

Descrição de metadados mínimos para integração e preservação digital dos acervos científicos e culturais da Fundação Oswaldo Cruz: um estudo de caso.

Description of minimum metadata for integration and digital preservation of the scientific and cultural collections of the Fundação Oswaldo Cruz: a case study.

Descripción de metadatos mínimos para integración y preservación digital de los acervos científicos y culturales de la Fundação Oswaldo Cruz: un estudio de caso.

Aline da Silva Alves | aline.alves@icict.fiocruz.br

Fundação Oswaldo Cruz. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT). Rio de Janeiro, Brasil.

Barbara Cristina Euzebio Pereira Dias de Oliveira | barbaradias@ioc.fiocruz.br

Fundação Oswaldo Cruz. Instituto Oswaldo Cruz (IOC). Rio de Janeiro, Brasil.

Carolina Sacramento | carol@coc.fiocruz.br

Fundação Oswaldo Cruz. Casa de Oswaldo Cruz (COC). Rio de Janeiro, Brasil.

Caroline Ghirardelli | caroline.ghirardelli@fiocruz.br

Fundação Oswaldo Cruz. VPPLR

Claudete Fernandes de Queiroz | claudete.queiroz@icict.fiocruz.br

Fundação Oswaldo Cruz. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT). Rio de Janeiro, Brasil.

Inês Nogueira | inesn@coc.fiocruz.br

Fundação Oswaldo Cruz. Casa de Oswaldo Cruz (COC). Rio de Janeiro, Brasil.

Ivone Pereira de Sá | ivonesa@fiocruz.br

Fundação Oswaldo Cruz. Casa de Oswaldo Cruz (COC). Rio de Janeiro, Brasil.

Karina Veras Praxedes | karina.gagliano@fiocruz.br

Fundação Oswaldo Cruz. Casa de Oswaldo Cruz (COC). Rio de Janeiro, Brasil.

Marise Terra Lachini | mariseterra@fiocruz.br

Fundação Oswaldo Cruz. Casa de Oswaldo Cruz (COC). Rio de Janeiro, Brasil.

Marcos José de Araujo Pinheiro | mjap@coc.fiocruz.br

Fundação Oswaldo Cruz. Casa de Oswaldo Cruz (COC). Rio de Janeiro, Brasil.

Rodrigo Ferrari | rodrigo.ferrari@icict.fiocruz.br

Fundação Oswaldo Cruz. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (ICICT). Rio de Janeiro, Brasil.

Resumo

A partir da criação de um Grupo de Trabalho multidisciplinar para a elaboração dos documentos institucionais normativos “Plano de Preservação Digital” e “Manual de Digitalização” voltados para a implementação de ações referentes à preservação digital nos acervos científicos e culturais da Fundação Oswaldo Cruz (Fiocruz), observou-se a necessidade de se estabelecer padrões mínimos de metadados descritivos, para integrar e dinamizar o acesso aos distintos objetos digitais que representam os acervos arquivístico, bibliográfico, museológico e de coleções biológicas que compõem o patrimônio da instituição. Considerando o caráter histórico (alguns datam do século XVIII), probatório, informativo e a função científica e cultural desses objetos, serão fixados também metadados específicos voltados ao apoio tanto de sua preservação a longo prazo quanto dos próprios metadados descritivos atribuídos a eles. Cabe ressaltar que, embora sejam tratados distintos objetos (peças museais, itens bibliográficos, documentos arquivísticos e acervos biológicos). A proposta desta comunicação será apresentar a metodologia aplicada para a definição dos metadados integrativos baseados no uso do padrão Dublin Core (DC) e para a definição dos metadados de preservação a longo prazo, fundamentado no Preservation Metadata: Implementation Strategies (PREMIS), de acordo com as premissas de autenticidade, confiabilidade e acessibilidade.

Palavras-chave: Metadados descritivos e de preservação; Acervos Científicos e Culturais; Preservação Digital; Fundação Oswaldo Cruz (FIOCRUZ).

Abstract

From the creation of a multidisciplinary Working Group for the elaboration of normative institutional documents “Digital Preservation Plan” and “Digitalization Manual”, aimed at the implementation of actions related to digital preservation in the scientific and cultural collections of the Fundação Oswaldo Cruz (Fiocruz), it was observed the need to establish minimum standards of descriptive metadata, to integrate and streamline access to the different digital objects that represent the archival, bibliographic, museological and biological collections that make up the institution’s patrimony. Considering the historical character (some dating from the eighteenth century), evidence, informative, and the scientific and cultural function of these objects, specific metadata will also be set to support both their long-term preservation and the very descriptive metadata attributed to them. It should be noted that although different objects (museum pieces, bibliographic items, archival documents and biological collections) are treated, the proposal of this communication will be to present the methodology applied for the definition of integrative metadata based on the use of the Dublin Core (DC) standard and for the definition of long-term preservation metadata, based on the Preservation Metadata: Implementation Strategies (PREMIS) with the authenticity, reliability and accessibility premises.

Keywords: Descriptive and preservation metadata; Scientific and Cultural Collections; Digital preservation; Fundação Oswaldo Cruz (FIOCRUZ).

Resumen

A partir de la creación de un Grupo de trabajo multidisciplinar para la elaboración de los documentos institucionales normativos “Plan de Preservación Digital” y “Manual de Digitalización” orientados a la implementación de acciones referentes a la preservación digital de los acervos científicos y culturales de la Fundação Oswaldo Cruz (Fiocruz), se observó la necesidad de establecer estándares mínimos de metadatos descriptivos, para integrar y dinamizar el acceso a los distintos objetos digitales que representan los acervos archivísticos, bibliográficos, museológicos y de colecciones biológicas que componen el patrimonio de la institución. En cuanto al carácter histórico (algunos datan del siglo XVIII), probatorio, informativo y la función científica y cultural de esos objetos, se fijarán también metadatos específicos orientados al apoyo tanto de su preservación a largo plazo como de los propios metadatos descriptivos atribuidos a ellos. Cabe resaltar que, aunque sean tratados distintos objetos (piezas museales, elementos bibliográficos, documentos archivísticos y acervos biológicos). La propuesta de esta comunicación será presentar la metodología aplicada para la definición de los metadatos de integración basados en el uso del estándar Dublin Core (DC) y para la definición de los metadatos de preservación a largo plazo, basado en el Preservation Metadata: Implementation Strategies (PREMIS) con las premisas de autenticidad, confiabilidad y accesibilidad.

Palabras clave: Metadados descriptivos e de preservação; Acervos Científicos e Culturais; Preservação Digital; Fundação Oswaldo Cruz (FIOCRUZ).

Introdução

A Fundação Oswaldo Cruz (Fiocruz), uma das mais relevantes instituições de ciência e tecnologia em saúde da América Latina, tem por missão “Produzir, disseminar e compartilhar conhecimentos e tecnologias voltados para o fortalecimento e a consolidação do Sistema Único de Saúde (SUS) e que contribuam para a promoção da saúde e da qualidade de vida da população brasileira [...]”¹. Dentre as finalidades indicadas em seu estatuto, destaca-se a necessidade de preservar, valorizar e divulgar o patrimônio histórico, cultural e científico da instituição e contribuir para a preservação da memória da saúde e das ciências biomédicas².

No entanto, a instituição vem enfrentando desafios para se adequar às novas tendências e realidades do campo da preservação de acervos. Neste sentido, foi firmado, no ano de 2014, um contrato de apoio financeiro com o Banco Nacional de Desenvolvimento Econômico e Social (BNDES), com um projeto para a preservação e difusão dos acervos culturais e científicos da instituição, intitulado Preservo: Complexo de Acervos da Fiocruz, realizado em parceria com três unidades da Fiocruz: a Casa de Oswaldo Cruz (COC), o Instituto Oswaldo Cruz (IOC) e o Instituto de Comunicação e Informação Tecnológica em Saúde (ICICT), e contempla os acervos Arquitetônico, Urbanístico e Arqueológico, Bibliográfico, Biológico, Museológico e Arquivístico³.

Dentre outras iniciativas do Preservo, destaca-se a realização da digitalização de acervos e a criação de catálogos virtuais que ficarão disponíveis para acesso à sociedade e à comunidade científica. Segundo o coordenador geral do projeto:

O Preservo se integra aos esforços nacionais realizados nos últimos anos para a preservação do patrimônio científico e cultural brasileiro, que abrange tanto os acervos que registram as atividades científicas, quanto os arquivos que são fontes de pesquisa para as ciências humanas, biológicas e sociais³.

Diante da complexidade do projeto, tornou-se necessária a criação de diferentes grupos de trabalho (GT) para tratar de questões envolvidas na temática da preservação digital. Um dos grupos ficou responsável por definir diretrizes para digitalização dos acervos culturais e científicos da instituição, indicando equipamentos, procedimentos, sistemas de armazenamento, metadados, etc. adequados às diferentes tipologias dos acervos existentes na Fiocruz. Essas diretrizes serão consolidadas em dois documentos institucionais normativos denominados “Plano de Preservação Digital” e “Manual de Digitalização”.

Neste contexto, para consubstanciar os processos de preservação digital e de integração dos acervos arquivístico, bibliográfico, museológico e coleções biológicas, foi verificada a necessidade de especificar de maneira mais detalhada algumas ações, como o estabelecimento de metadados mínimos descritivos e de preservação, o que mereceu maior atenção e análise antes que passasse a constar no Manual. Para isso, foi criado um SubGrupo de Trabalho (SubGT) voltado à identificação e padronização desses metadados.

Composto por bibliotecários, arquivistas, analistas de sistemas, microbiologistas e historiadores, especialistas, que realizam o tratamento técnico desses acervos e conhecem suas características e especificidades, o SubGT tem como objetivos: estabelecer uma padronização mínima dos metadados descritivos para orientar a descrição e a representação dos acervos (baseado no padrão Dublin Core); identificar e propor o uso de um padrão de metadados de preservação digital para os acervos (baseado no padrão PREMIS - Preservation Metadata: Implementation Strategies); promover a interoperabilidade entre os acervos, de forma a permitir o acesso, a integração e a recuperação dos dados; e atuar em consonância com as diretrizes estabelecidas pela “Política de Preservação dos Acervos Científicos e Culturais da Fiocruz”.

Ressalta-se que, os objetos digitaisⁱ (peças museais, itens bibliográficos, documentos arquivísticos e coleções biológicas) compõem o patrimônio da instituição, tanto pelo seu caráter histórico (alguns datam do século XVIII), probatório e informativo, como pela sua função científica e cultural e, embora tratem-se

ⁱ Nesta pesquisa considera-se como objetos digitais o conjunto de representantes digitais oriundos da digitalização e os documentos nato digitais produzidos pela Fiocruz.

de distintos objetos, com suas especificidades e características próprias, é fundamental o estabelecimento de um padrão mínimo de metadados descritivos e de preservação digital.

A padronização semântica dos metadados descritivos viabilizará a integração entre os acervos, permitindo assim a interoperabilidade em ambientes informacionais, já os metadados relativos à preservação digital apoiarão a manutenção do acesso, da integridade e da autenticidade desses objetos ao longo do tempo.

Na Fiocruz existem diversos sistemas de informação que utilizam distintos padrões de metadadosⁱⁱ descritivos. Para os itens bibliográficos, utiliza-se o padrão Dublin Coreⁱⁱⁱ (DC) (no caso dos repositórios), para os documentos arquivísticos é adotado o padrão Encoded Archival Description^{iv} (EAD) e para as coleções biológicas utiliza-se o Darwin Core^v (DwC). O acervo museológico ainda não dispõe de um sistema digital de registro de informações, mantendo sua descrição em meio físico através de fichas de catalogação. Vale ressaltar que essa integração não irá alterar as fontes de informação existentes e seus metadados. Isso ocorrerá em um outro ambiente, de busca e recuperação integrada, que permitirá buscar essas informações em suas fontes originais, realizar a compatibilização e trazer os resultados num formato padronizado.

A riqueza de informações contidas em cada um dos sistemas citados e sua potencialidade de pesquisa evidencia a importância em integrá-los. Todavia, além de não interoperarem, esses diferentes sistemas ainda não são aderentes a padrões de metadados de preservação.

Outrossim, a definição de metadados mínimos de preservação contribuirá para que um objeto digital seja acessado e corretamente interpretado com o passar dos anos pois, ao contrário dos documentos impressos, cujas leitura e interpretação podem ser feitas diretamente, sem intermediação, na relação entre o usuário e um objeto digital existe uma fronteira: o ambiente tecnológico formado por camadas de software e hardware (Sayão, 2010). Portanto, a preservação digital não se restringe a simplesmente preservar o objeto digital, mas também os meios de apresentar e de usar os objetos⁴, o que torna essencial, e ainda mais complexo, o estabelecimento e atribuição desses metadados.

Deste modo, a proposta desta comunicação é apresentar a sequência de atividades utilizada pelo SubGT para identificação e padronização dos metadados mínimos de integração, baseados no uso do padrão Dublin Core (DC), e de preservação digital, fundamentado no padrão PREMIS. Para tanto, foram estabelecidas nove etapas de trabalho:

Etapa 1: Identificar os metadados utilizados em cada tipo de acervo, ou seja, verificar os metadados utilizados atualmente nas diferentes bases de dados (fontes de informação).

Etapa 2: Criar planilhas individuais (por tipo de fonte de informação) com cada metadado descritivo e a sua definição.

Etapa 3: Comparar os metadados identificados com o padrão Dublin Core.

Etapa 4: Analisar a definição de cada metadado descritivo e identificar os metadados que, embora possuam diferentes rótulos (nomenclatura), sejam utilizados para descrever o mesmo tipo de informação.

Etapa 5: Realizar a compatibilização entre o padrão DC e os padrões utilizados nas fontes de informação.

Etapa 6: Identificar o conjunto mínimo de metadados descritivos a ser proposto para utilização nos acervos Fiocruz.

Etapa 7: Estudar o padrão de metadados PREMIS.

ii Metadados podem ser definidos como: Dados sobre dados⁵.

iii O padrão Dublin Core é definido como “um conjunto de elementos descritivos com a finalidade de facilitar a descrição de recursos digitais disseminados na Internet”⁶. Disponível em: <http://dublincore.org/>

iv Disponível em: <https://www.loc.gov/ead/>

v O padrão Darwin Core é definido como “um conjunto de padrões, composto por um conjunto de termos (que em outros contextos podem ser chamados de propriedades, elementos, campos, colunas, atributos ou conceitos) desenvolvidos para facilitar o compartilhamento de informações sobre diversidade biológica provendo definições de referência, exemplos e comentários”. Disponível em: <http://rs.tdwg.org/dwc/>

Etapa 8: Identificar o conjunto mínimo de metadados de preservação a ser proposto para utilização nos acervos da Fiocruz, no contexto do Preservo.

Etapa 9: Consolidar e descrever os metadados mínimos, descritivos e de preservação, no Plano de Preservação Digital

A partir disso, com o estudo e compatibilização dos metadados, será possível estabelecer um padrão de metadados descritivos e de preservação nos diferentes acervos, possibilitando a integração e o compartilhamento de informações entre os sistemas e a preservação dos objetos digitais.

Este trabalho é de crucial importância para a definição dos requisitos necessários ao desenvolvimento e/ou adoção de sistemas de informação que viabilizem ações de busca e recuperação integrada de acervos além de focar na preservação desses objetos digitais a longo prazo.

Finalmente, entende-se que a divulgação desse trabalho desenvolvido pelo SubGT contribuirá para a compreensão teórica e a aplicação de práticas precisas no que tange ao uso de metadados como ferramenta para viabilizar a integração entre diferentes tipos de acervos e, conseqüentemente, dinamizar a recuperação da informação, incrementar a pesquisa científica e cultural e fomentar a garantia de preservação dos objetos digitais ao longo do tempo, considerando toda a complexidade envolvida na manutenção desses registros, tanto dos próprios documentos quanto de seus metadados atribuídos. Por isso, em escopo geral, considera-se que o trabalho ora exposto favorecerá no aprofundamento da reflexão crítica sobre a utilização de metadados descritivos e de preservação e estimulará o aumento de debates sobre o tema.

Referências

1. Fundação Oswaldo Cruz. Portal Fiocruz. Perfil Institucional. Rio de Janeiro, 2017. [acesso em 07 abr. 2017]. Disponível em: <https://portal.fiocruz.br/pt-br/content/perfil-institucional>
2. Fundação Oswaldo Cruz. VII Congresso Interno: plenária extraordinária nov./2015. Conhecimento e inovação para a saúde, o desenvolvimento e a cidadania: relatório final. Rio de Janeiro, 2016. 47 p. [acesso em 07 abr. 2017]. Disponível em: <http://congressointerno.fiocruz.br/sites/congressointerno.fiocruz.br/files/documentos/VII%20Congresso%20Interno%20-%20Relat%C3%B3rio%20Final%20-%20Carta%20Pol%C3%ADtica%2C%20Estatuto%2C%20Mo%C3%A7%C3%B5es%20e%20Pend%C3%Aancias.pdf>
3. Fundação Oswaldo Cruz. Agência Fiocruz de Notícias. Com apoio de R\$ 5 milhões do BNDES, Fiocruz lança o Preservo. Rio de Janeiro, 2014. [acesso em 07 abr. 2017]. Disponível em: <https://agencia.fiocruz.br/com-apoio-de-r-5-milh%C3%B5es-do-bndes-fiocruz-lan%C3%A7a-o-preservo>
4. Lavoie B., Gartner, R. Preservation Metadata. 2.ed. Oxford: Online Computer Library Center. Great Britain, 2013. 40 p. (DPC Technology Watch Report). [acesso em 20 abr. 2017]. Disponível em: <http://www.dpconline.org/docman/technology-watch-reports/894-dpctw13-03/file>
5. Sayão, LF. Uma outra face dos metadados: informações para a gestão da preservação digital. Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf. 2010; 15(30): 1-31. [acesso em 05 abr. 2017]. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1>
6. Souza, MIF, Vendrusculo, LG, Melo, GC. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. Ci. Inf. 2000; 29(1): 93-102. [acesso em 11 abr. 2017]. Disponível em: http://www.scielo.br/scielo.php?pid=S0100-1965200000100010&script=sci_abstract&tlng=pt