: O CASO DO REPOSITÓRIO INSTITUCIONAL UNESP

Fabício Silva Assumpção
Renata Eleuterio da Silva
Jaider Andrade Ferreira
Flávia Maria Bastos



SNBU 2014
Belo Horizonte - MG

XVIII Seminário Nacional de
Bibliotecas Universitárias
16 a 21 de novembro

RESUMO

O Repositório Institucional UNESP foi criado em 2013 e, para sua implantação, foi povoado com dados obtidos de forma automática. Considerando a experiência realizada na UNESP, este trabalho tem por objetivo apresentar os processos utilizados para a conversão dos registros coletados de três diferentes fontes de dados (*Web of Science*, SciELO e *Scopus*) para inclusão no repositório. A partir da coleta dos registros, os padrões de metadados da *Web of Science*, da SciELO e da *Scopus* foram mapeados para o perfil de aplicação de metadados utilizado no repositório. Os registros foram coletados como arquivos XML e, para sua conversão, foram elaboradas folhas de estilo utilizando a linguagem XSLT. Após essa conversão, os arquivos XML foram convertidos em arquivos CSV e, então, importados no Repositório. Conclui-se que os processos de conversão utilizados permitiram alcançar as metas iniciais do Repositório e evitaram a necessidade de inclusão dos registros de forma manual.

Palavras-Chave: Repositório institucional; Conversão de registros; Folha de estilos XSLT; DSpace; Padrão de metadados.

ABSTRACT

The UNESP Institutional Repository was started in 2013 and, for its implementation, it was populated with data harvested automatically from three distinct sources (*Web of Science*, SciELO and *Scopus*). Based on UNESP's case, this paper aims to present the process used in the conversion of records from three distinct sources (*Web of Science*, SciELO and *Scopus*) for inclusion into the repository. After the harvesting, the *Web of Science*, the SciELO and the *Scopus* metadata standards were mapped to the metadata application profile used in the UNESP Institutional Repository. Since records were harvested as XML files, we created stylesheets using XSLT to transform these files in XML files in accordance with DSpace markup language and the application profile. From this point, the XML files were converted to CSV files in order to be imported into the Repository. We conclude that the conversion process allowed us to achieve the initial goals of Repository and to avoid the manual entering of records.

Keywords: Institutional repository; Record conversion; XSLT stylesheet; DSpace; Metadata standards.



1 Introdução

O desenvolvimento dos repositórios institucionais, de modo geral, tem ocorrido de forma “orgânica” a partir da criação de um ambiente digital designado para tal finalidade e da conscientização da comunidade usuária sobre a importância e a necessidade do depósito de sua produção em um repositório de acesso aberto. Assim, repositórios institucionais com um desenvolvimento “orgânico” estão presentes em diversos tipos de instituições, tais como as universidades, mesmo que, em muitos casos, não haja o autoarquivamento e a necessidade do depósito da produção não surja com a conscientização, mas com a obrigação.

Inserida num contexto global de universidades que visam à disseminação e à preservação de sua produção por meio de repositórios, a Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) iniciou em 2013 um projeto para a implantação do Repositório Institucional UNESP¹. Esse repositório, junto aos repositórios da Universidade de São Paulo (USP)² e da Universidade Estadual de Campinas (UNICAMP)³, faria parte do Repositório da Produção Científica do CRUESP (Conselho de Reitores das Universidades Estaduais Paulistas)⁴.

Para a inauguração do Repositório CRUESP, realizada em outubro de 2013, a UNESP definiu como objetivo incluir no Repositório o máximo possível da produção científica da Universidade. Assim, foi definida como meta inicial para a implantação a inclusão da produção científica de pesquisadores vinculados à Universidade indexada nas bases de dados *Web of Science* e *Scopus* e publicada em periódicos da *Scientific Electronic Library Online* (SciELO). Para alcançar essa meta de maneira eficiente, tendo em vista o prazo para a inauguração do Repositório CRUESP e do próprio Repositório Institucional UNESP, foram utilizados processos de coleta, de conversão e de importação automática dos registros referentes a essa produção científica.

Considerando a experiência realizada na UNESP, este trabalho tem por objetivo apresentar os processos utilizados para a conversão dos registros das três diferentes fontes de dados (*Web of Science*, SciELO e *Scopus*) para inclusão no Repositório Institucional UNESP.

¹ Disponível em: <<http://repositorio.unesp.br>>.

² Disponível em: <<http://producao.usp.br>>.

³ Disponível em: <<http://unicamp.sibi.usp.br>>.

⁴ Disponível em: <<http://cruesp.sibi.usp.br>>.



2 O Repositório Institucional UNESP

O Repositório Institucional UNESP surgiu a partir da Portaria UNESP número 88, de 28 de fevereiro de 2013, que instituiu o Grupo Gestor da Política do Repositório Institucional UNESP, “encarregado do desenvolvimento, implantação e manutenção do repositório institucional da universidade”, e definiu os propósitos do Repositório: “armazenar, preservar, divulgar e dar acesso à produção científica, acadêmica e administrativa da Universidade” (UNIVERSIDADE ESTADUAL PAULISTA, 2013, p. 47).

Entre os membros do Grupo Gestor e responsável pela coordenação executiva do projeto, está a Coordenadoria Geral de Bibliotecas da UNESP (CGB) na figura de sua coordenadora. Entre as atribuições da CGB está garantir a inclusão da produção no Repositório junto à Equipe Técnica, formada por bibliotecários e analistas de sistemas.

Durante a etapa de implantação, foram definidas quatro metas que refletiam o objetivo do Repositório em seu estágio inicial:

1. inclusão da produção científica institucional publicada no período de 2008 a 2012 e indexada na *Web of Science*⁵;
2. inclusão da produção científica institucional publicada em periódicos da SciELO⁶;
3. inclusão da produção científica institucional publicada no período de 1976⁷ a 2007 e indexada na *Web of Science*;
4. inclusão da produção científica institucional indexada na *Scopus*⁸.

Para a inauguração do Repositório, que ocorreu em outubro de 2013, apenas a primeira meta deveria ser alcançada, sendo que as demais poderiam ser alcançadas após a inauguração.

Considerando o prazo para a inauguração, a quantidade de documentos abrangidos pela primeira meta (aproximadamente 16.400 documentos distribuídos principalmente entre artigos e trabalhos publicados em anais de eventos) e a indisponibilidade de recursos humanos para a inclusão de forma manual, o Grupo Gestor e a Equipe Técnica buscaram estabelecer processos que possibilitassem a inclusão automática.

A partir desses processos, foi possível alcançar a primeira e a segunda metas antes de

⁵ Disponível em: <<http://wokinfo.com>>.

⁶ Disponível em: <<http://scielo.br>>.

⁷ O ano de 1976 foi definido como data limite para a coleta da produção científica da UNESP por ser o ano de criação desta universidade.

⁸ Disponível em: <<http://scopus.com>>.



SNBU 2014
Belo Horizonte - MG

XVIII Seminário Nacional de
Bibliotecas Universitárias
16 a 21 de novembro

outubro de 2013, de modo que, em sua inauguração, o Repositório contava com cerca de 28.400 registros, aproximadamente 16.400 provenientes da *Web of Science* e 12 mil da SciELO. Utilizando-se desses processos, foi possível alcançar a terceira e quarta metas em fevereiro e em maio de 2014, respectivamente.

Os processos utilizados para a inclusão dos dados de forma automática compreenderam diversas atividades, entre elas a coleta, a conversão e a importação dos registros. Para a atividade de coleta de registros da *Web of Science* foi utilizado seu *web service*⁹, que permitiu recuperar registros em *Extensible Markup Language* (XML) (Linguagem de Marcação Extensível). Os registros da SciELO também foram coletados em XML a partir da interface da própria SciELO. Já os registros da *Scopus*, também em XML, por não estarem disponíveis gratuitamente, foram adquiridos mediante compra. Os processos utilizados na conversão dos registros, que são o foco deste trabalho e têm em sua base os registros em XML, são descritos nas seções seguintes.

3 A conversão de registros utilizando folhas de estilo XSLT

Os registros das três fontes de dados (*Web of Science*, SciELO e *Scopus*) foram coletados como arquivos XML. Segundo o *World Wide Web Consortium* (W3C) (2012, 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”. Essa tecnologia, desenvolvida pelo W3C na década de 1990, é amplamente disseminada, relacionada a outras tecnologias e utilizada em uma diversidade de aplicações de informática.

Uma das tecnologias relacionadas à XML é a *Extensible Stylesheet Language for Transformation* (XSLT) (Linguagem Extensível para Folhas de Estilo de Transformação), linguagem utilizada na transformação de arquivos XML. Com essa linguagem são criadas folhas de estilo que permitem transformar arquivos XML estruturados com uma linguagem de marcação em arquivos XML estruturados de acordo com outras linguagens de marcação ou mesmo em arquivos que não sejam XML, por exemplo, arquivos de texto simples (W3C,

⁹ Um *web service* é um método de comunicação entre dois dispositivos eletrônicos em uma rede. O *web service* da *Web of Science* é um serviço sem custos adicionais, disponível aos assinantes da base de dados. Mais informações sobre o serviço podem ser encontradas em:

<http://wokinfo.com/products_tools/products/related/webservices>.

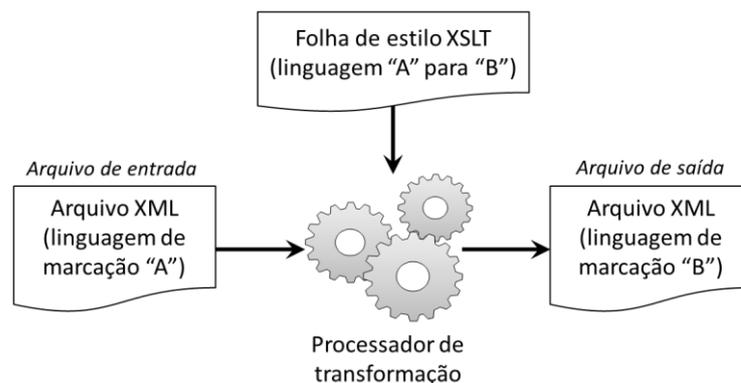


2007).

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.

O software responsável pela transformação, também chamado de processador de transformação, recebe o arquivo XML a ser transformado e a folha de estilo XSLT que contém as regras de transformação. Com base nas regras presentes na folha de estilo, o software realiza as transformações gerando um novo arquivo. Entre essas transformações estão desde o simples acréscimo ou a alteração de um dado até a reorganização de todo o conteúdo do arquivo de modo que o mesmo fique em acordo com outra linguagem de marcação. A transformação utilizando uma folha de estilo XSLT é ilustrada na Figura 1.

Figura 1: Transformação utilizando uma folha de estilo XSLT.



Fonte: Elaborada pelos autores.

Ao receber o arquivo XML de entrada, o processador de transformação consulta a folha de estilo e, com base em suas regras, gera o arquivo de saída, seja ele um arquivo XML ou não. Cabe notar que, ao realizar a transformação, o arquivo de entrada não é alterado, seu conteúdo é que é enviado para o arquivo de saída de acordo com as regras de transformação.

Cada uma das fontes de dados utilizadas (*Web of Science*, *SciELO*, *Scopus*) faz uso de uma linguagem de marcação diferente para a descrição de seus registros utilizando a XML. A Figura 2 ilustra as diferenças entre as três linguagens de marcação por meio de exemplos de como o título de um artigo é representado em cada uma delas.



Figura 2: Título de um artigo em um registro da *Web of Science*, da *SciELO* e da *Scopus*.

Web of Science	<pre><titles count="4"> <title type="source">PERSPECTIVAS EM CIENCIA DA INFORMACAO</title> <title type="source_abbrev">PERSPECT CIENC INF</title> <title type="abbrev_iso">Pespect. Cienc. Inf.</title> <title type="item">Decision making models and their relationship with organic information</title> </titles></pre>
SciELO	<pre><title-group> <article-title xml:lang="pt"><![CDATA[Modelos de tomada de decisão e sua relação com a informação orgânica]]></article-title> <article-title xml:lang="en"><![CDATA[Decision making models and their relationship with organic information]]></article-title> </title-group></pre>
Scopus	<pre><citation-title> <titletext original="n" xml:lang="eng">Decision making models and their relationship with organic information</titletext> <titletext original="y" xml:lang="por">Modelos de tomada de decisão e sua relação com a informação orgânica</titletext> </citation-title></pre>

Fonte: Elaborada pelos autores a partir de registros coletados da *Web of Science*, da *SciELO* e da *Scopus*.

Como mostrado na Figura 2, as linguagens de marcação das três fontes utilizam diferentes *tags* XML para representar o título do artigo: *title*, *article-title* e *titletext*. Observa-se também a utilização de diferentes normas para a indicação do idioma do título (atributo *xml:lang*): a *SciELO* utiliza a norma ISO 639-1 (dois caracteres: pt, en) e a *Scopus* utiliza a ISO 639-2 (três caracteres: eng, por), enquanto que a *Web of Science* não indica qual é o idioma do título. Há, nos registros da *Scopus*, uma indicação se o título é o título original do artigo (*original="y"*) ou se é um título traduzido (*original="n"*).

Para a descrição dos documentos no Repositório é utilizado um conjunto de metadados criado a partir do padrão de metadados *Dublin Core* e levando em consideração as necessidades de descrição em tal repositório e as possibilidades do software *DSpace*. Esse



conjunto de metadados é denominado Perfil de Aplicação de Metadados do Repositório Institucional UNESP.

Para os propósitos de importação e de exportação no Repositório, os registros criados com o Perfil de Aplicação podem ser representados de duas maneiras: como arquivos XML e como arquivos de texto simples com valores separado por vírgulas (*Comma-Separated Values*) (CSV). Na importação e na exportação via XML, é utilizada junto aos metadados do Perfil de Aplicação uma linguagem de marcação do próprio *DSpace*. Um fragmento de um registro criado com o Perfil de Aplicação e com a linguagem de marcação do *DSpace* é apresentado na Figura 3.

Figura 3: Registro criado com o Perfil de Aplicação e com a linguagem de marcação do *DSpace*.

```
<dublin_core schema="dc">
  <dcvalue element="contributor" qualifier="author" language="pt">Manziona, Rodrigo Lilla [UNESP]</dcvalue>
  <dcvalue element="contributor" qualifier="author" language="pt">Marcuzzo, Francisco Fernando
  Noronha</dcvalue>
  <dcvalue element="contributor" qualifier="author" language="pt">Wendland, Edson Cezar</dcvalue>
  <dcvalue element="date" qualifier="issued">2012-09-01</dcvalue>
  <dcvalue element="identifier">http://dx.doi.org/10.1590/S0100-204X2012000900022</dcvalue>
  <dcvalue element="language" qualifier="iso">pt</dcvalue>
  <dcvalue element="relation" qualifier="isPartOf" language="pt">Pesquisa Agropecuária Brasileira</dcvalue>
  <dcvalue element="subject" language="pt">água subterrânea</dcvalue>
  <dcvalue element="subject" language="pt">geoestatística</dcvalue>
  <dcvalue element="subject" language="en">groundwater</dcvalue>
  <dcvalue element="subject" language="en">geostatistics</dcvalue>
  <dcvalue element="title" language="pt">Integração de modelos espaciais e temporais para predições de níveis
  freáticos extremos</dcvalue>
  <dcvalue element="title" qualifier="alternative" language="en">Integration of spatial and temporal models to
  predict extreme water table depths</dcvalue>
  <dcvalue element="type" language="pt">Artigo</dcvalue>
</dublin_core>
```

Fonte: Elaborada pelos autores.

Uma vez que os registros coletados estavam em arquivos XML utilizando distintos padrões de metadados e distintas linguagens de marcação e que, para a importação no *DSpace*, era necessário obter os registros de acordo com o Perfil de Aplicação e com a linguagem de marcação do *DSpace*, optou-se pela utilização de folhas de estilo XSLT como meios para a conversão desses registros.

A utilização de folhas de estilo XSLT na conversão de registros está presente em alguns estudos que relatam os procedimentos e os instrumentos utilizados na conversão



(KEITH, 2004; KEENAN, 2010; KURTH; RUDDY; RUPP, 2004; RUDIĆ; SURLA, 2009) e no modelo para conversão de registros para os Formatos MARC 21 elaborado por Assumpção (2013).

Considerando que os registros coletados utilizavam linguagens de marcação diferentes, precisariam ser criadas três folhas de estilo, cada uma tendo como origem uma linguagem de marcação e como destino a linguagem de marcação do *DSpace*. A partir da escolha pela utilização de folhas de estilo XSLT, foram definidas e executadas as etapas do processo de conversão: mapeamentos, elaboração das folhas de estilo, conversão e importação no Repositório. Essas três etapas são apresentadas nas seções seguintes.

4 Mapeamentos

Para os propósitos deste trabalho, os elementos ou campos para a descrição apresentados nas linguagens de marcação da *Web of Science*, da *SciELO* e da *Scopus* são entendidos como metadados, enquanto que os conjuntos desses campos são entendidos como padrões de metadados. De forma semelhante, o Perfil de Aplicação é considerado um padrão de metadados no contexto do Repositório e seus campos são considerados metadados.

Considerando que a conversão utilizando folhas de estilo tem por base o conceito de “de - para” (de um metadado no padrão de origem para um metadado no padrão de destino), foi necessário mapear os metadados dos três padrões de origem para os metadados do Perfil de Aplicação (padrão de destino) para que as folhas de estilo pudessem ser elaboradas.

Esse processo de identificação de relacionamentos existentes entre metadados semanticamente equivalentes de diferentes padrões é denominado mapeamento ou *crosswalking* e resulta em um mapa (*crosswalk*) que provê 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).

A partir do mapeamento dos três padrões foram estabelecidos três mapas:

- *Web of Science* para o Perfil de Aplicação;
- *SciELO* para o Perfil de Aplicação;
- *Scopus* para o Perfil de Aplicação.

Esses mapas foram representados em quadros, como o exemplificado no Quadro 1.


Quadro 1: Fragmento do mapa “Web of Science para o Perfil de Aplicação”.

Web of Science (padrão de origem)	Perfil de Aplicação (padrão de destino)	Repetível	Observações
names/name/last_name names/name/first_name	dc.contributor.author	Sim	Unir os dois metadados separando-os utilizando uma vírgula.
addresses/address_name/ address_spec	dc.contributor.institution	Sim	
pub_info/@sortdate pub_info/@pubyear	dc.date.issued	Não	Utilizar o @pubyear apenas se o @sortdate não estiver presente no registro.
abstracts/abstract/abstract_text	dc.description.abstract	Sim	
fund_ack/grants/grant/grant_agency	dc.description.sponsorship	Sim	
identifiers/identifier[@type='issn']	dc.identifier.issn	Sim	
titles/title[@type='source']	dc.relation.isPartOf	Não	O título do periódico está todo em maiúsculas, corrigir deixando em maiúsculas apenas a inicial de cada palavra.
keywords/keyword	dc.subject	Sim	
titles/title[@type='item']	dc.title	Não	

Fonte: Elaborado pelo autor.



Como mostrado no Quadro 1, foram acrescentadas aos mapas informações que auxiliaram na elaboração das folhas de estilo, tais como:

- a indicação sobre a possibilidade de múltiplas ocorrências de um metadado, ou seja, se o metadado é repetível ou não;
- a indicação sobre como proceder quando o conteúdo de dois ou mais metadados do padrão de origem precisa reunido em um único metadado do padrão de destino;
- a indicação sobre como proceder quando há no padrão de origem mais de um metadado correspondente a um único metadado no padrão de destino (correspondência muitos-para-um);
- a indicação sobre as alterações necessárias ao conteúdo do metadado, por exemplo, a capitalização e a transformação de códigos em valores textuais e vice-versa.

5 Elaboração de folhas de estilo XSLT e conversão de registros

A partir dos mapas indicando as correspondências entre os metadados utilizados pelas distintas fontes de dados e pelo Perfil de Aplicação, foram elaboradas três folhas de estilo. Para essa atividade foi utilizada versão 2.0 da linguagem XSLT (W3C, 2007) e o Oxygen XML Editor¹⁰, software utilizado para a criação e a edição de documentos XML e que dispõe de processadores de transformação necessários à conversão dos registros.

Seguindo as correspondências indicadas nos mapas, os conteúdos de alguns metadados dos padrões de origem foram simplesmente enviados para os metadados do padrão de destino (Perfil de Aplicação). Nesses casos em que não foi necessária a alteração do conteúdo, foram incluídas na folha de estilo regras de transformação semelhantes à apresentada na Figura 4.

Figura 4: Regra de transformação em que não há alteração no conteúdo do metadado.

```
<xsl:for-each select="head/citation-info/author-keywords/author-keyword">
  <dcvalue element="subject">
    <xsl:value-of select="."/>
  </dcvalue>
</xsl:for-each>
```

Fonte: Elaborada pelos autores.

¹⁰ Disponível em: <<http://oxygenxml.com/>>.



A regra de transformação apresentada na Figura 4 indica ao processador de transformação: para cada ocorrência do metadado *author-keyword* em um registro da *Scopus*, crie um metadado *subject* e insira nesse metadado o conteúdo do metadado *author-keyword*. Com essa regra, cada palavra-chave presente no registro da *Scopus* foi inserida em um metadado *subject* do registro de destino.

Em alguns casos, no entanto, o conteúdo do metadado de origem não pôde ser simplesmente enviado para o metadado de destino, sendo necessária a adequação desse conteúdo durante a transformação de modo que ficasse de acordo com o padrão para o preenchimento dos campos estabelecido para o Repositório. Alguns desses casos incluíram: a devida capitalização das palavras que estavam com todas as letras em maiúsculo; a adequação dos códigos de idiomas e dos tipos de documentos; e a padronização do nome da UNESP, das principais instituições com as quais ela mantém parcerias e das principais agências de fomento/financiamento nacionais.

Para a padronização do nome da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por exemplo, foi dada ao processador de transformação a seguinte instrução: caso a agência de fomento presente no registro de origem seja “Fundação de Amparo à Pesquisa do Estado de São Paulo”, “*São Paulo Research Foundation*” ou “FAPESP”, insira no registro de destino “Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)” (forma padronizada para a inserção no Repositório).

Para garantir que o resultado da conversão estivesse de acordo com o Perfil de Aplicação e com o padrão para o preenchimento dos campos, durante e após a elaboração das folhas de estilo foram realizados testes para verificar a adequação dos registros convertidos aos resultados esperados.

Após a finalização de cada folha de estilo, que ocorreu em diferentes momentos da implantação do Repositório, os registros coletados em XML foram inseridos no processador de transformação. O processador de transformação, por sua vez, processou os registros e, com base nas regras de transformação da folha de estilo, converte-os, apresentando como resultado arquivos XML contendo registros de acordo com o Perfil de Aplicação e seguindo a linguagem de marcação do *DSpace*.



6 Conversão para CSV e importação

Após a conversão, os registros precisavam ser importados no Repositório. Entre os métodos para importação de registros disponíveis no software utilizado no Repositório, o *DSpace*, há a importação em massa por meio de arquivos de texto simples contendo valores separados por vírgula (CSV). Esse arquivos assemelham-se às planilhas dos programas *Microsoft Office Excel* e *LibreOffice Calc*, em que cada coluna representa um metadado e cada linha representa um registro.

Considerando a necessidade da importação de registros em massa no Repositório, foi decidido pela utilização de arquivos CSV como o meio para a inserção dos registros convertidos. Desse modo, os arquivos XML resultantes da conversão dos registros da *Web of Science*, da *SciELO* e da *Scopus* precisariam ser convertidos em arquivos CSV.

Como mencionado, as folhas de estilo XSLT podem ser utilizadas para a conversão de arquivos XML tanto em arquivos XML de acordo com outras linguagens de marcação, quanto em arquivos de outros tipos, por exemplos, arquivos de texto simples. Essa possibilidade fez com que as folhas de estilo fossem escolhidas também para a segunda conversão (de arquivos XML para arquivos CSV).

Cabe observar que a segunda conversão não incluiu alterações nos metadados e nem em seus conteúdos, podendo ser entendida apenas como uma reorganização dos registros para atender a uma necessidade de importação do *DSpace*.

Após a elaboração da folha de estilo e a segunda conversão, os registros foram então importados no Repositório. A partir da importação a Equipe Técnica prosseguiu com as demais atividades para a implantação do Repositório.

7 Considerações finais

O processo de conversão de registros para inclusão em repositórios institucionais, assim como todo o processo de implantação e de gestão desses repositórios, deve ser realizado de forma interdisciplinar, congregando, na medida do possível, conhecimentos de profissionais de distintas áreas, entre elas a Biblioteconomia e a Ciência da Computação.

A participação de bibliotecários no processo de conversão ocorre principalmente no estudo e no mapeamento dos padrões de metadados e na validação dos resultados obtidos com a conversão. Essas atividades requerem a participação de bibliotecários não apenas por



demandarem aspectos da disciplina de catalogação, disciplina presente no núcleo da Biblioteconomia, mas, também, pelos bibliotecários terem o conhecimento, ou a capacidade de obtê-lo, sobre o papel da produção científica nas universidades e, de forma mais ampla, sobre a comunicação científica e os agentes nela envolvidos (pesquisadores, leitores, publicadores, bases de dados, universidades, agências de fomento, etc.).

Apesar de terem seu papel definido, os bibliotecários não devem ter sua atuação limitada a essas atividades na conversão de registros. O que se observa, a partir da experiência da UNESP, é a necessidade dos bibliotecários ultrapassarem os limites tradicionalmente colocados pela Biblioteconomia em busca de conhecimentos de outras áreas que possam ser úteis ao desenvolvimento de suas atividades e a um reposicionamento da Biblioteconomia e da biblioteca universitária diante das necessidades das universidades no cenário informacional atual.

No caso do Repositório Institucional UNESP, os bibliotecários da Equipe Técnica puderam desempenhar algumas atividades que, por sua natureza, seriam delegadas aos analistas de sistemas, tais como a coleta dos dados pelo *web service* da *Web of Science* e a elaboração das folhas de estilo. Isso se deu, principalmente, em razão desses bibliotecários buscarem em outras áreas os conhecimentos necessários ao desenvolvimento dessas atividades.

Após a conversão e a importação dos registros, foram necessárias ações para a padronização, a inserção e a remoção de dados, considerando, principalmente, que nem todos os dados necessários ao Repositório puderam ser obtidos a partir dos registros coletados da *Web of Science*, da *SciELO* e da *Scopus*, e que alguns dados, os nomes dos autores e das instituições, por exemplo, nem sempre são indicados de forma consistente, sendo possível, em uma mesma fonte de dados, encontrar diferentes formas desses nomes. Essas ações, em sua maior parte, tiveram de ser realizadas manualmente pela Equipe Técnica.

Além dessas ações de padronização, de inserção e de remoção de dados, foram executadas pela Equipe Técnica ações para a verificação das condições de acesso (acesso aberto ou acesso restrito) e das permissões para o arquivamento (arquivamento em repositórios institucionais permitido ou não) e para a coleta dos documentos, nos casos em que o arquivamento em repositórios institucionais era permitido. Essas ações foram realizadas considerando uma característica essencial dos repositórios institucionais: prover o acesso aberto e preservar os documentos, não apenas seus registros ou metadados.



Após a inclusão da produção científica indexada na *Web of Science* e na *Scopus* e publicada em periódicos da SciELO, uma das próximas etapas do Repositório será a inclusão das teses e das dissertações defendidas na Universidade. Para essa inclusão serão reutilizados os registros no Formato MARC 21 para Dados Bibliográficos criados pela Rede de Bibliotecas da UNESP e também o processo de conversão por meio de folhas de estilo XSLT, sendo conduzidas as etapas de mapeamento e de elaboração de folhas de estilo de forma semelhante à descrita neste trabalho.

Além de ter possibilitado o alcance das quatro metas definidas para o estágio inicial do Repositório Institucional UNESP, poupando esforços manuais tanto dos pesquisadores quanto dos servidores, a inclusão dos registros de forma automática permite projetar para a comunidade da UNESP os benefícios da existência de um repositório institucional. Nesse sentido, é possível, então, fomentar na comunidade uma cultura de depósito da produção científica que auxilie no desenvolvimento do Repositório enquanto produto e serviço de informação que representa a Universidade.

Essa inclusão automática também se apresenta como um diferencial, não necessariamente exclusivo da UNESP, que favorece o convencimento dos gestores da Universidade quanto à necessidade de recursos específicos que possibilitem a continuidade do trabalho iniciado com a implantação do Repositório.

Considerando os resultados da conversão e da inclusão automática dos registros no Repositório Institucional UNESP, espera-se que este trabalho contribua com as instituições que almejam realizar o povoamento de seus repositórios por meio da coleta de dados de maneira automática.

Referências

ASSUMPCÃO, F. S. *Conversão de registros em XML para MARC 21: um modelo baseado em folhas de estilo XSLT*. 135 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013. Disponível em: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/assumpcao_fs_me_mar.pdf>. Acesso em: 14 maio 2014.

KEENAN, T. M. Why Purchase When You Can Repurpose? Using Crosswalks to Enhance User Access. *Code4Lib Journal*, n. 11, 2010. Disponível em: <<http://journal.code4lib.org/articles/3604>>. Acesso em: 14 maio 2014.



SNBU 2014
Belo Horizonte - MG

XVIII Seminário Nacional de
Bibliotecas Universitárias
16 a 21 de novembro

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: 14 maio 2014.

KURTH, M.; RUDDY, D.; RUPP, N. Repurposing MARC metadata: using digital project experience to develop a metadata management design. *Library Hi Tech*, v. 22, n. 2, p. 153-165, 2004. Disponível em: <<http://dx.doi.org/10.1108/07378830410524585>>. Acesso em: 14 maio 2014.

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: 14 maio 2014.

ST. PIERRE, M.; LAPLANT, W. P. *Issues in Crosswalking Content Metadata Standards*. Baltimore: National Information Standards Organization, 1998. Disponível em: <http://www.niso.org/publications/white_papers/crosswalk>. Acesso em: 14 maio 2014.

UNIVERSIDADE ESTADUAL PAULISTA. Portaria n.º 88, de 28 de fevereiro de 2013. *Diário Oficial do Estado de São Paulo*, Executivo, São Paulo, 01 mar. 2013. Caderno 1, p. 47.

W3C. *XML*. Last modified: 2012/01/24. Cambridge, 2012. Disponível em: <<http://www.w3.org/XML>>. Acesso em: 14 maio 2014.

W3C. *XSL Transformations (XSLT) Version 2.0*: W3C Recommendation 23 January 2007. Cambridge, 2007. Disponível em: <<http://www.w3.org/TR/xslt20>>. Acesso em: 14 maio 2014.

WOODLEY, M. S. Crosswalks, Metadata Harvesting, Federated Searching, Metasearching: Using Metadata to Connect Users and Information. In: BACA, M. (Org.). *Introduction to Metadata*. 2nd ed. Los Angeles: Getty Research Institute, 2008. Disponível em: <http://www.getty.edu/research/publications/electronic_publications/intrometadata/path.pdf>. Acesso em: 14 maio 2014.