

A CONTRIBUIÇÃO DA ORGANIZAÇÃO DO CONHECIMENTO PARA A INTEROPERABILIDADE SEMÂNTICA: ALTERNATIVAS PARA REPOSITÓRIOS INSTITUCIONAIS¹

LA CONTRIBUCIÓN DE LA ORGANIZACIÓN DEL CONOCIMIENTO PARA LA INTEROPERABILIDAD SEMÁNTICA: ALTERNATIVAS PARA LOS REPOSITORIOS INSTITUCIONALES

Morgana Carneiro de Andrade - morganaandrade@gmail.com
Mestre em Gestão da Informação - Universidade Estadual de Londrina
Doutoranda em Tecnologia e Sistemas de Informação – Universidade do Minho (Portugal)
Bibliotecária da Universidade Federal do Espírito Santo

Brígida Maria Nogueira Cervantes- brigidacervantes@gmail.com
Doutora em Ciência da Informação pela Faculdade de Filosofia e Ciências da
Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de Marília
Professora do Departamento de Ciência da Informação da UEL.

RESUMO

Introdução: Os repositórios institucionais tornam-se relevantes para as instituições de pesquisa e ensino por reunirem o conhecimento produzido em um único local, facilitando a gestão da produção intelectual e a sua socialização. Entretanto, como pré-requisito para cumprir esse papel, é necessário que exista interoperabilidade semântica, não só entre os repositórios como em relação a outros serviços de informação.

Objetivo: Analisar o elemento de metadado *Subject* do *Dublin Core (dc.subject)*, sob a perspectiva da organização do conhecimento para fornecer subsídios para Repositórios Institucionais.

Metodologia: Pesquisa documental e descritiva, com abordagem quali-quantitativa, empregando o método de estudos de casos múltiplos. Para a coleta de dados, foram utilizadas uma planilha como roteiro de pesquisa, que orientou a identificação das informações registradas nos *sites* dos repositórios selecionados, e o questionário eletrônico EncuestaFacil.

Resultados: Identificou-se que o padrão de metadados usado pela totalidade desses repositórios é o *Dublin Core*. Apontou para alguns fatores, como a falta de controle de vocabulário na maioria dos

repositórios – que poderia ser evitada com a disponibilização de instrumentos que facilitam a identificação de descritores pelos depositantes; o baixo índice de descritores traduzidos para o inglês e a ausência de exposição de vocabulários controlados e de esquemas de classificação.

Conclusões: Conclui-se que o panorama atual apresentado pelos repositórios institucionais nacionais necessita de um alinhamento com iniciativas internacionais em prol da interoperabilidade semântica. É essencial que os procedimentos adotados por gestores e pelo pessoal envolvido com os RIs sigam boas práticas em relação aos metadados e à Organização do Conhecimento.

Palavras-chave: Interoperabilidade semântica. Metadados. Repositório institucional. Organização do conhecimento.

¹ Resultado parcial de dissertação de mestrado, defendida em 2012.

1 INTRODUÇÃO

A globalização, a complexidade existente no mundo em relação às diversas atividades desenvolvidas por vários setores e o aumento da quantidade de conhecimento necessário para enfrentar os constantes desafios científicos e tecnológicos levam a parcerias internacionais, em que o compartilhamento de informações e documentos de forma rápida e eficaz se torna essencial. Os repositórios digitais, na concepção de vários autores, são vistos como uma possibilidade de otimizar esse compartilhamento (CROW, 2002; MARCONDES; SAYÃO, 2009).

O repositório institucional (RI), um dos tipos de repositórios digitais, apresenta-se como uma perspectiva de reunir, organizar, preservar, divulgar e garantir o acesso confiável e permanente da produção intelectual gerada na universidade em um único local (LEITE, 2009). Contudo, para que a preservação, a divulgação e o acesso a esses documentos sejam possíveis, é necessário que os repositórios possam “conversar” entre os diversos sistemas de informação existentes na Rede. O estabelecimento desse diálogo é viabilizado por vários tipos de interoperabilidade que buscam o desenvolvimento de regras e padrões (BOTERAM, 2010; PATEL et al., 2005; ZENG, 2010),

A interoperabilidade permite a integração de processos bibliográficos e documentais heterogêneos entre diferentes sistemas de comunicação, ou seja, que serviços de busca e recuperação de informação direcionados aos usuários sejam disponibilizados a partir de sistemas que integram *hardwares*, *softwares*, estruturas de dados e interfaces diferentes, de forma mais eficaz (NATIONAL INFORMATION STANDARDS ORGANIZATIONS, 2004; SAYÃO; MARCONDES, 2008).

Há vários níveis de interoperabilidade – técnica, semântica, organizacional, política e humana, intercomunitária, legal e internacional –, mas este estudo optou por explorar a interoperabilidade semântica na perspectiva da Organização do Conhecimento, por entender que a possibilidade de usuários pesquisarem por uma única interface, utilizando uma linguagem comum ao acessar diversas bases de dados e coleções ao mesmo tempo, proporciona a otimização e o aumento da qualidade de acesso e recuperação por diferentes comunidades.

A interoperabilidade semântica no contexto deste estudo é interpretada como

[...] a capacidade dos sistemas de informação (bibliotecas digitais e repositórios institucionais) para intercambiar informações baseando-se em um significado comum dos termos e expressões contidas nos metadados e documentos, com o objetivo de assegurar a consistência, a representação e a recuperação dos conteúdos. Isto envolve o uso de esquemas semânticos que incluem vocabulários controlados [...] (GÓMEZ DUEÑAS, 2012, p. 3, tradução nossa).

De acordo com McCulloch (2004), a prevalência de diferentes terminologias e as limitações causadas aos usuários na recuperação das informações têm levado instituições e pesquisadores a procurar uma solução para esse problema. O uso de diversas terminologias compromete a interoperabilidade semântica entre os sistemas, o que não é desejável em um ambiente onde os usuários buscam obter resultados consultando várias fontes simultaneamente. Boteram (2010) acrescenta, ainda, que não apenas a abundância de informação e a heterogeneidade de sistemas utilizados para o acesso aos documentos comprometem a recuperação de informações pelos usuários, mas também questões relacionadas aos diferentes idiomas, vocabulários e esquemas de organização da informação (OI).

A falta de padronização terminológica tende a prejudicar a consecução da interoperabilidade semântica entre os repositórios e também com outros sistemas de informação, em razão das inconsistências semânticas – diferentes terminologias e significado – e dos esquemas de metadados.

No caso do esquema de metadados Dublin Core, o elemento de metadado *Subject* (*dc.subject*) e os qualificadores utilizados para a organização do conhecimento contribuem para a potencialização da interoperabilidade semântica entre repositórios institucionais.

A necessidade de estabelecer um padrão de interoperabilidade semântica para o RI da Universidade Federal do Espírito Santo (RIUFES), especificamente em relação aos valores do elemento de metadado *Subject* (Assunto) do Dublin Core (*dc.subject*), norteou a escolha do tema deste estudo, que, inicialmente, utilizou a pesquisa bibliográfica. Em um segundo momento, foi desenvolvido um estudo de casos múltiplos envolvendo RIs contemplados pelo projeto-piloto do IBICT/FINEP e pelos Editais IBICT/FINEP/PCAL/XBDB n. 02/2009 e 03/2009, que totalizam 38 repositórios de instituições de ensino/pesquisa. Analisou-se, por meio de questionário e análise dos *sites* dos repositórios institucionais brasileiros contemplados pelos editais do IBICT/FINEP, os metadados descritivos e qualificadores utilizados para a organização do conhecimento, de forma a fornecer subsídios para a potencialização da interoperabilidade semântica, no nível de dados categorizados.

2 REVISÃO DE LITERATURA

De acordo com Barité (2001), a Organização do Conhecimento (OC) busca oferecer subsídios para o tratamento da informação, em especial, o tratamento temático da informação, assim como para a gestão do uso social da informação. Broughton et al. (2005) acrescentam que a OC representa os sistemas que organizam o conhecimento (SOC), tais como registros bibliográficos, sistemas de classificação, tesauro, redes semânticas e processos de organização do conhecimento (classificação e descrição de documentos, catalogação descritiva, indexação e análise de assunto).

Os SOC e os serviços oferecidos a partir do seu uso estão relacionados à interoperabilidade semântica (PARK; TOSAKA, 2010; TUDHOPE; NIELSEN, 2006). Eles têm a habilidade de conectar os recursos existentes em bibliotecas digitais com outros itens. A base dessa conexão está na identificação da informação inserida (metadados) nos recursos digitais, que podem ser extraídos e utilizados para pesquisa e localização dentro de um SOC (HODGE, 2000).

De acordo com a NISO (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004, p. 1), os

Metadados são informações estruturadas que descrevem, explicam, localizam e, ainda, facilitam a recuperação, o uso ou o gerenciamento de recursos de informação. Constantemente, metadados são denominados de dados sobre dados ou informação sobre informação (Tradução nossa).

No entendimento de Relvão (2003, p. 5), os metadados aumentam o potencial informativo dos objetos, por se tratar de dados que fornecem informação ou documentação de outros dados dentro de uma aplicação ou um ambiente, podendo conter informação descritiva em relação a contexto, qualidade, condição ou características dos dados.

Existem vários padrões de metadados que atendem às necessidades das comunidades e às especificidades dos objetos. No caso dos repositórios institucionais, é utilizado o padrão Dublin Core (DC), que propicia a interoperabilidade semântica, flexibilidade e modularidade na *Web*. Há um consenso internacional da comunidade científica em relação ao emprego desse padrão na obtenção da interoperabilidade (PEREIRA; RIBEIRO JÚNIOR; NEVES, 2005).

Os 15 elementos básicos do Dublin Core podem ser implementados pelos usuários ou pelo sistema, e cada um possui um valor (ex.: Esquemas de codificação Vocabulary Encoding Scheme - VES e Syntax Encoding Scheme - SES) (CATARINO, 2009). Eles têm uma ocorrência ilimitada e a forma de distinguir seus valores é a utilização de qualificadores. Pode haver um identificador (esquema) e/ou um modificador que informa, aos usuários e aos programas, como o valor daquele elemento deve ser interpretado. Como exemplo, há o elemento Assunto (*Subject*) e o qualificador palavras-chave (LCSH). A utilização dos sistemas de organização do conhecimento (SOC) aumentam as especificidades dos dados (DUBLIN CORE METADATA INICIATIVE, 2011; HODGE, 2000; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2001, 2004).

O fato dos metadados darem um significado semântico aos dados, por meio de procedimentos técnicos como catalogação, indexação e categorização dos conteúdos informacionais, permite que haja uma integração entre sistemas heterogêneos de informação (SOUZA; ALVES, 2009). Para que o acesso e a recuperação dos conteúdos digitais sejam eficazes, é necessário caracterizar os documentos por meio da padronização da descrição dessas informações (CASTRO; SANTOS, 2007).

A partir dessa padronização, tornam-se possíveis a obtenção e a potencialização da interoperabilidade semântica, cujo principal objetivo é apoiar processamentos

complexos e avançados de consulta aos recursos em um contexto de informações heterogêneas.

A interoperabilidade semântica está relacionada não só com a forma com que o dado pode ser entendido por humanos e por programas de computador, mas com a diversidade de idioma; a sintática (diferenças na codificação e na representação da informação) e os esquemas de metadados. Entende-se que os elementos de metadados representam as propriedades dos recursos, e que recurso é o objeto/item e pode conter descrições mais específicas, atribuídas por meio de qualificadores que adicionam valores. Por sua vez, os qualificadores são termos que, associados ao elemento, fornecem informações adicionais mais precisas em relação ao significado do elemento (BAPTISTA; MACHADO, 2001; SOUZA; ALVES, 2009). Esse nível de interoperabilidade contém as linguagens documentárias utilizadas nas áreas de Ciência da Informação (CI).

De acordo com as Diretrizes DRIVER 2.0 (DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH, 2009, p. 11), para se obter uma recuperação simultânea do registro e do texto integral, é preciso o desenvolvimento de serviços avançados e integrados, como a “[...] pesquisa por assuntos combinada com a navegação através de classificações, análise de citações [...]”.

Neste estudo, são abordadas as recomendações do DRIVER em relação aos elementos de metadados do Dublin Core, *Subject*. Essas diretrizes aconselham a utilização de vocabulários e semântica, com o objetivo de reduzir a ambiguidade dos conceitos e termos empregados na comunicação científica, e a adoção do idioma inglês para informação descritiva, com indicação também do idioma português, aspecto sugerido pela autora (BAPTISTA, 2010).

O elemento *Subject* retrata o assunto do recurso, geralmente expresso por palavras-chave, frases ou códigos de classificação. Pode-se utilizar um ou outro, mas, se houver dois, eles devem ser registrados separadamente. Recomenda-se evitar o uso de palavras genéricas para descrever o recurso.

De acordo com Baptista (2010), os esquemas de classificação adotados no contexto OAI são Classificação da Biblioteca do Congresso Americano (LCC), CDD e CDU. Já os sistemas de cabeçalhos de assuntos são LCSH e Schlagwortnormdatei (SWD). Os metadados OAI contêm códigos de áreas a partir de esquemas como Mathematics Subject Classification (MSC) e MeSH.

Uma das observações do DRIVER (DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH, 2009, p. 120) é que esses esquemas devem ser expostos pelo RI. Com o uso desses elementos de classificação, torna-se possível implementar serviços de navegação por classificação, substituir “[...] códigos de classificação por termos em inglês, tradução de termos para outros idiomas e fazer uma junção dos códigos de classificação utilizando regras de mapeamento”.

De acordo com Tudhope, Kock e Heery (2006), a utilização dos SOC melhora a interoperabilidade semântica, e esse tipo de interoperabilidade no ambiente de bibliotecas digitais está relacionada ao compartilhamento consistente do uso de terminologias, de modo que a interoperabilidade no seu nível mais alto ou domínio possa ser obtida por meio de negociação e aceitação dos conceitos compartilhados, dos termos e seus significados. Nesse sentido, os serviços de terminologia desempenham papel importante para facilitar o emprego consistente e coerente de termos comuns e sua semântica (PATEL et al., 2005).

De acordo com Koch (2006), um número reduzido de repositórios e de agregadores de serviços adota medidas para harmonizar a heterogeneidade das informações ou oferece serviços que facilitem a pesquisa e a navegação por assunto ou tópicos para a recuperação dos recursos.

3 INTEROPERABILIDADE SEMÂNTICA E A ORGANIZAÇÃO DO CONHECIMENTO: RESULTADOS DA PESQUISA

A pesquisa com os RIs foi realizada após aprovação pelo Comitê de Ética da Universidade Estadual de Londrina e com a assinatura do Termo de Consentimento Livre Esclarecido (TCLE). Do universo de 38 repositórios selecionados, só foram analisados 26 RIs. Dois dos RIs não foram identificados, seis não responderam e quatro justificaram a não participação na pesquisa em razão de os repositórios se encontrarem em fase de implantação ou mudança de coordenação.

Os resultados obtidos com a pesquisa foram analisados levando em consideração os aspectos apontados na literatura consultada. A seguir, serão explicitadas algumas questões avaliadas.

3.1 Esquema de Metadados

Dos 26 (100%) respondentes, 12 (38%) utilizam o esquema de metadados Dublin Core Qualified; 5 (16%) Dublin Core simplificado; 1 (3%) adota o *contex_object*; 1 (3%), o MARC; 3 (9%), o METS; 6 (20%) não souberam responder. Três (10%) repositórios responderam que adotam o Dublin Core (entende-se, mediante análise do *site*, que se trata do Dublin Core simplificado).

O Dublin Core possui algumas características que o tornaram um dos mais utilizados pela comunidade internacional (NAGAMORI; SUGMOTO, 2006): é descritivo e codificado em XML, o que permite uma melhor interoperabilidade (SHINTAKU; FERREIRA; ROBREDO, 2011); interage com o RDF (BAPTISTA; MACHADO, 2001); atende ao modelo *Linked Data*; aumenta a visibilidade de recursos em coleções entre os setores e domínios de assuntos (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2001).

Com o conjunto de elementos de descrição do DC, é possível que catalogadores ou não catalogadores utilize-o para simples descrição de recursos de informações, tornando-as mais visíveis aos motores de busca e sistemas de recuperação (WEIBEL, 1997). Podem ser reproduzidos de forma ilimitada utilizando qualificadores que informam o usuário sobre como cada item pode ser interpretado (DUBLIN CORE METADATA INITIATIVE, 2011).

3.2 Administração do RI

A administração dos RIs é, na sua maioria, realizada pela biblioteca (17, 61%), enquanto 5 (18%) dos RIs utilizam a administração compartilhada – biblioteca e outro órgão –, 6 (21%) indicaram que o RI é administrado por outro órgão da universidade.

Do total de respondentes (26), identificou-se que o depósito dos recursos realizado pelos autores com controle do profissional da informação corresponde a 4 (12%); em 1 (3%) o depósito é realizado pelo autor; 12 (35%) recebem dos autores e o depósito é feito por profissionais da informação; em 12 (35%) a coleta é feita pela administração e o depósito efetuado por profissionais da informação. Cinco (15%) respondentes indicaram outras possibilidades, como: os recursos são coletados e depositados por bolsistas de iniciação científica; a coleta e o controle de qualidade são realizados pelo administrador da comunidade; e o depósito é feito por pessoa indicada pelo administrador ou por outros órgãos responsáveis pelo acervo.

Um outro aspecto abordado pela questão indicou que a maior parte da coleta e da submissão dos itens é realizada pelo pessoal administrativo, corroborando a pesquisa de Melero et al. (2009). O índice de autoarquivamento ficou muito abaixo dos encontrados por Darby et al. (2008) em pesquisa realizada com RIs no Reino Unido. A ausência do autoarquivamento diverge dos princípios norteadores do movimento de acesso aberto aos RIs (via verde) (KURAMOTO, 2011; MARCONDES; SAYÃO, 2009).

Dados internacionais mostram que apenas 15% a 20% dos depósitos são praticados dessa forma (ENABLING OPEN SCHOLARSHIP, [2010]). A realidade brasileira é observada por Kuramoto (2011), que alerta sobre o baixo índice de depósito.

3.3 Metadados

Dezessete (65%) dos 26 RIs indicaram que a administração controla ou gerencia os metadados; 4 (15%) fazem esse controle de forma parcial, 1 (4%) não participa; e 4 (15%) não souberam responder.

O padrão de conteúdo de metadados utilizado por 14 (54%) dos repositórios é o AACR2/RDA; 3 (12%) utilizam padrão interno; 1 (4%) adota o Dublin Core Application Profile (DCAP) e 7 (27%) não sabem.

Já em relação ao padrão de conteúdo de metadados, 14 (54%) repositórios utilizam o AACR2/RDA, embora a análise de registros pelo *sítes* de alguns RIs tenha constatado a não utilização do AACR2/RDA nos metadados de título e autor, e 3 (12%) adotam padrão interno. O alto índice de utilização do AACR2/RDA corrobora o resultado da pesquisa de Park e Tosaka (2010). O uso desses padrões está em consonância com as orientações da NISO (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2007), entretanto a utilização de um padrão interno contraria as considerações feitas por Nagamori e Sugimoto (2006), Park e Tosaka (2010) no que diz respeito à interoperabilidade.

Para Souza e Alves (2009), esses padrões possibilitam a integração e o compartilhamento de informações e aplicações pelos sistemas de informação. A NISO (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2007) e Park e Tosaka (2010) despertam para a importância de desenvolver a explicitação do perfil de aplicação para que a comunicação possa ser estabelecida entre as diferentes comunidades.

3.4 Indexação dos Documentos

O controle de vocabulário ainda tem um índice baixo. Dos 25 respondentes, 2 (7%) informaram que não é feita a indexação de assunto; 8 (28%) disseram que a atribuição de palavras-chave é livre e utiliza o idioma português; 7 (24%) permitem livre atribuição das palavras-chave e adotam tanto o idioma do documento como o inglês; 6 (21%) afirmaram que são atribuídas palavras-chave ou sistema de classificação e idioma do país; em 4 (14%) as palavras-chave ou sistema de classificação são atribuídas ao idioma do país e em inglês. Foi informado ainda que as palavras-chave são atribuídas pelo autor em outros idiomas e não apenas no idioma do documento; e são retiradas do documento, empregando-se o inglês, o português e o idioma do documento.

Diante do crescimento no número de RIs e, conseqüentemente, dos depósitos, principalmente praticados pelo autoarquivamento, quando não existe controle da padronização de terminologias ou vocabulários e identificação dos valores dos metadados, há um prejuízo na obtenção da interoperabilidade semântica entre os vários sistemas de informação. A tradução dos termos é um fator relevante para a recuperação em relação ao contexto global. A simples reutilização das palavras-chave adotadas pelas publicações também pode ser um fator prejudicial na recuperação, já que nem todas as publicações empregam o vocabulário controlado para indexação dos documentos.

A falta de controle de vocabulário por parte dos RIs brasileiros foi abordada pela Professora Malgorzata Lisowska, durante a Segunda Conferência de Acesso Aberto em 2011, no Rio de Janeiro, ao apresentar o relato “Estrategia regional y marco de interoperabilidad y gestión para una Red Federada Latinoamericana de Repositorios Institucionales de documentación científica”. Lisowska exteriorizou sua preocupação em relação a esses aspectos para ações futuras que contemplam os repositórios nacionais.

Existe, por parte de vários organismos internacionais (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2001, 2004; DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH, 2009) e autores (BAPTISTA, 2010; PATEL et al., 2005; PARK; TOSAKA, 2010; PEREIRA; RIBEIRO; NEVES, 2005), orientação para que sejam utilizados vocabulários controlados e o SOC adotado seja exposto, assim como o uso do idioma inglês (DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH, 2009). O desenvolvimento

de vocabulário controlado para um domínio específico gera preocupação para alguns autores (PARK; TOSAKA, 2010). Para Castro e Santos (2011), o tratamento das informações deve visar à interoperabilidade entre os sistemas de informação.

Zeng e Chan (2004) veem nos SOC a possibilidade da consulta “*one stop*”, ou seja, o acesso a informações por uma única opção de busca, consultando diferentes sistemas e informações heterogêneas. Nesse sentido, Botheram (2010) afirma que a interoperabilidade semântica está estritamente relacionada aos SOC.

Em relação à responsabilidade de atribuir as palavras-chave, identificou-se que, dos RIs que responderam a essa questão, 19 (46%) indicaram os autores; 18 (44%) citaram bibliotecários; 1 (2%) indicou pessoal administrativo e 3 (7%) explicaram: autores indicam os termos e os bibliotecários revisam com o objetivo de identificar termos autorizados; o recurso é inserido por responsável pelo depósito com revisão do bibliotecário; ou utilizam os termos do próprio documento. Fica demonstrado um equilíbrio na responsabilidade da atribuição das palavras-chave, mas, quando comparado com os índices de indexação de assunto, pode-se deduzir que parte dos termos atribuídos pelos autores não é checada pelos bibliotecários.

Vinte e seis RIs responderam à questão sobre controle de autoridade de assunto, em que a CDU (11%) aparece como o sistema de classificação mais utilizado, em seguida vem a CDD (8%) e a LCC (3%), adotada apenas por um repositório. O vocabulário controlado mais utilizado na área das Ciências da Saúde é o DECS (11%), demonstrando que não existe uma tradução dos termos para o inglês, o que corresponde aos índices sobre o controle de vocabulário (Tabela 7). Os vocabulários controlados empregados pela maioria dos RIs são o da Biblioteca Nacional (24%) e da Bibliodata (8%), enquanto o LCSH (5%), que refletiria a tradução dos termos em idioma inglês, é adotado por apenas dois repositórios.

A utilização desses vocabulários controlados está em conformidade com a maioria dos trabalhos citados neste estudo, porém, ao analisar os *sites*, apenas um repositório adota a identificação e a exposição do vocabulário utilizado, como orientam as diretrizes DRIVER e autores como Crow (2002), Zeng e Chan (2004), Koch (2006). Ao expor os vocabulários, os RIs facilitam a utilização de mecanismos e serviços de terminologia que melhoram a interoperabilidade semântica. Outra vantagem é recorrer a esquemas de classificação, como tesouros e taxonomias, que proporcionam a melhoria na qualidade de acesso aos assuntos (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2007).

Vinte e cinco RIs responderam à questão sobre a existência de instrumentos ou procedimentos que auxiliam a descrição dos recursos, demonstrando uma heterogeneidade no tipo de SOC utilizado pelos repositórios. Entretanto, não foi mencionada a existência de tabelas de classificação (CDU, CDD, LCC), folksonomia, ontologia ou indexação automática, e nove repositórios não adotaram nenhum SOC. Eis a questão, eles usam CDU/CDD internamente, mas não disponibilizam esses instrumentos para auxiliar na descrição, ou seja, são utilizados apenas pelos bibliotecários.

Quando ofertados ao usuário, tanto no processo de busca como de depósito, esses instrumentos viabilizam a harmonização terminológica necessária para a interoperabilidade semântica (KOCH, 2006; MCCULLOUGH, 2004), como a possibilidade de desenvolvimento de ontologias a partir desses instrumentos (GOMES; GUIMARÃES, 2012). Não são identificadas facilidades, como a disponibilização de instrumentos, que ajudam a identificar termos empregados pelos sistemas, conforme pode ser conferido no menu de alguns repositórios.

A utilização de termos livres atende ao que Koch (2006) chama de preservação da riqueza das fontes participantes, ou seja, a linguagem natural inserida pelo autor pode ser conservada, mas deve-se incentivar a opção por essa linguagem ser traduzida para uma linguagem controlada. As diretrizes Driver 2.0 apresentam recomendações de como isso pode ser feito (DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH, 2009).

4 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

A interoperabilidade semântica, no que concerne à recuperação de conteúdo, é um dos aspectos mais discutidos por profissionais e estudiosos de várias áreas, como Ciência da Informação, Informática e Linguística.

Aqui buscou-se identificar como a organização do conhecimento pode contribuir para esse tipo de interoperabilidade. Em relação aos repositórios, a interoperabilidade semântica pode se referir a dois aspectos: à semântica dos elementos e à semântica dos valores. Neste estudo, o foco foi a semântica dos valores, relacionada ao uso de elementos da organização do conhecimento. Assim, a interoperabilidade semântica

abordada está baseada no significado comum dos termos contidos nos metadados e nos documentos.

Nessa perspectiva, a OC, por meio dos SOC, permite que os sistemas trabalhem não apenas com o controle terminológico, mas com a possibilidade de utilizar a linguagem natural para a extração de termos controlados, proporcionando a interoperabilidade semântica em nível de dados categorizados.

A heterogeneidade das informações existentes nos repositórios pode ser administrada pelos gestores a partir do estabelecimento de boas práticas em relação aos metadados e da criação e da disponibilidade de perfis de aplicação, como o Scholarly Work Application Profile (SWAP).

No contexto nacional, assim como ocorre em outros países, a interoperabilidade semântica ainda não atingiu o nível aceito como ideal. A partir dos objetivos propostos, é possível destacar os resultados provenientes da análise da literatura e os obtidos pela pesquisa desenvolvida com os 26 RIs que participaram deste estudo: a) os metadados descritivos utilizados pelos RIs permitem a identificação do título (elemento *Title*), autor (elemento *Creator*), do assunto (elemento *Subject*) e do resumo (elemento *Description*). O número de RIs que adotam o controle de vocabulário e a tradução dos termos para o inglês, considerado relevante por ser um idioma mundialmente aceito, ainda é baixo; b) elementos da representação temática, como sistemas de classificação e vocabulários controlados, são utilizados, porém não explicitados como valores/qualificadores do metadado *Subject*, o que favoreceria a descoberta dos recursos, assim como o uso de serviços de terminologias, otimizando esse tipo de interoperabilidade; c) instrumentos da representação temática, como vocabulário controlado, tesouro esquema de categorização, taxonomia e lista de descritores de assunto, foram indicados pelos administradores, porém poucos repositórios os disponibilizam para o usuário, o que poderia facilitar a utilização de termos adotados no sistema pelos depositantes e por aqueles que buscam os recursos; e d) a utilização dos SOC possibilita o desenvolvimento de serviços de terminologia, que, por sua vez, potencializam a interoperabilidade semântica.

Questões práticas e de qualidade em relação ao emprego de palavras-chave, assunto e classificações são preocupantes, não apenas pela prática do autoarquivamento, quando esse não é seguido de práticas que viabilizam o controle terminológico, mas também pelos procedimentos adotados pelos profissionais da

informação. A disponibilidade de instrumentos e mecanismos baseados na necessidade de recuperação da informação deveria ter em vista a harmonização de terminologias e a utilização de serviços de terminologias que podem agregar valor aos conteúdos dos repositórios por meio de uso dos SOC.

Vale salientar que a adoção dos SOC resulta em algumas possibilidades, como a utilização de *add-on* – recurso extra do DSpace – pelos repositórios, facilitando o depósito e a busca, além do desenvolvimento de ontologias.

Estudos revelaram que muitas práticas estão sendo adotadas em nível internacional e devemos avaliar essas orientações, procurando empregá-las mediante as condições de cada repositório, mas pensando sempre no conjunto.

Com base no que foi exposto e nas ações já adotadas pelo RiUFES, recomenda-se que os administradores do repositório disponibilizem ferramentas que propiciem a identificação e a utilização de termos controlados empregados no repositório, tanto pelos usuários depositantes como para os utilizadores, buscando sempre a padronização e a harmonização dos termos, e desenvolvam o perfil de aplicação.

REFERÊNCIAS

BAPTISTA, Ana Alice. A falar nos entendemos: a interoperabilidade entre repositórios digitais. In: GOMES, Maria João; ROSA, Flávia (Org.). **Repositórios institucionais: democratizando o acesso ao conhecimento**. Salvador: EDUFBA, 2010. p. 71-90.

BAPTISTA, Ana Alice; MACHADO, Altamiro Barbosa. Um gato preto num quarto escuro: falando sobre metadados. **Revista de Biblioteconomia de Brasília**, Brasília, v. 25, n. 1, p. 77-90, 2001.

BARITÉ, Mario. Organización del conocimiento: um nuevo marco teórico-conceptual em bibliotecología y documentación. In: CARRARA, Kester (Org.). **Educação, universidade e pesquisa**. Marília: UNESP, 2001. p. 35-60.

BOTERAM, Felix. "Content architecture". Semantic interoperability in an international comprehensive knowledge organization system. **Aslib Proceedings**, London, v. 62, n. 4-5, p. 406-414, 2010.

BROUGHTON, Vanda et al. Knowledge organization. In: KAJBERG, Leif; LORRING, Leif (Org.). **European curriculum reflections on library and information science education**. Copenhagen: Royal School of Library and Information Science, 2005. p. 133-148. Disponível em: <<http://biblis.db.dk/uhtbin/hyperion.exe/db.leikaj05>>. Acesso em: 2 jul. 2011.

CASTRO, Fabiano Ferreira; SANTOS, Plácida L. V. A. C. A consistência da informação descritiva em repositórios digitais: caminhos para a interoperabilidade. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24., Maceió, 2011. **Anais...** Maceió, 2011. Disponível em: <<http://www.febab.org.br/congressos/index.php/cbbd/xxiv/paper/view/327>>. Acesso em: 20 set. 2011.

_____. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da Web Semântica. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 22., 2007, Brasília. **Anais...** Brasília: ABDF/UNB, 2007.

CATARINO, Maria Elisabete. **Integração das folksonomias nos metadados: identificação de novos elementos como contributo para a descrição de recursos em repositórios**. 2009. Tese (Doutorado em Tecnologias e Sistemas de Informação) – Departamento de Sistemas de Informação, Universidade do Minho, 2009.

CROW, Raym. **Sparc**. Institutional repository checklist & resource guide. Washington: SPARC, 2002.

DARBY, Rob et al. Increasing the productivity of interactions between subject and institutional repositories. **New Review Information Networking**, United Kingdom, v. 14, p. 117-135, 2008.

DIGITAL REPOSITORY INFRASTRUCTURE VISION FOR EUROPEAN RESEARCH. **Directrizes DRIVER 2.0**: directrizes para fornecedores de conteúdos: exposição de recursos textuais com o protocolo OAI-PMH. Portugal: Serviços Documentação da Universidade do Minho, 2009. Disponível em: <http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final__PT.pdf>. Acesso em: 5 out. 2010.

DUBLIN CORE METADADA INICIATIVE. DCMI. **Metadada Terms**: Section 4: Vocabulary encoding schemes. Disponível em: <<http://dublincore.org/documents/dcmi-terms/#H4>>. Acesso em: 2 fev. 2011.

ENABLING OPEN SCHOLARSHIP. **Formulating an institutional open access policy**. Liège: *Université* of Liège, [2010].

GOMES, H.E.; GUIMARAES, L. S. Organização do conhecimento e recuperação da informação na era das tecnologias digitais. **Bibliotextos**, 15 mar. 2012. Disponível em: <<http://bibliotextos.files.wordpress.com/2012/03/organizac3a7c3a3o-do-conhecimento-e-recuperac3a7c3a3o-da-informac3a7c3a3o-na-era-das-tecnologias.pdf>>. Acesso em: 10 mar. 2012.

GÓMEZ DUEÑAS, Laureano Felipe. **Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales**: caso Biblioteca Digital Colombiana BDCOL. Disponível em: <http://eprints.rclis.org/handle/10760/14878#.T57tRLOm_NU>. Acesso em: 2 mar. 2012.

HODGE, Gail. **Systems of knowledge organization for digital libraries**: beyond traditional authorities files. Washington, DC: CLIR, 2000. Disponível em: <<http://www.clir.org/pubs/reports/pub91/contents.htm>>. Acesso em: 24 maio 2008.

KOCH, Traugott. Subject access and semantic interoperability in repositories. In: HEERY, Rachel et al. (Org.). **An evaluation study on the development and implementation of community repositories to support research and learning and teaching**. 2006. p. 38-49. Disponível em: <<http://delos-wp5.ukoln.ac.uk/project-outcomes/WP-5-0-1/d-5-1-1-report.pdf>>. Acesso em: 12 mar. 2011.

KURAMOTO, Helio. **Peter Suber fala sobre o estágio atual das iniciativas OA**. Disponível em: <<http://kuramoto.blog.br/2011/07/07/peter-suber-fala-sobre-o-estagio-atual-das-iniciativas-oa/>>. Acesso em: 7 jul. 2011.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: Ibict, 2009.

MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Introdução: repositórios institucionais e livre acesso. In: SAYÃO, Luis et al. (Org.). **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009.

MCCULLOCH, Emma. Multiple terminologies: an obstacle to information retrieval. **Library Review**, Bradford, v. 53, n. 6, p. 1723-1737, 2004.

MELERO, Remedios et al. The situation of *open access* institutional repositories in Spain: 2009 report. **Information Research**, Sweden, v. 14, n. 4, p. 1-21, Dec. 2009.

NAGAMORI, Mitsuharu; SUGIMOTO, Sshigeo. Metadata schema registry as a tool to enhance metadata interoperability. **TCDL Bulletin**, v. 3, n. 1, 2006. Disponível em: <<http://www.ieee-tcdl.org/Bulletin/v3n1/nagamori/nagamori.html>>. Acesso em: 5 jan. 2011.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **A framework of guidance for building good digital collections**. 3. ed. Bethesda: NISO, 2007.

_____. **The dublin core metadata element set**. Bethesda: NISO Press, 2001.

Disponível em:

<http://www.niso.org/apps/group_public/download.php/6578/The%20Dublin%20Core%20Metadata%20Element%20Set.pdf>. Acesso em: 9 set. 2011.

_____. **Understanding metadata**. Bethesda: NISO, 2004.

PARK, Jung-Ran; TOSAKA, Yuji. Metadata creation practices in digital repositories and collections: schemata, selection criteria, and interoperability. **Information Technology and Libraries**, Chicago, v. 29, n. 3, p. 104-116, 2010. Disponível em: <http://www.ala.org/ala/mgrps/divs/lita/ital/292010/2903_sep/park.cfm>. Acesso em: 5 jun. 2011.

PATEL, Manjula et al. **Semantic interoperability in digital library systems**. Bath: Network of Excellence on Digital Libraries, 2005. (Project n.º 507618 DELOS).

PEREIRA, Ana Maria; RIBEIRO JUNIOR, Divino Ignácio; NEVES, Guilherme Luiz Cintra. Metadados para a descrição de recursos da internet: as novas tecnologias desenvolvidas pra o padrão Dublin Core e sua utilização. **Revista ACB: Biblioteconomia em Santa Catarina**, Florianópolis, v. 10, n. 1, p. 241-249, jan./dez. 2005.

RELVÃO, Rafael. **Elementos de metadados para a aprendizagem à distância**. 2003. 47 f. Relatório de Projeto (Mestrado em Sistemas de Informação) – Departamento de Sistemas de Informação, Universidade do Minho, Guimarães, 2003.

SAYÃO, Luís Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 2, p. 133-148, 2008.

SHINTAKU, Milton; FERREIRA, Sueli Mara Mara Soares Pinto; ROBREDO, Jaime. Repositórios brasileiros implementados com DSpace. In: ENCONTRO NACIONAL DE ENSINO E PESQUISA EM INFORMAÇÃO., 10., Salvador, 2011. **Anais...** Salvador, 2011. Disponível em: <<http://www.cinform2011.ici.ufba.br/modulos/submissao/Upload/37558.pdf>>. Acesso em: 2 mar. 2012.

SOUZA, Maria Izabel Fugisawa; ALVES, Maria das Dores Rosa. Representação descritiva e temática de recursos de informação no sistema agência Embrapa: uso do padrão Dublin Core. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 7, n. 1, p. 190-205, 2009.

TUDHOPE, Douglas; KOCH, Traugott; HEERY, Rachel. **Terminology services and technology. JISC state of the art review**. 2006. Disponível em: <http://www.jisc.ac.uk/media/documents/programmes/capital/terminology_services_and_technology_review_sep_06.pdf>. Acesso em: 2 fev. 2011.

TUDHOPE, Douglas; NIELSEN, Marianne Lykke. Introduction to knowledge organization systems and services. **New Review Hypermedia and Multimedia**, London, v. 12, n. 1, p. 3-9, June 2006.

WEIBEL, Stuart. The Dublin core: a simple content description model for electronic resources. **Bulletin of the American Society for Information Science and Technology**, Silver Spring, v. 24, n. 1, p. 9-11, Oct./Nov. 1997.

ZENG, Marcia Lei. Linguagens de marcação específicas por domínio e metadados descritivos: funções para a descoberta de recursos científicos. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. esp., jul. 2010. Disponível em: <<http://br.vlex.com/vid/linguagens-especificas-descritivos-descoberta-264811162>>. Acesso em: 2 maio 2011.

ZENG, Marcia Lei; CHAN, Lois Mai. Trends and issues in establishing interoperability among knowledge organization systems. **Journal of the American Society for Information Science and Technology**, v. 55, n. 5, p. 377-395, 2004.

Title

The contribution of knowledge organization for semantic interoperability: alternatives for the Institutional Repositories

Abstract

Introduction: The institutional repositories are alternatives for the research and education institutions. They optimize the scientific and technological advance and assemble the generated knowledge in a single spot which facilitates the intellectual production management and its socialization. However, as the requirements for the repositories to fulfill their role is to enable the semantic interoperability, not only among repositories, but also concerning other information services.

Objective: This study aimed to analyze the Dublin Core metadata element set – Subject (dc.subject) based on the perspective of the knowledge organization in order to provide support for the institutional repository.

Methodology: Descriptive and documentary research with approach quali-quantitative and the method adopted was the multiple case study. For the data collection were used a research script which contains aspects that identify the repositories and a spreadsheet that guided the identification of the information recorded on the sites of the selected repositories and EncuestaFacil electronic query.

Results: The pattern of metadata adopted by all those repositories is Dublin Core. The research point out to some problems such as: the lack of vocabulary control in most repositories that could be avoided by making the instruments to identify the descriptors accessible to people in putting data; small rate of descriptors translated into English as well as absence of controlled vocabulary or classification schemas.

Conclusion: This study demonstrates that current situation presented by the national institutional repositories needs alignment with international initiatives in favor of semantic interoperability. It is necessary that managers and staff who are involved in the IRs adopt good practices regarding the metadata and the knowledge organization.

Keywords

Semantic interoperability. Metadata. Institutional repository. Knowledge organization.

Título

La contribución de la Organización del Conocimiento para la Interoperabilidad Semántica: alternativas para los Repositorios Institucionales

Resumen

Introducción: Los repositorios institucionales se vuelven relevantes para las instituciones de investigación e enseñanza a la vez que el conocimiento que se produce en un solo lugar, facilitando la gestión de la producción intelectual y la socialización. Sin embargo, como un requisito previo para cumplir esta función, debe haber interoperabilidad semántica no sólo entre los repositorios como en relación con otros servicios de información.

Objetivo: Analizar el elemento de metadatos: *Asunto* del Dublin Core (dc.subject), desde el punto de vista de la organización del conocimiento para proveer subsidios para los Repositorios Institucionales.

Metodología: Investigación documental y descriptiva, con abordaje cuali-cuantitativo, empleando el método de estudio de casos múltiples. Para la recopilación de los datos, se utilizó una hoja de cálculo como de ruta de investigación, que guiaron la identificación de las informaciones registradas en los sitios web de los repositorios seleccionados y el cuestionario electrónico EncuestaFacil.

Resultados: Se identificó que el estándar de metadatos utilizado por todos los repositorios analizados es el Dublin Core. Se refirió a algunos factores, como la falta de control del vocabulario en la mayoría de los repositorios – que podrían evitarse con la disponibilidad de herramientas que faciliten la identificación de los descriptores por los depositantes–; la baja tasa de descriptores traducidos al inglés y la falta de exposición vocabularios controlados y esquemas de clasificación.

Conclusiones: Se concluye que la situación actual de los repositorios institucionales nacionales necesita una alineación con las iniciativas internacionales en favor de la interoperabilidad semántica. Es esencial que los procedimientos adoptados por los administradores y por el personal involucrado con los RIs contengan las mejores prácticas con relación a los metadatos y con la Organización del Conocimiento.

Palabras clave: Interoperabilidad semántica. Metadatos. Repositorio Institucional. Organización del Conocimiento.

Recebido em: 15/12/2012

Aceito em: 15/01/2013