

Reutilização de Metadados para o Povoamento de um Repositório Institucional: Procedimentos Aplicados no Repositório Institucional UNESP

Silvana Aparecida Borsetti
Gregorio Vidotti
UNESP – Univ Estadual
Paulista, Brasil
vidotti@reitoria.unesp.br

Flávia Maria Bastos
UNESP – Univ Estadual
Paulista, Brasil
fmbastos@reitoria.unesp.br

Juliano Benedito Ferreira
UNESP – Univ Estadual
Paulista, Brasil
julianoferreira@reitoria.unesp.br

Ana Paula Grisoto
UNESP – Univ Estadual
Paulista, Brasil
grisotoana@reitoria.unesp.br

Fabrcio Silva Assumpção
UNESP – Univ Estadual
Paulista, Brasil
fabricio@reitoria.unesp.br

Renata Eleutério da Silva
UNESP – Univ Estadual
Paulista, Brasil
renata_silva@marilia.unesp.br

Vitor Silvério Rodrigues
UNESP – Univ Estadual Paulista,
Brasil
vitorsrodrigues@reitoria.unesp.br

Oberdan Luiz May
UNESP – Univ Estadual
Paulista, Brasil
oberdan@reitoria.unesp.br

Palavras-chave: Reutilização de metadados; repositório institucional.

1. Introdução

Os esforços para a implantação de um repositório institucional na Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP) tiveram início em 2013, quando a Universidade optou pelo uso do software DSpace e definiu, a partir do padrão Dublin Core, um conjunto de metadados para alcançar suas necessidades no que diz respeito à produção científica de seus pesquisadores.

A meta inicial do Repositório Institucional UNESP (<http://repositorio.unesp.br>) era incluir os artigos publicados no período de 2008 a 2012 e indexados na Web of Science. Para alcançar essa meta, optou-se pela reutilização dos metadados já existentes na Web of Science para criar registros para importação no DSpace. Após o alcance dessa meta inicial, os metadados de outras bases de dados também foram reutilizados.

Este trabalho apresenta alguns procedimentos para a reutilização de metadados a partir da Web of Science, da SciELO, da Scopus e da Plataforma Lattes na criação de registros para importação no Repositório Institucional UNESP.

2. Coleta

Para coletar os metadados em um formato XML, foram utilizados diferentes procedimentos para cada base de dados: os metadados da Web of Science foram coletados utilizando o *web service* disponibilizado pela Web of Science; os metadados da SciELO foram coletados utilizando um software criado por um membro da equipe do Repositório; os metadados da Scopus foram comprados pela Universidade; e os metadados da Plataforma Lattes foram coletados utilizando outro software criado pela equipe. A ferramenta criada para coletar os metadados da SciELO também coletou os objetos digitais (ou seja, os arquivos PDF).

3. Conversão

As bases de dados utilizadas como fonte de dados possuem seus próprios padrões de metadados, assim, foi necessário mapear os metadados dos quatro padrões para os metadados do perfil de aplicação utilizado no Repositório e converter os registros de modo a obtê-los de acordo com esse perfil de aplicação. Uma vez que os registros foram coletados em XML, foram criadas folhas de estilo com a linguagem Extensible Stylesheet Language for Transformation (XSLT) para realizar a conversão. A conversão ocorreu em dois passos: (1) conversão do arquivo XML original em um arquivo XML de acordo com o perfil de aplicação e (2) conversão do arquivo XML de acordo com o perfil de aplicação em um arquivo CSV. Para os registros da Scopus e da Plataforma Lattes foi necessário um passo adicional antes do primeiro passo para juntar todos os arquivos XML coletados em um único arquivo XML.

4. Verificação

Durante a etapa de verificação, primeiramente foram removidos os registros duplicados. Para isso, os registros foram comparados entre si e com os registros já presentes no Repositório. Para a comparação foram utilizados o DOI e o título e ano de publicação juntos.

Após a remoção das duplicações, os registros foram verificados pela equipe para checar se a Universidade estava mencionada nos dados de afiliação dos autores, para corrigir erros e incluir os dados ausentes. Após essa verificação, foram verificadas também as permissões de acesso (acesso aberto ou acesso restrito) e de arquivamento do objeto digital (se o arquivamento em repositórios institucionais era permitido ou não). Nos casos em que o arquivamento era permitido, uma cópia do objeto digital era salva e nomeada com um ID obtido a partir do registro (o ID da Web of Science, da SciELO ou da Scopus, por exemplo).

Ao final desta etapa, foi executado um programa que distribuiu os registros entre as coleções do Repositório a partir das informações presentes nos metadados de afiliação e de autor. Esse programa, criado pela equipe, incluiu em uma coluna do arquivo CSV o código “handle” das coleções nas quais o registro deveria ser incluído.

5. Importação

O arquivo CSV verificado foi importado no DSpace de modo a inserir os registros no Repositório. Após a importação, um programa desenvolvido pela equipe incluiu cada objeto digital coletado em seu respectivo registro a partir da correspondência entre o nome do objeto digital e o ID presente no registro.

6. Considerações finais

Com a aplicação dos procedimentos apresentados neste trabalho, a reutilização de metadados permitiu o alcance de resultados positivos no Repositório Institucional UNESP: mais de 80 mil registros foram inseridos em cerca de um ano e meio. Esses procedimentos têm como principal característica os mapeamentos entre os padrões de metadados utilizados nas bases de dados (Web of Science, Scopus, SciELO e Plataforma Lattes) e o perfil de aplicação de metadados utilizado no Repositório criado a partir do padrão Dublin Core.

Por fim, com a demonstração dos procedimentos de reutilização dos metadados para importação de itens em um repositório institucional, este trabalho provê contribuições para as instituições que almejam aumentar as coleções de seus repositórios e, conseqüentemente, sua visibilidade acadêmica.