

Esquemas de metadados utilizados por repositórios digitais científicos e de objetos educacionais: estudo preliminar

Morgana Carneiro Andrade

Doutoranda em Tecnologias e Sistemas de Informação - Universidade do Minho, Guimarães, Portugal.

morgana@dsi.uminho.pt

Décio Wey Berti Junior

Doutorando em Ciência da Informação - Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

decowbj@gmail.com

Gleice Pereira

Departamento de Biblioteconomia da Universidade Federal do Espírito Santo

Email: gleiceufes@gmail.com

Ana Alice Baptista

Centro de Investigação ALGORITMI / Dep. de Sistemas de Informação
Universidade do Minho, Portugal

analice@dsi.uminho.pt

Introdução

Os repositórios digitais (RDs) auxiliam no gerenciamento e no armazenamento da produção intelectual de universidades e institutos de pesquisa. Podem armazenar diferentes tipos de materiais para finalidades e usuários distintos e têm como função o apoio à pesquisa, ao ensino e aos processos administrativos (JOINT INFORMATION SYSTEMS COMMITTEE, 2005). A denominação do repositório digital pode variar de acordo com o tipo de recurso armazenado, como é o caso dos repositórios digitais científicos (RDCs), direcionados para o armazenamento de publicações científicas de instituições de ensino, e dos repositórios de

objetos educacionais (ROEs) ou temáticos, voltados para uma determinada área de conhecimento. Para tanto, têm nos metadados o fator-chave para que essa função seja alcançada. Os metadados podem conter grande diversidade de informação e ser adotados para diferentes projetos. A estrutura de metadados desenvolvida para descrever determinados recursos em contextos específicos é denominada esquema de metadados (Chan; Zen, 2006).

Os RDs podem utilizar diferentes esquemas de metadados, elementos e valores para descrições dos seus recursos que podem interferir no nível de interoperabilidade semântica desses repositórios. A interoperabilidade semântica é definida, neste estudo, como uma funcionalidade que “[...] permite a agentes, serviços e aplicações trocar informação, dados e conhecimento de forma significativa tanto na Web como fora dela” (USHOLD, 2005). A interpretação inequívoca desses elementos pelos repositórios possibilita, ao pesquisador, o acesso aos variados tipos de repositórios, facilitando a pesquisa interdisciplinar e contribuindo para as abordagens multidisciplinares (Chan; Zen, 2006; Crow, 2002).

Objetivo

Identificar os esquemas de metadados utilizados para descrição de recursos em RDCs de objetos de aprendizagem. Objetivo justificado pela escassez de pesquisas com esse foco, embora esse tipo de estudo possibilite: o fomento da interoperabilidade semântica; o mapeamento de esquemas e seus elementos que permitem suportar desenhos de novos elementos de metadados; e o desenvolvimento de perfis de aplicação.¹

Metodologia

A pesquisa foi dividida em duas etapas. Na primeira, foram analisados 40 RDCs europeus e os 20 latino-americanos com maior número de depósitos. A diferença da amostra considerou a proporção de repositórios existentes nos dois continentes. Para tal seleção, utilizou-se o *site* openDoar, diretório mundial de repositórios científicos. Foram incluídos na pesquisa apenas RDCs, sendo excluídas as bibliotecas digitais. Os dados referentes aos esquemas de metadados utilizados pelos RDCs foram obtidos por meio de informações disponibilizadas nos seus *sites*, artigos, relatórios e descrições recuperadas via *Google* e OAI-PMH. Na segunda etapa, decidiu-se redimensionar a população de repositórios a serem estudados devido a dúvidas que surgiram ao se analisar a população inicial. Realizou-se um *Survey* com 2.153 repositórios científicos e de objetos educacionais, de um total de 2.720 repositórios digitais cadastrados no OpenDoar, em que se utilizou o questionário eletrônico SurveyMonkey. Neste caso, foram incluídos todos repositórios digitais cadastrados com contato, independentemente da sua tipologia e localização geográfica.

Resultados

Na altura de submissão deste resumo, os números eram os seguintes: 2.153 questionários enviados; 268 respondidos; 51 não entregues ou rejeitados. Os principais resultados (parciais) obtidos com este questionário demonstram que: 223 são institucionais, 25, temáticos; 11 de objetos educacionais; 25 de teses e dissertações; e 16, organizacionais. Os repositórios organizacionais foram incluídos na pesquisas pelo fato deles armazenarem materiais científicos. Alguns repositórios se identificaram como pertencentes a mais de uma categoria. Foram considerados predoninantes os seguintes esquemas de metadados: Dublin Core, 170; DIM, 8; ETDMS, 24; LOM, 11; MARC, 17; METS, 28; MODS, 24; MPEG-21, 4; OAI-DC, 84. Foi possível identificar também que alguns repositórios utilizam mais que um esquema de metadados. Tipos de documentos: livros; artigos científicos; teses; projetos de graduação; objetos de aprendizagem; patentes; relatórios; dados de pesquisa; documentos históricos; materiais audiovisuais e *softwares*. Também foi possível identificar que 34 dos 263 gestores dos repositórios afirmam possuir perfil de aplicação (PA). Alguns gestores disponibilizaram o *link* do PA e outros forneceram informações insuficientes ou não perceberam a questão. Em vista desse resultado, sugeriu-se que *sites* como OpenDoar e OpenRoar possibilitem a disponibilização de *links* para os perfis de aplicação existentes e que os repositórios indiquem os *links* dos seus perfis de aplicação nesses *sites*, o que facilitaria o acesso às informações em relação aos esquemas de metadados utilizados e outras informações relevantes para a interoperabilidade semântica.

Conclusão

O estudo mostra que alguns RDCEs utilizam mais de um esquema de metadados para descrever os diferentes tipos de recursos que armazenam. Sem surpresa, confirma-se que o esquema mais utilizado é o Dublin Core simples.

Futuros trabalhos são necessários no que diz respeito: a) à análise dos elementos de metadados utilizados pelos repositórios pesquisados com o objetivo de propor perfil de aplicação para repositórios digitais; b) ao estudo dos perfis de aplicação existentes como futura base para proposição de perfil de aplicação direcionado para repositórios digitais científicos. Em relação às estratégias adotadas para coleta de dados (análise do *site*, OAI-PMH e questionário), percebeu-se que nem sempre os dados são coincidentes. Em um trabalho futuro, será necessário utilizar outras técnicas de recolha de dados para efeitos de triangulação.

Palavras-chave: Metadados. Repositórios digitais científicos. Interoperabilidade semântica.

Referências

CHAN, L. M.; ZENG, M. L. (2006) – Metadata interoperability and standardization: a study of methodology Part I. Achieving interoperability at the schema level. *D-Lib Magazine*, v. 12, n. 6, June 2006. [Consult. 12 maio 2014]. Disponível na Internet: URL:

<<http://www.dlib.org/dlib/june06/zeng/06zeng.html>>

CROW, R. (2002) – *Sparc*. Institutional repository checklist & resource guide. Washington: SPARC

JOINT INFORMATION SYSTEMS COMMITTEE (2005) – *Digital repositories: helping universities and colleges. Briefing paper: higher education sector*. Bristol: JISC

USHOLD, M. et al. *Semantic integration & interoperability using RDF and OWF*. W3C Editor's Draft 3 November 2005. [Consult. 20 Maio 2014] Disponível na Internet: URL

<<http://www.w3.org/2001/sw/BestPractices/OEP/SemInt/>>

ⁱ Ver <http://dublincore.org/usage/documents/profiles/index.shtml>