

Definição de metadados para registros de áudio em repositórios digitais de acesso aberto

Maria Imaculada da Conceição, Maria do Carmo Avamilano Alvarez, Hálida Cristina Rocha Fernandes

Resumo

Introdução: A Biblioteca/CIR da Faculdade de Saúde Pública da Universidade de São Paulo (FSP/USP) está desenvolvendo um repositório digital para a área da saúde pública, e possui uma coleção de áudios produzidos pela própria instituição, denominada Audioteca. O objetivo deste trabalho é definir metadados para registros de áudio em repositórios digitais de acesso aberto, garantindo que essas informações tenham maior visibilidade no extenso universo da internet. **Método:** Foram selecionados registros de áudio das bases Dedalus e LILACS e identificados os metadados dessas bases, que utilizam os formatos MARC21 e Metodologia LILACS, respectivamente. Em seguida, estabeleceu-se um escopo mínimo de metadados no esquema Dublin Core (DC), que correspondiam aos metadados dos formatos anteriores. Em uma etapa final, utilizou-se a possibilidade de adaptação do DC para a criação dos metadados com base em estudo da Library of Congress, além de refinamentos que atendiam as necessidades específicas da instituição e da coleção Audioteca, entre eles o entrevistador, o editor de som, e termos técnicos tais como o áudio *bits*. **Resultados:** O resultado deste trabalho é um formulário específico para os materiais de coleções de áudio, mantendo a compatibilidade e interoperabilidade do esquema Dublin Core. **Conclusão:** Dar visibilidade às coleções de áudios, que podem ser veiculadas por programas de rádios comunitárias e educativas, vai ao encontro do papel de promoção de pesquisas das instituições e contribui para a construção da cidadania.

Palavras-chave

Audioteca. Metadados. Bibliotecas digitais. Repositórios digitais.

Introdução

A sociedade atual exige inovadores ambientes de trabalho para lidar efetivamente com uma carga cada vez maior de informações heterogêneas. A produção e divulgação dos dados e novos conhecimentos também ocorrem de forma rápida e ampla (CANDELA *et al.*, 2010).

Com o objetivo de dar acesso a esse conhecimento que é produzido, bibliotecas e repositórios digitais crescem em tamanho, número e diversidade (TANI; CANDELA; CASTELLI, 2013) com conteúdos que podem ser armazenados localmente e acessados remotamente por meio da rede de computadores e de forma simultânea por diversos usuários. Além disso, há uma multiplicidade de suportes e uma informação pode estar em um livro, artigo, filme ou arquivo de áudio.

Para a descrição e representação dos recursos informacionais utilizam-se metadados, que segundo a definição de Jordan (2006), podem ser entendidos como a informação estruturada que descreve ou que se encontra associada individualmente a itens da coleção. Essa definição é ampla e leva em consideração as informações, descrevendo uma estrutura de objeto digital, e o histórico de como seus arquivos constituintes têm sido migrados para novos formatos. Descrever o item e seus atributos básicos (título, autor e origem) é tarefa que as bibliotecas têm feito por um longo tempo, sendo tão importante para uma coleção digital quanto para coleções impressas. Entretanto, documentos digitais são diferentes dos impressos, sendo necessário o apropriado uso dos tipos de metadados para seu gerenciamento.

A adoção adequada de padrão de metadados propicia a interoperabilidade entre aplicações

e o compartilhamento de dados entre sistemas (ALVES; SOUZA, 2007). Um dos padrões de metadados mais utilizado em repositórios digitais é o Dublin Core (DC), que pode ser definido como sendo o conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos. Ele permite a descrição do acervo e pode ser acrescido de adaptações introduzidas diante da necessidade de adequar-se a especificidades institucionais (SOUZA *et al.*, 2000). Ainda sobre o DC, Marcondes e Sayão (2002) ressaltam que é resultado de intenso trabalho de discussão e padronização em nível internacional, mantida por um ativo grupo e fórum internacionais, e é usado em diferentes sistemas, inclusive na Open Archives Initiative.

Dentro desse contexto, a Biblioteca/CIR da Faculdade de Saúde Pública da Universidade de São Paulo (FSP/USP) está desenvolvendo um repositório digital para a área da saúde pública, e possui uma coleção de áudios produzidos pela própria instituição, denominada Audioteca. Os arquivos de áudio são gerados a partir de dissertações e teses defendidas na universidade e são registrados em bases de dados. Os registros das teses são associados a arquivos de áudio (*spots* de rádio), cuja finalidade, segundo Gallo (2001), é ampliar o acesso a essa informação por radialistas ou comunicadores de rádio e inseri-los nas programações diárias das rádios comunitárias e educativas. Dessa forma, a informação é disseminada em linguagem acessível ao público em geral.

O objetivo deste trabalho é definir metadados para registros de áudio em repositórios digitais de acesso aberto, garantindo que essas informações tenham maior visibilidade no extenso universo da internet.

Procedimentos metodológicos e resultados

O Repositório Digital da FSP/USP utiliza a plataforma DSpace e o padrão de metadados Dublin

Core com adaptações, entre elas, a criação de um conjunto de metadados de interesse específico da instituição e que não ficam visíveis para o usuário final.

A coleção Audioteca fará parte do acervo do Repositório Digital da FSP/USP. Essa coleção foi iniciada em 1999 e é formada atualmente por 166 registros¹ (em mp3), a maioria dos quais integrados por dois áudios: um depoimento e uma entrevista, ambos com no máximo 5 minutos de duração. As teses e dissertações que dão origem ao material gravado estão depositadas no acervo impresso da Biblioteca/CIR da FSP/USP e na forma digital nos portais de teses, de acesso gratuito, o que garante ao usuário consultar o texto completo.

Para a definição dos registros de áudio no repositório digital, inicialmente se realizou um levantamento dos metadados dos registros atuais que a Biblioteca utiliza para a coleção Audioteca, tanto no Dedalus – que é o Banco de Dados Bibliográficos da USP – como na base de dados LILACS², com a qual a Biblioteca/CIR colabora enviando registros. A base LILACS é matida pela BIREME³. É importante destacar que as duas bases utilizam padrões de metadados diferentes.

O Dedalus utiliza o formato MARC 21 – *MAchine-Readable Cataloguing* – da Library of Congress, que é um conjunto de códigos e designações de conteúdos definido para codificar registros que serão interpretados por máquina. Sua principal finalidade é possibilitar o intercâmbio de dados, ou seja, importar dados de diferentes instituições ou exportar dados de uma instituição para outros sistemas ou redes de bibliotecas através de programas de computador desenvolvidos especificamente para isto (MARANHÃO; MENDONÇA, 2010).

A base LILACS utiliza o formato de descrição bibliográfica desenvolvido pela BIREME e que se fundamenta no “*Reference Manual for Machine-readable Bibliographic Description*” do Unisist/

¹ Consulta realizada em 01/11/2013.

² Literatura Latino-Americana e do Caribe em Ciências da Saúde.

³ Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde.

UNESCO, com adaptações baseadas no MARC21 Lite (BIREME, 2008). A utilização desses metadados é parte da Metodologia LILACS.

Inicialmente, foram selecionados alguns registros de áudio das bases Dedalus e LILACS para o estudo comparativo dos metadados nos formatos MARC21, LILACS e Dublin Core (Quadro 1). Para essa comparação foi utilizado o DCMÍ metadata terms (DUBLIN CORE METADATA INITIATIVE, 2012).

A etapa seguinte consistiu na definição dos metadados do padrão Dublin Core pertinentes aos registros de áudio e, na etapa final, na elaboração do formulário para entrada de dados de registros

de áudio em repositórios digitais de acesso aberto (Quadro 2).

Recursos em áudio e vídeo são menos usuais em bibliotecas do que documentos textuais. Tanto o seu processamento quanto sua preservação, ou mesmo a definição de formato para entrega final ao usuário, ainda não são consenso (JORDAN, 2006). Ao disponibilizar a informação e o conteúdo da coleção Audioteca em um repositório digital, a Biblioteca/CIR sentiu a necessidade de novos metadados para uma melhor descrição, administração e preservação desse tipo de recurso.

Assim, a partir da identificação dos metadados utilizados, verificou-se que o DC atendeu a maio-

Quadro 1 – Comparação e equivalência de campos MARC 21, Metodologia LILACS e Dublin Core, utilizados nos registros analisados

MARC 21 (DEDALUS)	Metodologia LILACS	Dublin Core (Elementos)	Conteúdo
Líder	09	N/A	Tipo de registro
Líder	06	N/A	Nível de tratamento
Base	04	N/A	Identificação da base de dados
007	N/A	N/A	Código para descrição física
008	09	N/A	Dados fixos / Tipo de registro
040	01	N/A	Fonte catalogadora
041	40	N/A	Idioma
044	67	Idioma	Idioma do documento
080 / 082 / 084	03	Assunto	Códigos de classificação
100	16	Criador	Nome pessoal
110	17	Criador	Nome corporativo
245	18	Título	Título
242	19	Título alternativo	Título traduzido
246	19	Título alternativo	Variações do título
260	62	Publicador digital	Responsável pela publicação
260	64	Data	Data da publicação
260	65	Data	Data padronizada
260	66	Publicador digital	Local de publicação
300	38	Formato	Descrição física
500	61	Descrição	Texto livre
650	87	Assunto	Descritores, palavras-chave, vocabulários controlados
650	88	Assunto	Descritores, palavras-chave, vocabulários controlados
700		Colaborador	Nome pessoal
710		Colaborador	Nome corporativo
856	08	Identificador	Uniform Resource Locator (URL)
945	05	Tipo	Tipo de recurso

Nota: N/A = Não aplicável.

Fonte: os autores, 2013.

Quadro 2 – Proposta de formulário para coleções de áudio

DUBLIN CORE (Elementos)	Refinamento	CONTEÚDO
Contributor	author ¹	Responsável pelo conteúdo do recurso
Contributor	interviewer ¹	Jornalista responsável pela entrevista
Contributor	soundeditor ¹	Responsável pela edição sonora
Date	Created	Data da gravação (padronizada)
Description	note ¹	Notas relativas ao item
Format		Informações sobre a parte física do recurso
Format	Extent	Tamanho do item
Format	Medium	Suporte físico do item
Format	Audiobits ²	Número de bits do arquivo digital (amostragem)
Format	Duration ²	Duração do áudio ou vídeo
Format	Extension ²	Formato do arquivo digital
Identifier	Citation	Citação bibliográfica, descrição de dados de identificação do recurso
Identifier	url	Identificador da URL (Uniform Resource Locator) do item
Identifier	Arquichesisid ²	Identificação para localização dos arquivos digitais
Identifier	Deviceid ²	Identificação dos equipamentos utilizados para a captura digital
Language	Isso	Informação sobre o idioma do conteúdo do item
Publisher		Editora responsável pela publicação
Publisher	Local	Local da editora
Publisher	País	País da editora
Subject	DeCS ¹	Assuntos extraídos do Descritores em Ciências da Saúde
Subject	vcusp ¹	Assuntos extraídos do vocabulário controlado da USP
Subject		Palavras-chave do autor
Rights		Informação sobre os direitos de acesso da publicação
Rights	Holder	O proprietário dos direitos autorais
Type		Tipo ou natureza do conteúdo
Title		Indicação do título do item
Title	Alternative	Indicação do título alternativo
Relation	References	Documentos citados

Notas: (1) Refinamentos que não estão indicados na documentação do Dublin Core, mas pertinentes a este tipo de material; (2) Refinamentos baseados em estudo da Library of Congress para o desenvolvimento do seu repositório digital (<http://www.loc.gov/standards/metatable.html>).

Fonte: os autores, 2013.

ria das necessidades de dados em comparação aos outros formatos, com a vantagem de possibilitar a inserção de novos metadados. No caso da Biblioteca/CIR foram acrescentados aos elementos existentes refinamentos, tais como: entrevistador e editor de som, por identificarem a responsabilidade desses profissionais na construção do áudio. Além desses, foram incluídos dados técnicos, como: áudio *bits*, duração e identificação do arquivo, com a finalidade de atender a necessidade de descrição e preservação digital, e fornecer parâmetros de uso no momento da recuperação da informação.

Conclusões

Este trabalho apresenta um modelo de formulário para inclusão de registros de áudio em repositórios digitais. Além do escopo mínimo de metadados no padrão DC, foram sugeridos refinamentos específicos para coleções de áudios, além de refinamentos baseados em estudos da Library of Congress, cuja finalidade é melhorar e completar a descrição dos registros desse tipo de coleção. Uma das vantagens das coleções em áudio é o fato de que a linguagem utilizada na expressão oral, especialmente em entrevistas, tende a ser mais coloquial e por isso torna a informação

mais acessível aos ouvintes não especializados e aproxima a universidade da sociedade. Dar visibilidade a essas coleções vai ao encontro do papel de promoção de pesquisas das instituições.

O mundo está rapidamente sendo transformado pela proliferação de recursos de informação disponíveis e distribuídos pela rede de computadores com um número crescente de instituições, tais como a Faculdade de Saúde Pública da USP, que tem produzido coleções digitais em áudio e vídeo. Por sua vez, repositórios digitais respondem a demandas da sociedade da informação com a disponibilização, na *web*, de dados e publicações científicas, o que segundo Abreu (2012) apresenta novos espaços de aprendizagem. Os trabalhos futuros envolverão outros tipos de materiais para as coleções de áudio, além de teses e dissertações, e o desenvolvimento de formulário específico para a coleção de vídeos produzidos pela FSP/USP.

Referências

- ABREU, J. S. Repositório institucional ou rede social de aprendizagem? **DataGramZero**: Revista de Informação, Rio de Janeiro, v. 13, n. 2, abr. 2012. Disponível em: <http://www.dgz.org.br/abr12/Art_06.htm>. Acesso em: 25 jun. 2014.
- ALVES, M. D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: DUBLIN CORE e MARC 21. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 4, n. 2, p. 20-38, jan./jun. 2007. Disponível em: <<http://143.106.108.14/seer/ojs/index.php/rbci/article/view/358>>. Acesso em: 25 jun. 2014.
- BIREME/ OPAS/OMS. **Manual de descrição bibliográfica**. São Paulo: BIREME/OPAS/OMS, 2008. Disponível em: <<http://metodologia.lilacs.bvsalud.org/download/P/LILACS-2-ManualDescricao-pt.pdf>>. Acesso em 25 jun. 2014.
- CANDELA, L.; CASTELLI, D.; FOX, E. A.; IOANNIDIS, Y. On digital library foundations. **International Journal on Digital Libraries**, v. 11, n. 1, p. 37-39, 2010.
- DUBLIN CORE METADATA INITIATIVE. **DCMI Metadata Terms**. 2012. Disponível em: <<http://dublincore.org/documents/2012/06/14/dcmi-terms/>>. Acesso em: 25 jun. 2014.
- GALLO, P. R. Radiodifusão comunitária: um recurso a ser valorizado no âmbito da educação em saúde. **Saúde em debate**, v. 25, n. 59, p. 59-66, set./dez. 2001.
- JORDAN, M. **Putting content online**: a practical guide for libraries. Oxford, U.K.: Chandos Publishing, 2006.
- MARANHÃO, A. M. N.; MENDONÇA, M. L. S. **MARC 21**: formato bibliográfico. [RJ], PUC Rio, [2010]. Disponível em: <<http://www.dbd.puc-rio.br/MARC21/index.htm>>. Acesso em: 25 jun. 2014.
- MARCONDES, C. H.; SAYÃO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. **Ciência da Informação**, v. 31, n. 3, Set. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n3/a05v31n3.pdf>>. Acesso em: 25 jun. 2014.
- SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, v. 29, n. 1, abr. 2000. Disponível em: <<http://dx.doi.org/10.1590/S0100-19652000000100010>>. Acesso em: 25 jun. 2014.
- TANI, A; CANDELA, L.; CASTELLI, D. Dealing with metadata quality: the legacy of digital library efforts. **Information Processing and Management**, v. 49, n. 6, p. 1194-1205, nov. 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0306457313000526>>. Acesso em: 25 jun. 2014.

Definition of metadata for audio recordings in open access digital repositories

Abstract

Introduction: The Library of the School of Public Health/ University of São Paulo (FSP/USP) is developing a digital repository for the public health area, and also has an audio collection, called Audioteca, produced by the institution itself. The objective of this study is to define the metadata for audio recordings in open access digital repositories, ensuring that the information attains greater visibility in the extensive internet universe. Method: Audio recordings of the databases Dedalus and LILACS were selected and the metadata used in these databases identified, using respectively MARC21 and LILACS methodology formats. Then a minimum metadata scope was established in accordance with of the Dublin Core (DC) schema, which corresponded to the metadata of the earlier formats. Finally, making use of the possibility of adjusting the DC, metadata refinements were created on the basis of a Library of Congress study, as well as some other refinements that met the specific needs of the institution and the Audioteca collection, were introduced, among them being: interviewer, sound editor and technical terms such as audio bits. Results: The result of this study is a specific form for the materials of audio collections, maintaining the compatibility and interoperability of Dublin Core scheme. Conclusions: It is expected that the results concurs to widen the visibility to audio collections which can be transmitted by community and educational radio programs. Such results also fulfill the purpose of promoting the research undertaken by institutions and contribute to the construction of citizenship.

Keywords

Audio library. Metadata. Digital libraries. Digital repositories.

Recebido em 14 de junho 2014

Aceito em 29 de junho 2014

Sobre as autoras:

Maria Imaculada da Conceição

Bacharel em Biblioteconomia e Documentação - ECA/USP. Supervisora Técnica. Biblioteca/Centro de Informação e Referência em Saúde Pública - FSP/USP. imak@usp.br

Maria do Carmo Avamilano Alvarez

Bacharel em Biblioteconomia - ECA/USP, Mestre em Saúde Pública - FSP/USP, Doutoranda em Saúde Pública - FSP/USP. Chefe técnica - Biblioteca/Centro de Informação e Referência em Saúde Pública - FSP/USP. malvarez@usp.br

Háilda Cristina Rocha Fernandes

Bacharel em Biblioteconomia - ECA/USP. Bibliotecária - Biblioteca/Centro de Informação e Referência em Saúde Pública - FSP/USP. halida@usp.br

Como citar este artigo:

CONCEIÇÃO, M. I. da; ALVAREZ, M. do C. A.; FERNANDES, H. C. R. Definição de metadados para registros de áudio em repositórios digitais de acesso aberto. **AtoZ**: novas práticas em informação e conhecimento, Curitiba, v. 3, n. 1, p. 72-77, jan./jun. 2014. Disponível em: <<http://www.atoz.ufpr.br>>. Acesso em:
