

Metadados para preservação digital de dados abertos: um estudo de identificação

Victória de Abreu e Silva

Universidade de Brasília, Brasil

Márcio Bezerra da Silva

Universidade de Brasília, Brasil

REVIEW

Resumo

Objetivo. Pesquisa que objetiva identificar padrões de metadados úteis à preservação e disponibilização digital de dados abertos conforme a literatura vigente.

Método. Caracteriza-se como uma pesquisa exploratória, na identificação de padrões de metadados, e bibliográfica, para o embasamento literário necessário sobre representação descritiva da informação, preservação digital e metadados, com fins de tratar, conservar e disponibilizar digitalmente os dados abertos, a partir da abordagem qualitativa de coleta de dados.

Resultados. Resulta, com base na pesquisa literária, no elenco de ambientes como sites governamentais, Institutos de pesquisa, repositórios digitais, catálogos eletrônicos e bibliotecas digitais que adotam padrões de metadados utilizáveis à preservação e disponibilização de dados abertos, sendo, especificamente, os esquemas: Dublin Core, utilizado em ambientes como o Portal Brasileiro de Dados Abertos; Preservation Metadata Implementation Strategies, presente em repositórios como o Carolina Digital Repository; Metadata Object Description Schema, adotado por sistemas como o Copac: UK and Irish Research Library Catalogue; e Metadata Encoding and Transmission Standard, acessado para gerir objetos digitais em bibliotecas e repositórios como a Biblioteca Nacional Digital de Portugal.

Conclusões. Concluiu-se que o único padrão de metadados pensado especificamente na preservação digital, e que se considera como o recomendado para a salvaguarda digital e disponibilização de dados abertos, é o Preservation Metadata Implementation, a partir de campos como rights, agents e events, enquanto que os demais esquemas identificados, mesmo sendo criados para outros fins, também possibilitam a preservação de dados, como é o caso do Dublin Core, no uso dos elementos source, relation, coverage e rights, do Metadata Object Description Schema, ao adotar os atributos location e recordInfo, e do Metadata Encoding and Transmission Standard, na adoção dos componentes structural link section e behavior section.

Palavras-chave:

Dados abertos; Metadados; Preservação digital; Preservation metadata implementation.

Metadata for the digital preservation of open data: an identification study

Abstract

Objective. Research that aims to identify useful metadata standards for the preservation and digital availability of open data according to current literature.

Method. Is characterized as an exploratory research, in the identification of metadata and bibliographic standards, for the necessary literary foundation on descriptive representation of information, digital preservation and metadata, with the purpose of digitally treating, conserving and making available the open data, from the qualitative approach of data collection.

Results. Based on literary research, results on the list of environments such as government websites, research institutes, digital repositories, electronic catalogs and digital libraries that adopt metadata standards that can be used to preserve and make available open data, specifically, Dublin Core, used in environments such as the Brazilian Open Data Portal; Preservation Metadata Implementation Strategies, present in repositories such as the Carolina Digital Repository; Metadata Object Description Schema, adopted by systems such as Copac: UK and Irish Research Library Catalog; and Metadata Encoding and Transmission Standard, accessed to manage digital objects in libraries and repositories such as the National Digital Library of Portugal.

Conclusions. Was concluded that the only metadata standard thought specifically of digital preservation, and that if considered as the recommended for the digital safeguard and availability of open data, is the Preservation Metadata Implementation, from fields like rights, agents and events, while that the other schemas identified, even if created for other purposes, also allow the preservation of data, such as the Dublin Core, in the use of the source, relation, coverage and rights elements, of the Metadata Object Description Schema, when adopting the location and recordInfo attributes, and the Metadata Encoding and Transmission Standard, in adopting the structural link section and behavior section components.

Keywords:

Digital preservation; Metadata; Open data; Preservation metadata implementation.

1 Introdução

Tomando como base o atual cenário de constante produção de informação, a preservação digital vem sendo levada em consideração ao funcionamento dos sistemas, pois assume “[...] uma importante etapa no gerenciamento de objetos digitais, [...] garantido o emprego de mecanismos que permitem o armazenamento em repositórios de objetos digitais que, garantem a autenticidade e a perenidade dos seus conteúdos” (MUNHOZ, 2011, p. 24). Nesse viés, de preservação digital, os objetos digitais (OD) compreendem-se como informações representadas por meio de uma sequência de dígitos binários, englobando, tanto os objetos criados originalmente em ambiente digital, quanto aqueles que passaram de um suporte analógico para o digital, por meio da digitalização, por exemplo, entre documentos textuais, fotografias, diagramas vectoriais, sequências de vídeo e áudio, websites, recursos de software etc. (FERREIRA, 2006).

Ao considerar a preservação digital como um cenário discursivo na Ciência da Informação (CI), deve-se ter em mente que a memória¹ do homem continua sendo desenvolvida, em um universo de produção informacional gerado e apresentado, tradicionalmente, via suportes como livros e revistas. No entanto, com os adventos computacionais², principalmente a partir da Internet, a humanidade se deparou com novos tipos de produção e disponibilização informacional, agora por meio de websites, por exemplo. Se antes, os periódicos ofereciam os seus artigos científicos em papel, agora, esses compêndios passaram a ser oferecidos de maneira digital³, criando-se um novo cenário, formalizado pela produção crescente e constante de conteúdo.

Os conteúdos são, tanto produzidos originalmente em computador, como são versões originalmente físicas e que migraram para uma versão digital (ARELLANO, 2004). Diante dessa questão, sobre a nascedouro do documento, independentemente de ser físico ou digital, no âmbito da biblioteconomia, documento é encarado como “qualquer unidade significativa de informação que tenha sido registrada em um suporte que permita seu armazenamento e sua posterior recuperação” (MUNHOZ, 2011, p. 21). Desta maneira, um documento digital (objeto digital – OD), permanece com a essência de trazer consigo informações, porém o armazenamento e/ou criação dessas passa a ser realizado essencialmente de maneira eletrônica.

Independentemente dos tipos de OD, a preocupação para com a preservação dos mesmos pode ser pensada a partir dos métodos que visem, inicialmente pela representação, até o acesso, por intermédio da preservação dos dados que os identificam. Todavia, para que ocorra a efetiva preservação da memória, faz-se uso de padrões de metadados, tradicionalmente conhecidos como “dados sobre dados ou informações sobre informações” (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004, p. 5) e configurado como um conjunto de recursos lógicos “[...] que descreve, explica, localiza e facilita a recuperação e o uso de um recurso de informação” (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004). Entre os exemplos, cita-se o padrão de metadados Dublin Core (DC)⁴, compreendido, de forma geral, como um “padrão de dados para catalogação de recursos eletrônicos da WEB” (GRÁCIO, 2002, p. 36), sendo este criado e implementado via linguagens web de marcação⁵, como o HyperText Markup Language (HTML)⁶ e o eXtensible Markup Language (XML)⁷, dentre outros padrões.

No caso das iniciativas de preservação digital, destaca-se o chamado Open Archival Information System (OAIS), estimado como um padrão-referência de metadados de preservação para arquivos que necessitam de salvaguarda em longo prazo, inclusive, sendo considerada uma International Organization for Standardization (ISO) desde 2003 e atestada pelo Consultive Committee for Space Data Systems (CCSDS). Mas para a efetiva preservação, ainda vale salientar que, seguindo a OAIS ou outro modelo, a estratégia adotada deve abordar questões como migração e conversão de formatos, recuperação de documentos etc. (ARELLANO, 2004).

Além de preocuparem-se com os OD, quanto ao suporte, as iniciativas de salvaguarda digital também promovem a preservação dos dados que representam esses objetos, em diferentes necessidades de armazenamento e acesso. Entretanto, diante das evoluções tecnológicas, discutidas e assumidas nos mais diversos campos do conhecimento, como na CI, compreender o que significa certo dado e representá-lo devidamente, com fins de recuperação, se mostra como um debate necessário ao âmbito dos dados abertos, disponibilizados em demasia, por vezes, mas sem a preocupação necessária quanto a formas de torná-los compreensíveis a quem deles necessita como também contribuir a identificação pelos meta-buscadores.

Donec finibus arcu quis ipsum fringilla sagittis. Sed massa ligula, rutrum vel augue a porttitor ultrices ex. Integer quis dignissim erat, eget mattis dolor. Etiam blandit diam ac ante sodales posuere. Nam lacinia et augue et tempor. Pellentesque quis mauris at neque dapibus malesuada id efficitur erat. Morbi et ultricies est. Morbi odio ante malesuada fringilla elit vitae, mollis elementum quam.

2 Fundamentação teórica

A contemporânea e constante produção de informação mencionada incide na necessidade de organizá-la para o posterior acesso e utilização. Neste contexto, em especial no viés da CI, estar atento as formas como os elementos que constituem a informação, neste caso, os dados, parece se apresentar como uma preocupação eminente, especificamente em como são representados, descritos. De acordo com as palavras de Aganette, Teixeira e Aganette (2017), também conhecido como catalogação, a representação descritiva da informação (RDI) visa representar os dados referentes a características físicas ou bibliográficas de um objeto, como um livro, por exemplo, na indicação de autor, título edição dentre outros, obedecendo regras e padrões.

Considerar a representação antes da disponibilização e compartilhamento dessas informações, como ocorre no meio acadêmico/científico, juntamente com os dados que as descrevem, de forma livre e acessível, é caminhar ao sentido dos chamados dados abertos. Pressupondo que esses dados, “livremente acessados, usados, modificados e compartilhados por qualquer pessoa para qualquer finalidade – sujeitos apenas, no máximo, aos requisitos para fornecer atribuição e/ou compartilhamento” (OPEN KNOWLEDGE FOUNDATION, 2019), necessitam estar descritivamente representados.

Para serem considerados dados abertos, devem ter duas premissas fundamentais. Em primeiro, para que sejam legalmente abertos, os dados devem possuir uma licença aberta (licença para uso livre dos dados) que permita, portanto, a qualquer pessoa acessar, reutilizar e redistribuir. Como segundo fundamento, os dados devem ser tecnicamente abertos, o que significa que devem estar disponíveis em um valor que não exceda o custo de produção, e que estejam em formato legível por máquina e passíveis de compartilhamento (OPEN KNOWLEDGE FOUNDATION, 2019a).

A partir da RDI garante-se a compreensão, o significado, dos dados, ao descrever as características do objeto representado, através de padrões de descrição já estabelecidos, como, por exemplo, o formato Machine Readable Cataloging 21^o (MARC 21), com enfoque em objetos bibliográficos, e o padrão de metadados DC. Além dessa compreensão, sobre o que significam os dados de um ambiente digital, os dois meios de padronização supracitados poderão garantir, tanto que os OD sejam acessados, quanto que os seus correspondentes dados estejam disponíveis, ambos a partir de iniciativas de uso livre. E, ainda, a partir da RDI e posterior disponibilização, os dados dos seus OD estarão não apenas aptos ao acesso, mas viabilizados à modificação, reutilização e redistribuição. Desta forma, mesmo considerando que “[...] os dados abertos têm sido estimulados por movimentos globais, como a campanha feita pela Iniciativa Internacional de Transparência em Programas de Assistência [...] sobre a transparência nos registros dos gastos”, por exemplo, antes de alcançarem tal rótulo de “livremente dispostos”, de legalmente abertos, pressupõe-se sobre a necessidade de preocupar-se com as temáticas de que se tratam esses dados; em seguida, que a adoção de padrões de metadados não onerem a descrição de dados, ao invés de códigos específicos serem criados, o que trariam a personalização ao invés da padronização; e, finalmente, que representar descritivamente, adotando padrões, no olhar da semântica, por exemplo, contribuirá na leitura dos dados pelas máquinas e conseqüente compartilhamento desses entre os mais diversos sistemas que usam da padronização descritiva.

Outra preocupação que deve ser envolvida aos dados abertos refere-se ao conhecimento aberto, ou ciência aberta, resumido no contexto deste trabalho como as produções do meio acadêmico/científico acessíveis e compartilháveis de forma gratuita. Em outras palavras, ciência aberta resume-se da seguinte maneira, a partir da adjunta intenção:

[...] preocupação primordial é tornar a atividade de pesquisa mais transparente, mais colaborativa e mais eficiente. [...] reconhece, primordialmente, que o conhecimento científico é um patrimônio da humanidade e, que, portanto, deve estar disponível livremente para que as pessoas – cientistas ou não – possam usá-lo, reusá-lo e distribuí-lo sem constrangimentos tecnológicos, econômicos, sociais ou legais (SAYÃO; SALES, 2014, p. 77)

A ideia de conhecimento aberto vai ao encontro dos dados abertos, mas tratando-se do conhecimento em si, o qual é considerado aberto quando “[...] qualquer pessoa está livre para acessá-lo, utilizá-lo, modificá-lo, e compartilhá-lo — restrito, no máximo, a medidas que preservem a proveniência e abertura” (OPEN KNOWLEDGE FOUNDATION, 2019b). Sendo assim, a preocupação com o alcance e compartilhamento do conhecimento produzido, através, por exemplo, de pesquisas acadêmicas/científicas, vai além da intenção de torná-las acessíveis, isto é, apresentá-las compreensíveis, tanto aos que necessitam dos dados de pesquisa, como na descoberta por meta-buscadores.

É no contexto de debate sobre uma ciência aberta que se reforça a importância sobre a padronização de dados, que necessitam ser abertos para o acesso, mas que necessitam de uma descrição (representação) que a recuperação em sistemas em rede, a partir da interoperabilidade⁹, enquanto um ponto importante, não apenas para o compartilhamento livre do conhecimento, mas também para a preservação digital. Consequentemente, torna-se importante também que tais padrões sejam disponibilizados de forma livre, para que a interoperabilidade ocorra em um maior número possível de ambientes/Instituições que se preocupem com o compartilhamento e a preservação de seus conhecimentos, como é o caso do DC, adotado como o padrão de metadados oficial pelos sistemas que se fundamentam à Open Archives Initiative (OAI), neste caso, conforme Oliveira e Carvalho (2009), através do Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH), compreendido como um protocolo que possibilita aos participantes o compartilhamento de metadados.

Padrões de metadados abertos, como o DC, facilitam a troca e disponibilização da produção científica/acadêmica na Internet, especialmente pelo fato da não adoção desses esquemas dificultarem a comunicação entre áreas do conhecimento. Autores como Sayão e Sales (2014) afirmam que a comunidade científica/acadêmica já compreende a importância de se preocupar com o tratamento, gerenciamento e preservação dos dados que descrevem suas publicações, garantindo, tanto a fonte de informação em si, como a sua recuperação futura. Desta forma, os autores ainda afirmam que:

[...] quanto maior a capacidade dos sistemas de informação de oferecer dados de pesquisas livremente e que sejam tratados por metadados, de forma que possam ser interpretados e reutilizados pelo maior número possível de pesquisadores de diversas áreas, maior será o grau de transparência, de reprodutibilidade e de eficiência do processo de geração de conhecimento científico, e maior será a amplitude de aplicação dos projetos de pesquisa para a sociedade (SAYÃO; SALES, 2014 p. 80)

Variando entre necessidades, permeadas por TIC contemporâneas, tipos de metadados existem, considerando especificidades técnicas e as mais diversas ações, como a RDI e a preservação digital, ambas destacadas no presente trabalho.

2.1 Metadados

Os metadados são elementos essenciais na preservação digital pois, por meio da descrição dos dados que compõem o OD, como formato, assunto, direitos autorais dentre outros, auxiliam no armazenamento, colaborando, consequentemente, na recuperação. Previamente, para compreender o seu conceito, torna-se prudente levar em consideração o contexto de momento, “[...] principalmente pelos desafios do mundo digital, [aos quais] foram re-desenhando a ideia puramente descritiva de metadados, criando expansões para o seu conceito com o intuito de abrigar novos propósitos e funções” (SAYÃO, 2010, p. 4). Tais elementos, comumente chamados de dados sobre dados, ou informação sobre informação (NISO, 2004), podem ser compreendidos como simples dados catalográficos ou bibliográficos (CAMPOS, 2007), devido à função de descrição de informações. Sendo assim, de maneira ampla, um metadado é uma informação estruturada que descreve, explica e localiza um OD (NISO, 2004). Em outras palavras, é possível descrever as características de um recurso de informação, abordando produtos e serviços (SOUZA, 2000), por exemplo.

Destrinchando a tentativa supracitada de conceituar, os metadados podem ser identificados por tipos, sendo descritivos, estruturais ou administrativos, assim apresentados por Sayão (2010, p. 5):

- Metadados descritivos: é a face mais conhecida dos metadados, são eles que descrevem um recurso com o propósito de descoberta e identificação; podem incluir elementos tais como título, autor, resumo, palavras-chave e identificador persistente;
- Metadados estruturais: são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são vinculadas entre si e ordenadas para formar um capítulo;
- Metadados administrativos: fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação. Nessa categoria, estão metadados técnicos que explicitam as especificidades e dependências técnicas do recurso; inclui também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso.

Outra forma de categorizar os metadados tem a compreensão de Gilliland-Swetland, ao considerá-los como administrativos, descritivos, de preservação, técnico e de uso (LOURENÇO, 2007), esmiuçados da seguinte forma:

[...] os metadados administrativos são especificamente as informações a respeito da forma de aquisição do recurso informacional pela biblioteca digital, os direitos de reprodução relativos à permissão ou não de cópias e os critérios de seleção para digitalização relativos a acervos físicos que foram digitalizados para formarem uma biblioteca digital;

[...] os metadados administrativos são especificamente as informações a respeito da forma de aquisição do recurso informacional pela biblioteca digital, os direitos de reprodução relativos à permissão ou não de cópias e os critérios de seleção para digitalização relativos a acervos físicos que foram digitalizados para formarem uma biblioteca digital;

[...] os metadados descritivos são os metadados utilizados para o tratamento dos recursos informacionais, utilizando para isso o conjunto de técnicas de catalogação, classificação e indexação da Ciência da Informação;

[...] metadados de preservação [...] constituem os metadados utilizados para informar o usuário da biblioteca digital sobre o original físico daquele recurso digitalizado: seu estado de conservação, sua localização física, etc. Com relação aos recursos informacionais gerados na própria web, as informações de preservação serão os dados relativos ao organismo e/ou pesquisador responsável pelo seu conteúdo e outras formas de aquisição, no caso deste recurso informacional vir a não estar disponível na biblioteca digital;

[...] aos metadados técnicos, pode-se identificá-los pelas próprias linguagens de marcação que são os softwares para estruturação de bibliotecas digitais. Constituem as informações técnicas de hardware e software necessárias para desenvolvimento, organização, acesso e visualização dos recursos informacionais de uma biblioteca digital. São os metadados de documentação, de estruturação e de esquematização de uma biblioteca digital;

[...] metadados de uso forneceriam informações a respeito dos usuários que poderiam utilizar esse recurso, sobre quantas vezes esse mesmo recurso foi acessado e por quantos usuários (LOURENÇO, 2007, p. 68, grifos nossos).

A partir das óticas conceitual, funcional e tipológica, indo ao encontro da preservação digital, autores como Arellano (2004) e Sayão (2007) defendem que os metadados de preservação, além de pertencerem ao discurso administrativo, abordam técnicas e meios que proporcionam estocar a informação à posterioridade, ou seja, ao longo do tempo.

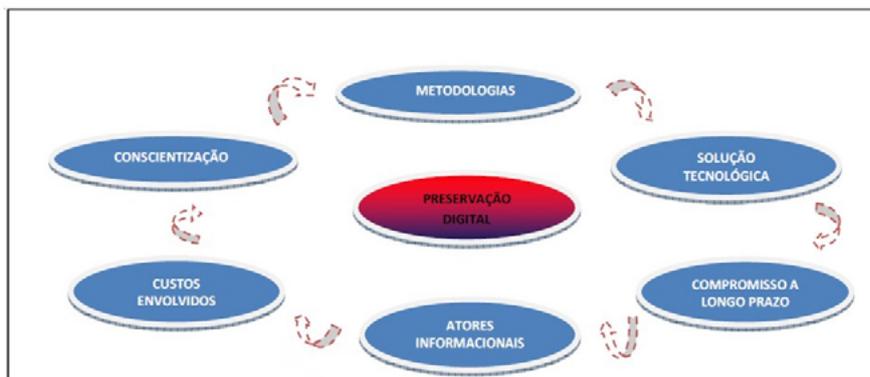
3 Preservação digital

Tendo-se em mente a importância da memória no processo de desenvolvimento do homem e da sociedade onde vive, compreende-se que a produção informacional¹⁰, mediante os mais variados tipos de registro, configura-se como um processo pertencente à própria construção da memória humana. Pensando neste processo, de produção informacional, destacam-se três momentos históricos: o advento da escrita, a prensa móvel de Gutemberg¹¹ e a segunda guerra mundial, que por sua vez apresentou o consequente aumento significativo na produção científica, como um texto científico, produzido em uma área do conhecimento específica, conforme Alves (2009) tendo este que ser publicado, para ser validado e julgado pelos pares.

Ao passo que o período da segunda guerra mundial proporcionou o avanço das ciências e das tecnologias, houve um aumento, em larga escala, no número de informações produzidas pelo homem, nas mais diversas formas, delineando uma espécie de nova explosão informacional. O avanço científico e tecnológico não se refreou, pelo contrário, vem aumentando exponencialmente até os dias de hoje, apresentando “a explosão da informação [...] como um problema social que teve seu início com o desenvolvimento das ciências, e hoje estendesse para todas as atividades humanas” (SARACEVIC, 1995, p. 2).

Em tempos recentes, proporcionado pelos avanços científico e tecnológico, apresenta-se um novo ambiente na contemporaneidade, ou seja, o universo digital, o qual presencia os mais diversos tipos de OD, nas formas de “[...] documentos de texto, fotografias digitais, diagramas vectoriais, bases de dados, sequências de vídeo e áudio, modelos de realidade virtual, páginas Web e aplicações de software” (FERREIRA, 2006, p. 21). Mas para que esses OD sejam acessados à posteriori, evoca-se a chamada preservação digital, ou seja, um conjunto de atividades com vistas a garantir o acesso no futuro à informação digital (GRÁCIO; FADEL, 2009), assegurando o “[...] armazenamento adequado, o acesso perene e a usabilidade dos objetos digitais para a posteridade” (ARAUJO; SOUZA; ARELLANO, 2015, p. 5). Mas para que a preservação digital transcenda do conceito à realidade de execução, algumas questões devem ser levadas em consideração, assim como apresentadas na figura 1.

Figura1: Elementos-chave na atividade da preservação digital



Fonte: Araújo, Souza e Arellano (2015, p. 5).

Conforme a figura 1, primeiramente, tem-se a metodologia, a qual abarca as decisões que devem ser tomadas pela Instituição a fim de realizar a preservação digital, sendo elas: formulação da política de preservação; como ocorrerá a garantia da proteção à propriedade intelectual; utilização das normas internacionais que regiram o processo; padrões e formatos a serem utilizados; e métodos de preservação a serem escolhidos. Na solução técnica, deve-se escolher a tecnologia utilizada para a realização do processo em si, inclusive sugerindo-se uma que faça parte de uma iniciativa já estabelecida no campo da preservação digital. Quanto à questão de compromisso a longo prazo, visa ser uma responsabilidade dos atores informacionais em garantir a continuidade do processo de preservação, independente de quaisquer problemas da Instituição. Logo, os atores informacionais são aqueles que interligam as demais ações, pois estes deverão responsabilizar-se pelo processo, pôr em prática cada etapa e garantir a continuidade da ação de preservação. Ademais, são também de responsabilidade dos atores os custos envolvidos no processo, pois deve-se pensar nos investimentos contínuos para a permanência das atividades. Por fim, a conscientização reforça sobre a necessidade/importância das atividades que contemplam o processo, pois a preservação dos OD faz parte do procedimento de preservação da memória (ARAUJO; SOUZA; ARELLANO, 2015).

A preservação digital objetiva, portanto, garantir que a informação digital esteja acessível, com qualidade e autenticidade, e podendo ser interpretada no futuro a partir de plataformas tecnológicas diferentes das utilizadas em sua criação (FERREIRA, 2006). Para tanto, a preservação digital leva em consideração três níveis de realização, os quais definirão um OD e a forma como este será preservado. Conforme a figura 2 é possível observar os níveis conceitual, lógico e físico no processo de preservação digital.

Figura 2: Níveis de um objeto digital na preservação digital



Fonte: Ferreira (2006, p. 24).

Conforme a figura 2, no nível conceptual (intelectual), a preservação tem como foco a garantia da integridade e autenticidade da obra, preocupando-se, portanto, com mecanismos que asseguram este nível de preservação (JESUS; KAFURE, 2010), isto é, que a obra seja representada de maneira fidedigna e compreensível a quem interessar nos processos de busca. No nível lógico, a preservação utiliza-se de métodos que objetivam converter o formato dos OD em versões mais acessíveis e atuais, evitando, assim, a obsolescência dos mesmos, devido à falta de manutenção dos recursos de software e hardware. Quanto ao nível físico, a preservação preocupa-se com a conservação das mídias físicas, como o Digital Video Disc (DVD), o Compact Disc (CD) e os disquetes (JESUS; KAFURE, 2010), além de cartões de memória, pen drives, blu-ray etc.

Os níveis supracitados farão uso de uma ou mais técnicas/métodos para que a preservação digital ocorra, pois se fará necessário “[...] na maioria dos casos, mudar, recriar, renovar: mudar formatos, renovar mídias, hardware e software” (SAYÃO, 2005, p. 118). Desta forma, a fala do autor configura-se nos métodos de preservação digital, sendo um conjunto de atividades que visam preservar o OD por intermédio de diferentes ações, como, por exemplo, conversão de formatos, transferência de suportes e digitalizações.

3.1 Métodos de preservação digital

Existem, na literatura especializada, duas concepções sobre a preservação dos OB, uma que visa à preservação do conteúdo e seu formato lógico e outra que se preocupa com sua conservação física. Em outras palavras, existem iniciativas que visam à preservação ou atualização no nível do hardware, enquanto outros visam preservar o conteúdo dos OD em si, isto é, o software e os seus dados. Arellano (2004) aponta que, tanto o dado, quanto a mídia que suporta a informação, devem estar em um estado que se permitam reproduzi-los a qualquer momento, a partir de métodos estruturais e operacionais (figura 3).

Figura 3: Métodos de preservação digital

Métodos Estruturais	Métodos Operacionais
Adoção de padrões	Conservação de <i>software</i> / <i>hardware</i>
Elaboração de normas	Migração de suporte
Metadados de preservação digital	Conversão de formatos
Montagem de infra-estrutura	Emulação
Formação de consórcios	Preservação do conteúdo

Fonte: Arellano (2004, p. 18).

Os métodos estruturais são aqueles utilizados na descrição e na guarda dos OD, ou seja, medidas preventivas de preservação. No caso dos métodos operacionais, são aspectos que garantem a reprodução da informação em formas diferentes de sua criação, mas tomando medidas que preservam a sua estrutura lógica, a partir de possíveis atualizações, transferências e simulações de formatos e suportes.

Na direção dos métodos estruturais, a adoção de padrões é o estabelecimento de formas uniformizadas de descrição e armazenamento de informações com o objetivo de facilitar os processos de recuperação e aplicação de métodos de preservação. A elaboração de normas vai ao encontro da adoção de padrões, pois são instruções, definidas pela unidade de informação, que visam certa padronização na descrição dos OD, bem como no armazenamento, preservação, transferências, conversões dentre outras ações. Os metadados de preservação digital são as iniciativas base no processo de salvaguarda, pois são eles que descreverão o OD de forma que possa ser recuperado futuramente, por meio dos dados que o identificam, além de ações voltadas a própria preservação como migrações, conversões dentre outras. Quanto à infraestrutura, tratam-se dos suportes estruturais, das unidades de informação, responsáveis por armazenar os OD com vistas à preservação, afetando diretamente na longevidade, ou seja, na conservação dos próprios OD, a partir dos seguintes pontos: armazenamento em ambiente estável/controlável; atualizações frequentes para cópias em novas mídias; cópias de preservação, inclusive assumindo licenças e permissões; procedimentos de manuseio adequados; e mídia de armazenamento padrão para transferência de conteúdo. Por fim, a formação de consórcios resume-se ao ato de aglutinar Instituições/unidades

de informação que fazem uso de um modelo de preservação em comum e, assim, garantir um padrão universal de preservação, bem como tornar a comunicação entre todos eficaz e duradoura (ARELLANO, 2004).

Com relação aos métodos operacionais, a conservação de software/hardware sugere a adoção de uma mídia estável para o armazenamento dos OD e, caso seja necessário, realizar uma cópia de dados para uma nova mídia (refreshed). No entanto, a cópia deve estar associada ao software original de criação do OD e ao sistema operacional (SO) equivalente. A migração de suporte é um método bastante utilizada por Instituições/unidades de informação que se preocupam com a preservação, pois consiste na migração periódica da informação para o funcionamento em recursos de hardware e software mais recentes, ao passo que garantirá a integridade da informação. Na conversão de formatos, como o próprio título sugere, os OD são convertidos para formatos mais recentes, auxiliando e acompanhando a atualização nas migrações, inclusive atuando em consonância a migração de suporte. Quanto a emulação, expressa-se em técnicas que preservam e acessam o OD em sua forma original, por meio de programas que simulam o funcionamento do hardware e do software utilizados em sua criação, como o SO da época, por exemplo. Para tanto, é sugerido que haja um encapsulamento¹² do OD junto ao software utilizado em sua criação para facilitar a simulação. Por fim, a preservação do conteúdo, como o próprio nome sugere, preocupa-se com a integridade do conteúdo do OD a partir de todas os métodos supracitadas, pois buscam a preservação do conteúdo, isto é, a adoção de técnicas de armazenamento, migração conversões e recuperação (ARELLANO, 2004).

Entre os métodos elencados, o que se destaca, por auxiliar na preservação dos dados descritivos e representativos ao OD, tanto à nível estrutural, quanto operacional, são os metadados de preservação, pois, além de realizarem a representação em si, permitem e garantem a autenticidade dos dados e possibilitam o acesso a esses entre sistemas, mesmo de características de software operacionais diversas, neste caso, a partir da interoperabilidade.

4 Padrões de metadados para preservação digital

Caracterizando-se como uma pesquisa exploratória, na identificação de padrões de metadados, e bibliográfica, para o embasamento literário necessário sobre RDI, preservação digital e metadados, com fins de considerar o tratamento, a conservação e a disponibilização digitalmente de dados abertos. A pesquisa ainda se caracteriza de abordagem de coleta de dados qualitativa, compreendida no período de junho de 2018 até fevereiro de 2019. No primeiro momento da pesquisa bibliográfica, o embasamento teórico de deu partir das seguintes expressões de busca: “dados abertos”, “representação descritiva da informação”, “RDI”, “preservação digital”, “preservação digital de dados”, “preservação digital de dados abertos”, “RDI e preservação digital”, “estratégias de preservação digital”, “métodos de preservação digital”, “metadados”, “metadados de preservação”, “ciência da informação e preservação digital”, “ciência da informação e metadados”, “modelos de preservação digital” e “ambientes de preservação digital”. No segundo momento da pesquisa bibliográfica, as expressões de busca corresponderam aos padrões de metadados identificados conforme a etapa anterior, a saber: “Dublin Core”, “PREMIS”¹³, “modelo PREMIS”; “modelo MODS”¹⁴; “MODS library of congress”; “METS”¹⁵; “METS library of congress”; e “METS AND aplicação”.

Entre os padrões de metadados está a Dublin Core Metadata Initiative (DCMI), ou apenas a iniciativa DC, voltado aos recursos da web, descrevendo e realizando a preservação dos OD. Conforme Souza (2000, p. 93), DC “[...] pode ser definido como sendo o conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos”. Trata-se de um esquema simples composto por 15 elementos¹⁶ essenciais que permitem serem mapeados em outros formatos, estruturando uma linguagem padrão, de fácil utilização (SAYÃO, 2010). A partir dos seus 15 elementos, o DC alcança a internacionalidade e objetiva a simplicidade, a flexibilidade, a semântica e a interoperabilidade entre sistemas (GRÁCIO, 2002). Entre os elementos do DC estão Title (título), Subject (assunto), Description (descrição), Language (idioma), Source (fonte) dentre outros.

Outro padrão de metadados voltado para a preservação é o Preservation Metadata: Implementation Strategies (PREMIS), grupo de trabalho internacional criado entre 2003 e 2005, resultado de em uma parceria entre a Online Computer Library Center (OCLC) e a Research Libraries Group (RLG) e atualmente mantido pelo Network Development and MARC Standards Office da LC (FORMENTON, 2015). A ideia desta parceria foi o estabelecimento de padrões de metadados para preservação digital (CAMPOS; SARAMAGO, 2007). Em outras palavras, o modelo PREMIS, ou Dicionário de Dados, foi o objeto exposto no relatório final da mencionada parceria, assim apresentado: “[...] define um conjunto de metadados básicos, implementáveis e de ampla aplicação, para apoiar a preservação digital em sistemas de repositórios” (FORMENTON, 2015, p. 81). Para tanto, o modelo entende que a representação do objeto, a ser preservado, se faz realizada a partir dos campos denominados de entidades,

sendo eles: Intellectual Entities (entidade intelectual), Objects (objetos), Events (eventos) e os Agents (agentes).

Um terceiro padrão identificado chama-se MODS, modelo desenvolvido pela LC em 2002, expresso em linguagem XML, derivado do padrão MARC e que objetiva traspor elementos selecionados do próprio MARC para a descrição de recursos. Atualmente, na versão 3.7, o esquema MODS resume-se a um conjunto de elementos para descrição bibliográfica, usualmente voltado para bibliotecas, com mais simplicidade que o padrão MARC, por exemplo (MODS, 2016). Da mesma forma que o MARC 21 e o DC, o esquema MODS possui elementos para a representação descritiva de objetos, como title Info (informações do título), abstract (resumo), physical Description (descrição física), related Item (item relacionado), identifier (identificador) etc.

O quarto e último o padrão de metadados identificado foi o METS, também criado pela LC em 2001. O esquema, mediante o uso da linguagem XML, objetiva gerir objetos digitais em bibliotecas e repositórios a partir de metadados descritivos, administrativos e estruturais. No uso desses metadados, o METS contempla as diversas e complexas ligações entre os OD, oferecendo, assim, uma codificação mais flexível (ARELLANO, 2010). A descrição ocorre a partir de seções, considerando as distintas partes do objeto. Essas seções são: Mets Header (cabeçalho METS), Structural Map section (mapa estrutural), Structural Link section (ligações estruturais) e Behavior Section (comportamento).

Atrelado aos padrões identificados, conseqüentemente, estão os ambientes digitais (na web) onde ocorrem as aplicações dos esquemas de metadados, considerados neste artigo como um conjunto que forma o campo da pesquisa e delineados pela característica comum de preservar digitalmente os seus dados. Neste caso, foram encontrados ambientes como sites governamentais, institutos de pesquisa, repositórios digitais (RD), catálogos eletrônicos e bibliotecas digitais. Entre os padrões de metadados identificados, e correspondentes aos ambientes de implementação, citam-se:

- DC: ao adotar campos como Format e Relation, e utilizado em ambientes como o Portal Brasileiro de Dados Abertos¹⁷, o qual oferta um conjunto de dados nacionais organizados por governo e política, saúde, educação, agricultura, economia, meio ambiente etc.;
- PREMIS: a partir de elementos como Intellectual Entities, Objects e Events, e presente em repositórios como o Carolina Digital Repository (CDR)¹⁸, o qual se apresenta como um arquivo digital que reúne a produção acadêmico-científica dos membros da Universidade da Carolina do Norte;
- MODS: no uso de elementos como originInfo, physicalDescription e accessCondition, adotado por sistemas como o Copac: UK and Irish Research Library Catalogue¹⁹ e disponibilizado como um catálogo de busca que interliga bibliotecas do Reino Unido e da Irlanda;
- METS: adota atributos como mets Header e Structural Map section, implementado para gerir OD em bibliotecas e repositórios como na Biblioteca Nacional Digital de Portugal (BND)²⁰.

Apresenta-se, portanto, o quadro 1 com a intenção de visualizar as características dos padrões de metadados identificados e de seus correspondentes elementos, neste caso, com foco nos campos que permitem a preservação digital de dados.

Quadro 1: Padrões de metadados identificados

Padrão	Características	Elementos de metadados		Características
		Element	Elemento	
DC		Description	Descrição	Relato do conteúdo do recurso, como, por exemplo, um índice, um resumo dentre outros.

	<p>- Esquema flexível e extensível de metadados descritivos para a descoberta de recursos web;</p> <p>- Mantido pela DCMI e que vem sendo muito usado por comunidades internacionais;</p> <p>- Pode ser representado em XML e propicia interoperabilidade de dados na web.</p>	<i>Date</i>	Data	Período de tempo associado com um evento no ciclo de vida do recurso, como a data de criação ou de procedimento de preservação etc.
		<i>Format</i>	Formato	É a manifestação física ou digital do recurso. Inclui o formato de arquivo, a forma física ou as dimensões (tamanho e duração por exemplo) do recurso
		<i>Identifier</i>	Identificador	Referência não ambígua para o recurso num dado contexto, como, por exemplo, uma sequência de números como o International Standard Book Number (ISBN) e o Digital Object Identifier (DOI).
		<i>Source</i>	Fonte	Referência a um outro recurso a partir do qual o recurso descrito é derivado no todo ou em parte.
		<i>Relation</i>	Relação	Referência para outro recurso do qual o recurso descrito é derivado ou está relacionado, como parte, versão, tradução, dependência de um recurso etc.
		<i>Coverage</i>	Cobertura	Âmbito espaço-temporal do conteúdo do recurso. Pode conter a designação de um lugar, de um período ou de uma entidade administrativa, além de datas, intervalos de datas e a especificação de um local via coordenadas geográficas.
		<i>Rights</i>	Direitos	Informação sobre os direitos de uso do recurso. Inclui uma indicação dos vários direitos de propriedade alusivos ao recurso, como os direitos de propriedade intelectual.
PREMIS	<p>- Esquema XML de metadados para suporte a preservação digital em repositórios;</p> <p>- É mantido pela Library of Congress (LC), focaliza o sistema e sua gerência e define unidades semânticas;</p> <p>- Pode ter apoio do DC, MODS, Encoded Archival Description (EAD) e Metadata for Images in XML (MIX) na verificação da autenticidade, integridade, procedência ou direitos relativos aos objetos mantidos em um RD.</p>	<i>Intellectual Entities</i>	Entidade Intelectual	Unidade descrita a partir de um conjunto coerente de dados, ou seja, é uma unidade de informação que pode ser um livro, uma revista, uma imagem e até uma base de dados, por exemplo.
		<i>Objects</i>	Objeto	Unidade de informação em formato digital, subdividida em ficheiros, fluxo de bytes (bits) e representação.
		<i>Events</i>	Eventos	Entidades que, durante o ciclo de vida do recurso, agregam metadados sobre quaisquer ações que o recurso venha a sofrer.
		<i>Agents</i>	Agentes	São pessoas, organizações ou aplicações informáticas associadas às ações de preservação de um dado OD.
		<i>Rights</i>	Direitos	Permissões legais de cada objeto ou agente.

MODS	<p>- Esquema XML aplicável aos objetos de bibliotecas digitais e mantido pela LC;</p> <p>- Elementos descritivos herdam a semântica dos elementos do MARC 21;</p> <p>- É mais simples que o MARC completo e mais rico que o DC qualificado²¹.</p>	<i>originInfo</i>	Informação de origem	Trata da origem do recurso, como dados do local de origem, edição, datas relacionadas, direitos autorais entre outras informações.
		<i>physicalDescription</i>	Descrição Física	Apresenta características físicas do recurso, dados de origem, dimensões, duração, formato entre outros.
		<i>tableOfContents</i>	Índice	Descreve os conteúdos do recurso, por exemplo, a indicação do conjunto de capítulos incluídos num livro ou em um documento textual.
		<i>relatedItem</i>	Item Relacionado	Traz as associações do recurso, ou seja, itens atrelados a ele, como referências, outras versões do recurso, outros formatos do mesmo entre outras informações.
		<i>identifier</i>	Identificador	Código ou número padrão único que identificará o recurso. Por exemplo, o International Standard Book Number (ISBN).
		<i>location</i>	Localização	Informações sobre o mantenedor do recurso, além da localização eletrônica Uniform Resource Locator (URL) do mesmo.
		<i>part</i>	Parte	Partes estruturais do recurso, a numeração de parte do recurso etc. Por exemplo: volume, edição, capítulo, seção, parágrafo entre outros, e suas respectivas medições, como a página, minuto etc.
		<i>accessCondition</i>	Condição de Acesso	Apresenta as restrições ou a falta destas para o acesso, o uso e a reprodução do recurso. Por exemplo: direitos de propriedade intelectual, limitações de distribuição entre outros.
		<i>recordInfo</i>	Informação de Registro	Informações necessárias ao gerenciamento e a interpretação dos metadados do registro, como a data da última modificação, a origem ou proveniência do registro, o idioma, as regras para descrição (Anglo American Cataloguing Rules, second edition – AACR2; Resource Description and Access – RDA; etc.) e qual instituição o criou ou alterou a sua originalidade.
		<i>mets Header</i> <metsHdr>	Cabeçalho METS	Seção onde estão os metadados que descrevem o documento em si, como, por exemplo, autor, criador, editor, datas entre outros.
		Structural Map section <structMap>	Mapa Estrutural	Parte central do documento METS, apresentando uma estrutura hierárquica para o OD e ligando os elementos dessa estrutura aos arquivos e metadados que correspondam à cada elemento.

METS	<ul style="list-style-type: none"> - Esquema XML mantido pela LC para gestão de objetos de biblioteca digital; - Pode implantar e padronizar os pacotes de submissão de informação (PSI)²², pacote de arquivamento de informação (PAI)²³ ou pacote de disseminação de informação (PDI)²⁴ no modelo sistema de arquivamento aberto de informação SAAI²⁵; - Organiza e expressa relações hierárquicas e hiperligações entre os objetos e suas partes; - Inclui DC, MODS, EAD e MIX na seção de Metadados Descritivos, e pode ter o PREMIS na seção de Metadados Administrativos 	<p><i>Structural Link section</i> <structLink></p>	<p>Ligações Estruturais</p>	<p>Seção onde os criadores METS registram se há hiperligações entre os elementos da hierarquia do mapa estrutural do documento, como também arquivam websites.</p>
		<p><i>Behavior Section</i> <behaviorSec></p>	<p>Comportamento</p>	<p>Utiliza-se para fazer associações entre um código executável (um programa ou um serviço) com um conteúdo do objeto (documento) METS, permitindo que esse conteúdo tenha uma característica específica, ou seja, um comportamento, além de estar disponível em fontes exteriores ao código METS.</p>

Fonte: Adaptado de Formenton (2015); Inclusões do autor (2018, 2019).

Diante do cenário exposto no quadro 1, observa-se aproximação na perspectiva da preservação digital entre os padrões de metadados elencados, especificamente, sendo o DC pelo uso dos elementos *Source*, *Relation*, *Coverage* e *Rights*; o PREMIS pelos campos *Rights*, *Agents* e *Events*; o MODS conforme os itens *location* e *recordInfo*; e o METS de acordo com os componentes *Structural Link section* e *Behavior Section*.

5 Conclusões

Buscou-se, neste trabalho, trazer a importância da preservação digital de dados enquanto resquícios da memória humana, tanto de conteúdos produzidas no próprio meio eletrônico, quanto os que são originários de maneira física e que se transformam em digitais por meio de métodos como os de migração, conversão e emulação. Neste sentido, o certame exploratório e bibliográfico permitiu identificar padrões de metadados que já vem sendo utilizados em ambientes digitais como o Portal Brasileiro de Dados Abertos, usando o DC; o CDR da Universidade da Carolina do Norte, o qual faz uso do PREMIS; o Copac do Reino Unido e da Irlanda, ao qual adota o MODS; e a BND portuguesa, identificada a partir do padrão de metadados METS.

Para a identificação de padrões de metadados, de maneira preliminar, foi considerada a RDI, pois descreve um OD conforme normas/padrões, e que, neste trabalho, se trata da primeira forma de compreensão dos metadados, inclusive, tendo a ação descritiva como finalidade de dois padrões de metadados identificados na pesquisa, sendo o DC e o MODS, enquanto que o METS, normalmente, é utilizado em RD, codificando dados descritivos, administrativos e estruturais de forma a torná-los comunicáveis entre sistemas.

Concluiu-se que o único padrão de metadados pensado especificamente na preservação digital, e que se considerada como o recomendado para a salvaguarda digital e disponibilização de dados abertos, é o Preservation Metadata Implementation, a partir de campos como *rights*, *agents* e *events*, enquanto que os demais esquemas identificados, mesmo sendo criados para outros fins, também possibilitam a preservação de dados, como é o caso do Dublin Core, no uso dos elementos *source*, *relation*, *coverage* e *rights*, do Metadata Object Description Schema, ao adotar os atributos *location* e *recordInfo*, e do Metadata Encoding and Transmission Standard, na adoção dos componentes *structural link section* e *behavior section*.

Finaliza inferindo que os supracitados exemplos de aplicações e uso de padrões de metadados para a preservação e disponibilização dos dados de objetos digitais correspondem apenas a realidade de um cenário maior, o

que culmina na expectativa de que mais estudos sejam realizados e, conseqüentemente, outros esquemas com tais finalidades de padronização sejam identificados, apresentados e exemplificados, como no âmbito dos repositórios dos dados abertos de pesquisa.

Referências

AGANETTE, Elisângela Cristina; TEIXEIRA, Livia Marangon Duffles; AGANETTE, Karina de Jesus Pinto. A representação descritiva nas perspectivas do século XXI um estudo evolutivo dos modelos conceituais. *Encontros Bibli, Santa Catarina*, v. 22, n. 50, p.176-187, set. 2017. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p176/34701>. Acesso em: 03 maio 2019.

ALVES, Bruno Henrique. Abordagens métricas: análise da produção científica de artigos e rede de colaboração científica dos docentes do programa de pós-graduação em ciência da informação, na linha de pesquisa organização da informação da UNESP/Marília. *Revista de Iniciação Científica da FFC, São Paulo*, v. 9, n. 2, p.104-115, 2009. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/ric/article/view/248>. Acesso em: 25 out. 2017.

ARAUJO, Priscilla Mara Bermudes; SOUZA, Rosali Fernandez de; ARELLANO, Miguel Ángel Márdero. Preservação digital e os periódicos científicos eletrônicos brasileiros em ciência da informação. In: Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB), 16, 2015, João Pessoa. Anais. Disponível em: <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2015/enancib2015/paper/viewFile/2829/1177>. Acesso em: 25 out. 2017.

ARELLANO, Miguel Ángel Márdero. Preservação de documentos digitais. *Ciência da Informação, Brasília*, v. 33, n. 2, p.15-27, maio 2004. Disponível em: <http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>. Acesso em: 09 ago. 2017.

BIBLIOTECA NACIONAL DIGITAL DE PORTUGAL. Apresentação. 2019. Disponível em: <http://purl.pt/index/geral/PT/about.html>. Acesso em: 03 maio 2019.

Brown University Library (2019). Welcome. 2019. Recuperado em 15 dezembro, 2018, de <https://library.brown.edu/>.

CAMPOS, Luiz Fernando de Barros. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. *Encontros Bibli*, n. 23, 1. sem. 2007. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12n23p16/390>. Acesso em 24 abr. 2017.

COLUMBIA DIGITAL LIBRARY. Welcome. 2018. Disponível em: <https://dlc.library.columbia.edu/>. Acesso em: 03 maio 2019.

CAROLINA DIGITAL REPOSITORY. About the Repository. 2018. Disponível em: <https://blogs.lib.unc.edu/cdr/>. Acesso em: 30 maio 2018a.

COPAC. About us. 2018a. Disponível em: <https://copac.jisc.ac.uk/about/>. Acesso em: 02 jun. 2018.

FERNANDES, Amaury. Notas Sobre a Evolução Gráfica do Livro. *Comum*. Rio de Janeiro: Faculdades Integradas Hélio Alonso, 2001. v.6, nº 17. p. 126 / 148. Disponível em: <http://www.producaografica.com.br/textos/NotEvGr-fLiv.pdf>. Acesso em: 11 out. 2017.

FERREIRA, Miguel. Introdução à preservação digital: conceitos, estratégias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/5820>. Acessado em: 17 out. 2017.

FONSECA FILHO, Clézio. História da Computação: O caminho do pensamento e da tecnologia. Rio Grande do Sul: Edipucrs, 2007.

FORMENTON, Danilo. Identificação de padrões de metadados para preservação digital. 2015. 103 f. Dissertação (Mestrado) - Curso do Programa de Pós-graduação em Ciência, Tecnologia e Sociedade, Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2015. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/7221>. Acesso em: 28 maio 2018.

GRÁCIO, José Carlos Abbud. Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002. 127 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Universidade Estadual Paulista, Marília, 2002. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/gracio_jca_dr_mar.pdf. Acesso em: 22 ago. 2017.

GRÁCIO, José Carlos Abbud; FADEL, Bárbara. Metadados para a preservação da informação digital. In: Encontro de Pesquisadores, 10, 2009, Franca, Anais, p. 354-362. Disponível em: <http://legacy.unifacef.com.br/novo/3fem/Encontro/Arquivos/Jose%20Gracio%20e%20Barbara%20Fadel.pdf>. Acesso em: 24 out. 2017.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. Dados conectados abertos. São Paulo: Editora Novatec, 2015.
JESUS, Joana D'arc Pereira de; KAFURE, Ivette. Preservação da informação em objetos digitais. *Biblionline*, João Pessoa, v. 6, n. 2, p.29-43, 2010. Disponível em: <http://periodicos.ufpb.br/index.php/biblio/article/download/7532/5516>. Acesso em: 17 out. 2017.

LOURENÇO, Cíntia Azevedo. Metadados: o grande desafio na organização da web. *Informação & Sociedade: Estudos*, João Pessoa, v. 17, n. 1, p.65-72, jan. 2007. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/466/1466>. Acesso em: 28 ago. 2018.

MILETTO, Evandro Manara; BERTAGNOLLI, Silvia de Castro (Org.). *Desenvolvimento de Software II: introdução ao desenvolvimento web com HTML, CSS, JAVASCRIPT e PHP*. Rio Grande do Sul: Bookman, 2014.

MUNHOZ, Augusto Maciel. Preservação digital de acervos: uma análise das políticas e estratégias utilizadas pelas universidades do Rio Grande do Sul. 2011. 63 f. Monografia (Graduação) - Curso de Biblioteconomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011. Disponível em: <http://www.lume.ufrgs.br/handle/10183/37473>. Acesso em: 19 ago. 2017.

MUCHERONI, Marcos Luiz; SILVA, José Fernando Modesto da. A interoperabilidade dos sistemas de informação sob o enfoque da análise sintática e semântica de dados na web. *Ponto de Acesso*, Salvador, v. 5, n. 1, p.03-18, abr. 2011. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/3622/3661>. Acesso em: 03 nov. 2017.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *Understanding Metadata*. 2004. Disponível em: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>. Acesso em: 05 set. 2017.

OLIVEIRA, Renan Rodrigues de; CARVALHO, Cedric Luiz de. Implementação de Interoperabilidade entre Repositórios Digitais por meio do Protocolo OAI-PMH. Goiás: Instituto de Informática Universidade Federal de Goiás, 2009. 56 p. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-09.pdf. Acesso em: 05 maio 2019.

OLIVEIRA, Zita Prates et al. O uso do dampo MARC 9XX para controle bibliográfico. *Ciência da Informação*, Brasília, v. 33, n. 2, p. 179-186, maio/ago. 2004. Disponível em: <http://www.scielo.br/pdf/ci/v33n2/a19v33n2>. Acesso em: 20 maio 2019.

OPEN KNOWLEDGE FOUNDATION. *Open Data*. 2019a. Disponível em: <http://opendatahandbook.org/glossary/en/terms/open-data/>. Acesso em: 19 maio 2019.

OPEN KNOWLEDGE INTERNACIONAL. *The Open Definition*. 2019b. Disponível em: <http://opendefinition.org/>. Acesso em: 03 maio 2019.

PORTAL BRASILEIRO DE DADOS ABERTOS. Portal. 2019. Disponível em: <http://dados.gov.br/>. Acesso em: 03 maio 2019.

RODRIGUES, Daniel. et al. *Descobrimo o OAIS*. 2013. Disponível em: <http://preservacaodigital5.blogspot.com.br/2013/01/descobrimo-o-oais.html>. Acesso em: 24 mar. 2018.

SARACEVIC, Tefko. Interdisciplinarity nature of Information Science. Tradução de: Durval de Lara Filho. *Ciência da Informação*, Brasília, 1995 v.24, n.1, p.36-41. Disponível em: <http://www.uff.br/ppgci/editais/saracevicnatureza.pdf>. Acesso em: 25 set. 2017.

SOUZA, Marcia Izabel Fugisawa; VENDRUSCULO, Laurimar Gonçalves; MELO, Geane Cristina. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. *Ciência da Informação*. 2000, vol.29, n.1, pp.93-102. Disponível em: http://www.scielo.br/scielo.php?pid=S0100-19652000000100010&script=sci_abstract&lng=pt. Acesso em: 24 out. 2017.

SAYÃO, Luís Fernando; SALES, Luana. Dados abertos para pesquisa: ampliando o conceito de acesso livre. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, Rio de Janeiro, v. 8, n. 2, p.76-92, jun. 2014. Disponível em: <https://www.reciis.iciict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 03 maio 2019.

SAYÃO, Luís Fernando. Metadados para preservação digital: Aplicação do modelo OAIS. Conselho Nacional de Energia Nuclear – Centro de Informações Nucleares – MCT/CNEN/CIN, 2007. 67 slides, color.

SAYÃO, Luís Fernando. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In: (ORG), Carlos H. Marcondes. Et al. *Bibliotecas Digitais: saberes e práticas*. 2. ed. Brasília: Edufba, 2005. p. 127-150. Disponível em: https://www.researchgate.net/publication/275031138_Preservacao_digital_no_contexto_das_bibliotecas_digitais_uma_breve_introducao. Acesso em: 25 set. 2017.

SAYÃO, Luís Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 15, n. 30, p.1-31, 2010. Universidade Federal de Santa Catarina (UFSC). Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1>. Acesso em: 03 maio 2019.

WORLD WIDE WEB CONSORTIUM BRASIL. Sobre o W3C. 2011. Disponível em: <http://www.w3c.br/Sobre/>. Acesso em: 1 maio 2018.

A interoperabilidade pode ser “[...] entendida como a capacidade que os sistemas de hardware e software têm para se comunicar e operar com outros sistemas no intercâmbio de dados” (MUCHERONI; SILVA, 2011, p. 5).

Notas al final

1 Conforme Le Goff (2013) na perspectiva coletiva, ou social, a memória é compreendida como aquela cuja formação vem por meio de produções realizadas pela humanidade como, por exemplo, a escrita e seus produtos, a música, as artes, as ciências dentre outras. Dito isto, o coletivo apresenta-se atrelado às ciências sociais e traz consigo o conceito de identidade coletiva.

2 Tratam-se de “[...] um corpo de conhecimentos formado por uma infraestrutura conceitual e um edifício tecnológico onde se materializam o hardware e o software” (FONSECA FILHO, 2007, p. 13).

3 A fim de evitar exclusões de conceitos, devido à questão semântica da palavra “documento”, que pode remeter o leitor a um objeto apenas textual, optou-se por utilizar, neste trabalho, o termo “objeto digital”, que por sua vez engloba documentos digitalizados, documentos criados em meio digital, websites, blogs, fóruns dentre outros.

4 De forma geral, DC é um “padrão de dados para catalogação de recursos eletrônicos da WEB” (GRÁCIO, 2002, p. 36).

5 Conjunto de ações utilizadas para a codificação de textos, visando a leitura e interpretação por computadores. (Almeida, 2002).

6 “[...] uma linguagem de marcação utilizada para criar páginas acessadas a partir de um navegador” (MILETTO; BERTAGNOLLI, 2014, p. 62), e que viabiliza a publicação de texto, imagem, vídeo, áudio dentre outros na web, tornando-se um padrão (W3C BRASIL, 2011).

7 Conforme Almeida (2002) é uma linguagem de marcação mais flexível que a HTML e que, por não ser predefinida, permite ao autor especificar a estruturação dos dados no documento, e ainda o definir semanticamente.

8 Desenvolvido pela LC na década de 1960, e que foi evoluindo ao longo do tempo até chegar na versão 21, o MARC é um formato eletrônico de descrição bibliográfica (OLIVEIRA et al., 2004).

9 A interoperabilidade pode ser “[...] entendida como a capacidade que os sistemas de hardware e software têm para se comunicar e operar com outros sistemas no intercâmbio de dados” (MUCHERONI; SILVA, 2011, p. 5).

10 Produção informacional, no presente trabalho, representa a produção de informação realizada pelo homem e pelas sociedades. É toda e qualquer informação produzida, por meio de qualquer meio e em qualquer formato, podendo ser uma partitura, um livro, um manual, um desenho dentre outros.

11 Gensfleisch von Guttenberg, xilografador e ourives por profissão, aperfeiçoou as técnicas de impressão por tipos móveis, introduzindo o uso das ligas metálicas, sendo estas mais resistentes e duráveis. Guttenberg inventou também a primeira máquina de impressão, revolucionando a imprensa (FERNANDES, 2001).

12 Baseia-se na ideia de que os objetos preservados devem ser auto descritos [sic.] e encapsulados em estruturas físicas ou lógicas com todas as informações necessárias para que seja decifrado e compreendido no futuro. (SAYÃO, 2010, p. 10).

- 13 Preservation Metadata Implementation Strategies. Website: <https://www.loc.gov/standards/premis/>.
- 14 Metadata Object Description Schema. Website: <http://www.loc.gov/standards/mods/>.
- 15 Metadata Encoding and Transmission Standard. Website: <http://www.loc.gov/standards/mets/>.
- 16 O esquema é escrito originalmente em língua inglesa.
- 17 Website: <http://dados.gov.br/>.
- 18 Website: <https://cdr.lib.unc.edu/>.
- 19 Website: <https://copac.jisc.ac.uk/>.
- 20 Website: <http://purl.pt/index/geral/PT/about.html>.
- 21 De forma geral, “o MODS é mais simples que o formato MARC completo e propicia uma descrição mais rica dos recursos quando comparado ao padrão de metadados DC qualificado” (FORMENTON, 2015, p. 67).
- 22 O PSI pertence a OAIS (SAAI), ou seja, é o conteúdo (objeto) do produtor enviado ao arquivo (sistema) (RODRIGUES et al., 2013).
- 23 O PAI também pertencente a OAIS (SAAI), refere-se ao pacote de informação efetivamente armazenado dentro do arquivo (sistema). (RODRIGUES et al., 2013).
- 24 O PDI também pertencente a OAIS (SAAI), o qual representa o pacote transferido do arquivo (sistema) para um consumidor em resposta a uma solicitação. (RODRIGUES, et al., 2013).
- 25 O PDI também pertencente a OAIS (SAAI), o qual representa o pacote transferido do arquivo (sistema) para um consumidor em resposta a uma solicitação. (RODRIGUES, et al., 2013).

Dados dos autores

Victória de Abreu e Silva

Bibliotecária pela Universidade de Brasília (UnB).

victoria.abreu95@gmail.com

Márcio Bezerra da Silva

Doutor em Ciência da Informação pela Universidade Federal da Bahia (2018). Mestre em Ciência da Informação pela Universidade Federal da Paraíba (2011). Graduação em Biblioteconomia pela Universidade Federal da Paraíba (2006). Tecnólogo em Processamento de Dados pela Escola Técnica Virginia Patrick-RJ (1996). Professor da Faculdade de Ciência da Informação (FCI) da Universidade de Brasília (UnB). Experiência e formação na área de Processamento de Dados (Computação). Atualmente se dedicando a direcionar e utilizar tal experiência na Educação (ensino), Ciência da Informação e áreas afins. Interesse por Tecnologias; Banco de Dados; Arquitetura da Informação, Acessibilidade, Usabilidade e Web; Representação, Organização e Recuperação da Informação. Membro dos grupos de estudos "G-Acervos manuscritológicos, bibliográficos, iconográficos, etnográficos: organização, preservação e interfaces das tecnologias da informação e comunicação" - UFBA; e "Biblioteca Digital" - UNB.

marciobdsilva@unb.br

Received: 2019-06-08

Accepted: 2021-07-31



This work is licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press