ISSN 1678-765X 10.20396/rdbci.v20i00.8666263

Artigo

RDBCI
Revista Digital de Biblioteconomia e Ciência da Informação
Digital Journal of Library and Information Science

Padrões de metadados no arquivamento da web: recursos tecnológicos para a garantia da preservação digital de websites arquivados[1]

Danilo Formenton [1]; Luciana de Souza Gracioso [2]

RESUMO

Introdução: A preservação digital no arquivamento da *Web* só será possível com o uso efetivo de padrões de metadados, pois são eles que determinaram a persistência, a coerência, a compreensibilidade, o acesso e a representação de *sites* selecionados, coletados e armazenados em arquivos da *Web*, além de definirem a arquivabilidade de *sites* e a interoperabilidade entre sistemas. Objetivo: Neste contexto, foi objetivo do artigo identificar e definir quais padrões de metadados poderiam ser julgados por instituições de memória e por universidades para que estas pudessem atender à preservação digital em arquivos da *Web*. Metodologia: Para isto, fez-se uma pesquisa qualitativa, exploratória e descritiva, que usa o método bibliográfico a partir de levantamento assistemático e de revisão e análise de conteúdo da literatura. Foram selecionados e analisados os padrões *Dublin Core*, MODS, EAD, VRA *Core*, PREMIS e METS. Resultados e Conclusão: A análise dos resultados aponta que *Dublin Core*, MODS, EAD e VRA *Core* amparam METS e PREMIS na descoberta e na documentação de aspectos técnicos dos *sites* e na comprovação de sua autenticidade, de seu contexto e de sua proveniência. O METS pode gerir *sites* arquivados, atuando como pacotes de informação OAIS, sendo que o *Dublin Core* mostrou ser um expoente para arquivamento da *Web* por seu uso em iniciativas notáveis da área.

PALAVRAS-CHAVE

Preservação digital. Arquivamento da Web. Metadados de preservação. Padrões de metadados. Ciência da Informação.

Metadata standards in web archiving technological resources for ensuring the digital preservation of archived websites

ABSTRACT

Introduction: Digital preservation in Web archiving will only be possible with the effective use of metadata standards. These standards are the ones that determine the persistence, consistency, comprehensibility, the access, and representation of selected sites, collected and stored in Web archives, besides defining the archivability of sites and the interoperability among systems. Objective: In this context, the objective of the article was to identify and define which metadata standards could

Correspondência do autor

[1] Universidade Federal de São Carlos, São Carlos, SP, Brasil / e-mail: formenton.danilo@gmail.com

[2] Universidade Federal de São Carlos, São Carlos, SP, Brasil / e-mail: lugracioso@yahoo.com.br

[1] O artigo origina-se de Tese de Doutorado em desenvolvimento, apresentando mudanças em relação ao texto original.

be judged by memory institutions and universities so that they could enable digital preservation in Web archives. Methodology: For this, a qualitative, exploratory, and descriptive research was done, using the bibliographic method from a non-systematic inventory together with a review and analysis of the literature content. The Dublin Core, MODS, EAD, VRA Core, PREMIS, and METS standards were selected and analyzed. Results and Conclusion: The analysis of the results indicates that Dublin Core, MODS, EAD, and VRA Core supported METS and PREMIS in detecting and documenting technical aspects of sites and proving their authenticity, context, and origin. METS can manage archived sites by acting as OAIS information packages, while Dublin Core proved to be an exponent for Web archiving through its use in remarkable area initiatives.

KEYWORDS
Digital preservation. Web archiving. Preservation metadata. Metadata standards. Information Science.

CRediT

- Reconhecimentos: Não é aplicável.
- Financiamento: O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.
- Conflitos de interesse: Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- Aprovação ética: Não é aplicável.
- Disponibilidade de dados e material: Não é aplicável.
- Contribuições dos autores: Conceituação, Curadoria de dados, Análise formal, Aquisição de Financiamento, Investigação, Metodologia, Administração de projetos, Recursos, Supervisão, Validação, Visualização, Escrita - Rascunho original: FORMENTON, D.; Revisão & edição: FORMENTON, D; GRACIOSO, L. S.

JITA: JH. Digital preservation

Artigo submetido ao sistema de similaridade

Submetido em: 03/07/2021 – Aprovado em: 16/08/2021 – Publicado em: 11/01/2022

1 INTRODUÇÃO

Proposta por Tim Berners-Lee em 1989, a *World Wide Web* é um registro exclusivo da vida no século XXI e um recurso de informação único, que hospeda milhões de *sites*, onde se conectam diferentes comunidades e indivíduos no mundo (PENNOCK, c2013). Porém, não só o ambiente da *Web* se desenvolve em uma velocidade intensa, como também as informações são publicadas e movem-se ao esquecimento rapidamente com a *Internet* e com o uso abundante de tecnologias. Do mesmo modo, os *websites* ao vivo (e as suas páginas da *Web*) são criados com rapidez e os seus *Uniform Resource Locator* (URLs) e conteúdos mudam regularmente e por vezes somem completamente, constituindo objetos digitais complexos, dinâmicos e efêmeros. Tudo isto representa uma ameaça muito real à nossa memória individual, organizacional, de fatos, ou cultural digital, bem como a seu legado técnico, a sua evolução e a sua história social (LIBRARY OF CONGRESS, [2021]; MASANÈS, c2006; PENNOCK, c2013; ROCKEMBACH; PAVÃO, 2018). Em frente a esta ameaça, conforme Costa, Gomes e Silva (2017), organizações do mundo todo – sobretudo universidades e instituições de patrimônio cultural, como bibliotecas, arquivos e museus – têm investido em políticas, em métodos e em tecnologias para coletar, preservar ao longo do tempo e tornar acessíveis cópias arquivadas do conteúdo da *Web*.

Nas últimas décadas, a preservação digital tornou-se um tema de estudo que se firmou na Ciência da Informação. Esse é um problema emergente, coletivo e atual nas produções nacionais e internacionais da área, exigindo análises inter e multidisciplinares e soluções sustentáveis, integradas e colaborativas. Uma das estratégias de preservação digital trata-se da adoção efetiva de padrões de metadados em apoio à gestão, à interpretação e à preservação de objetos digitais em meios informacionais, como repositórios.

Outra estratégia notável abrange a preservação digital do conteúdo de *websites*. Sendo um tema novo e carente de pesquisas e de iniciativas sistematizadas no Brasil (ROCKEMBACH; PAVÃO, 2018), o arquivamento da *Web* (*web archiving*) inclui cinco etapas, descritas em Kim e Lee (2007) e em Masanés (c2006), a saber: seleção (incluindo as fases de preparação – definir o objetivo da coleta, a política de captura e as ferramentas –, de descoberta – fixar os pontos de entrada para a captura, como a frequência e o escopo desta – , e de filtragem para reduzir o espaço aberto pela fase anterior aos limites na política de seleção), captura, arquivamento, acesso e revisão de qualidade; podendo este processo ser extensivo (não seletivo), intensivo (seletivo), centrado no tópico (temático) e/ou no domínio de *sites*. Unidos aos arquivos da *Web* surgidos, padrões e diretrizes para o avanço e para a preservação da *Web* são criadas por consórcios: o *World Wide Web Consortium* (W3C) e o *International Internet Preservation Consortium* (IIPC).

Integrados em páginas *Web*, na forma de *links* de uma página *Web* para outras e de registros do comportamento do usuário (RILEY, c2017), metadados (e padrões de metadados) têm a função de descrever unicamente um recurso informacional em ambientes digitais, multidimensionando suas formas de acesso e de uso, assegurando sua representação e a recuperação pelo usuário. Como exemplo, no domínio *Web*, o principal padrão é o *Dublin Core* (DC); no domínio arquivístico e museológico, há o *Encoded Archival Description* (EAD) e o *Visual Resources Association* (VRA) *Core* ou o *Categories for the Description of Works of Art* (CDWA); e, no domínio bibliográfico, frisamos o *Machine Readable Cataloging* (MARC) e o *Metadata Object Description Schema* (MODS). Para Formenton *et al*. (2017), os metadados ainda definem a garantia da preservação de um recurso/objeto digital (por exemplo, *sites* arquivados), através de padrões de metadados específicos, como o *PREservation Metadata: Implementation Strategies* (PREMIS) junto com o *Metadata Encoding and Transmission Standard* (METS).

Os padrões de metadados, sejam eles para descrição, gerência ou preservação de objetos digitais, são recursos tecnológicos-chave na interoperabilidade. Tal função é assegurada por práticas e por padrões de descrição que se traduzem nas sintaxes para codificação de dados,

como a *Extensible Markup Language* (XML) e a *Standard Generalized Markup Language* (SGML) *Document Type Definition* (DTD), além dos padrões de conteúdo (regras e códigos de catalogação), como *Cataloging Cultural Objects* (CCO), *General International Standard Archival Description* (ISAD(G)), *Anglo-American Cataloguing Rules* (AACR2) e *Resource Description and Access* (RDA) e dos padrões de valor de dados (vocabulários, tesauros e listas controladas), como *United States* (US) *Library of Congress Subject Headings* (LCSH). Estas práticas e padrões são indicadas por consórcios, órgãos de normalização e/ou líderes de comunidades, como, por exemplo, a *Online Computer Library Center* (OCLC), a *International Organization for Standardization* (ISO), a *National Information Standards Organization* (NISO) e o W3C.

Diante da carência de estudos nacionais de arquivamento da *Web* que investiguem, sistematizem e analisem a fundo os metadados e as características dos padrões de metadados aplicáveis na preservação de conteúdos *Web* em sistemas de arquivamento digital, é que se constatou a necessidade de identificar e de determinar quais padrões e esquemas de metadados poderiam ser julgados pelas organizações – sobretudo instituições de patrimônio cultural e universidades – que estão criando seus sistemas, para que estas pudessem atender à preservação digital em arquivos da *Web*. Objetiva-se, também, mais notadamente examinar em que grau os padrões de metadados no âmbito da preservação digital e do arquivamento da *Web* têm sido discutidos pela Ciência da Informação e pelas áreas afins, apontando os elementos de metadados que poderiam ser úteis às demandas de estruturação dos sistemas de arquivos da *Web* de forma mais apta à preservação de *websites* para fins históricos, culturais e de pesquisa.

Para isto, faz-se uma pesquisa qualitativa, de abordagem exploratória e descritiva (GIL, 2010; SILVA; MENEZES, 2005), que adota o método bibliográfico (MARCONI; LAKATOS, 2017; SEVERINO, 2016) a partir de um levantamento assistemático e de uma revisão da literatura específica, nacional e internacional, dos últimos vinte anos, dirigida e referente aos padrões e aos esquemas de metadados aplicados à preservação digital e ao arquivamento da *Web*. Assim, através da análise do conteúdo da revisão de produções científicas buscadas no *Google Scholar* e *Scientific Electronic Library Online* (SciELO) e nas bases *Scopus* e *ScienceDirect* (*Elsevier*), *Emerald Insight* (*Emerald Publishing*), *Web of Science* (*Clarivate Analytics*), *Library & Information Science Abstracts* (LISA) (*ProQuest*) e *Library, Information Science & Technology Abstracts with Full Text* (LISTA) e *Information Science & Technology Abstracts* (ISTA) (*EBSCO*) disponíveis no Portal de Periódicos CAPES, somada a análise de *sites*, relatórios e guias de consórcios e de órgãos de normalização e/ou líderes de comunidades, foram reconhecidas e sistematizadas uma definição, uma categorização e funções dos metadados; o conceito de metadados de preservação e as informações documentadas por metadados que apoiam a gestão da preservação digital de longo prazo e o arquivamento da *Web*; e os principais padrões de metadados usados na descrição e na preservação digital de conteúdos *Web* arquivados.

Assim, o presente trabalho se dispõe a expor os resultados e as análises dos conteúdos coletados, sendo que o produto deste mapeamento previu colaborar para prováveis delimitações de diretrizes e de políticas, as quais serão empregadas por instituições interessadas e/ou envolvidas com a captura, a retenção e o acesso permanente de um *website* ou de coleção de *sites* arquivados.

2 DEFINIÇÃO, CATEGORIZAÇÃO E FUNÇÕES DOS METADADOS

Como informações criadas, guardadas e partilhadas para descrever objetos, os metadados nos permitem interagir com eles para obtermos o conhecimento que necessitamos. Difundidos nos sistemas informacionais, os metadados aparecem de várias formas que nos mostram como os mesmos são todos estruturados até certo ponto, coletados para servirem a um objetivo útil e dispostos em categorias conhecidas. Na definição ampla e clássica de que

metadados significam “dados sobre dados”, é de se esperar que os metadados podem ser encontrados em qualquer lugar, e realmente são (RILEY, c2017). Entretanto, esta definição literal e minimalista do termo metadados não é satisfatória, visto que, pautando-se em Alves (2017) e em Sayão (2010), se faz inexpressiva e rasa ante a complexidade das funções designadas a eles nos contextos atuais da gestão da informação e também pelo fato de ser preciso entendê-los no domínio de aplicação onde estejam inseridos. Neste trabalho adotaremos a definição de metadados de Alves (2010, p. 47), por julgá-la aplicável ao domínio da *Web* e em domínios específicos, como o domínio bibliográfico, além de atender aos propósitos da presente investigação e fundamentar-se na construção padronizada e consistente de representações unívocas dos recursos informacionais em diferentes ambientes digitais estruturados. Desta maneira, os metadados (*metadata*) podem ser conceituados como:

> [...] elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades [...] com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação. (ALVES, 2010, p. 47).

Para a autora, a existência dos metadados dá-se através da sua codificação em estruturas de descrição padronizadas chamadas de padrões de metadados (*metadata statement*), sendo que o conjunto de metadados ou de elementos de metadados (*element sets*) integrará o esquema de metadados (*metadata schema*) do formato ou do padrão de metadados. Conforme Castro (2012) e Zeng e Qin (2008), o elemento de metadado (*metadata element*) corresponde a um termo formalmente definido para descrever uma das propriedades (ou atributos) do recurso de certo tipo ou com um propósito particular, como ‘o formato’ de um arquivo.

Além do conjunto de metadados (ou elementos prescritos, que são especificados através de declarações – *statements*), o esquema de metadados é composto pelo espaço de valor (*value spaces*), isto é, o conjunto de valores e regras de especificação para cada elemento e posição na estrutura descritiva, que são definidos por padrões externos ao esquema, como uma sintaxe para exprimir os valores nos elementos e esquemas de codificação que fixam regras de codificação, sintaxe dos dados e formas/valores aceitos. Tais componentes indicarão os aspectos estruturais (disposição dos atributos e relações entre elementos), de sintaxe (codificação dos elementos e ordem lógica dos valores) e semânticos (significado do atributo etc.) para a definição do esquema de metadados do padrão (ALVES, 2010; ZENG; QIN, 2008).

Para entender melhor a concepção de metadados, é útil separar metadados em categorias distintas que refletem aspectos-chave da funcionalidade deles em um sistema. Os tipos principais de metadados existentes são utilizados sob as particularidades do domínio (e as funções a serem realizadas), as demandas dos usuários e os tipos de recursos/entidades para representação (ALVES, 2017; GILLILAND, c2016; NATIONAL INFORMATION STANDARDS ORGANIZATION, c2004). A partir de *Digital Preservation Coalition* ([201-?]), Riley (c2017) e Sayão (2010) são consideradas várias categorias funcionais de tipos de metadados, sendo assim compreendidas:

- Metadados descritivos – detalham um recurso digital para localização, identificação ou compreensão. Podem incluir propriedades ou elementos, tais como título, autor e assunto, em que os usos primários são descobertos, apresentação e interoperabilidade.
- Metadados estruturais – explicitam a estrutura interna do arquivo digital e as relações hierárquicas de partes integrantes de recursos entre si. Podem ter propriedades, como ordem e lugar na hierarquia, em que os usos primários são navegação e apresentação.
- Metadados administrativos – fornecem informações que apoiam a gestão do ciclo de vida (criação, seleção, descrição etc.) dos recursos informacionais. Podem incluir propriedades, tais como tipo e tamanho de arquivo, data/hora de criação, evento de preservação, *status* dos direitos autorais e termos de licença, em que os usos primários são interoperabilidade, gerência de objetos digitais e preservação. Subdividem em:
  - Metadados técnicos – indicam os aspectos e as dependências técnicas de um

arquivo digital para decodificá-lo e renderizá-lo.
    - Metadados de preservação – incluem informações (por exemplo, as dependências de *hardware* e de *software*) exigidas para a gerência de um arquivo digital a longo prazo.
    - Metadados de direitos – documentam informações para apoio à gestão dos direitos de propriedade intelectual associados a um conteúdo.
- Linguagens de marcação – incluem metadados e sinalizadores para outros recursos estruturais ou semânticos no conteúdo. Podem conter propriedades, como parágrafo, nome, lista e data, em que os usos primários são navegação e interoperabilidade.

À vista disso, uma razão notável para criar metadados descritivos é facilitar a descoberta de recursos informacionais relevantes no domínio *Web* ou em domínios específicos; em adição, os metadados podem auxiliar a organizar recursos eletrônicos, promover a interoperabilidade, apoiar o arquivamento e a preservação, além de outras atividades comuns a serem feitas num sistema de informação digital, que, como afirmam Gilliland (c2016) e *National Information Standards Organization* (c2004), retratam algumas das funções primárias dos metadados. Para os objetivos deste trabalho, destacaremos a classe de metadados de preservação por serem vitais para a obtenção de uma efetiva gerência e para a preservação a longo prazo dos arquivos digitais e eletrônicos; e a classe de metadados descritivos, a faceta mais conhecida dos metadados (SAYÃO, 2010), que, apoiando-se nas indicações do grupo de trabalho internacional *Web Archiving Metadata* (WAM) da OCLC *Research* pelos estudos de Dooley e Bowers (c2018), Samouelian e Dooley (c2018) e Venlet *et al*. (c2018), serão abordados em conformidade com melhores práticas de criação de metadados descritivos coerentes e eficientes acerca de conteúdos da *Web* arquivados (ou melhor, *websites*) e para o arquivamento da *Web*.

3 METADADOS DE PRESERVAÇÃO E METADADOS DESCRITIVOS PARA ARQUIVAMENTO DA *WEB*

A julgar que a preservação digital é um processo de gestão, os metadados de preservação são categorizados principalmente como metadados administrativos, porém é admissível que os esquemas de metadados de preservação incluam elementos que se estendem em várias categorias, como descritivos, estruturais e administrativos. Tais metadados compõem uma parte crucial das estratégias de preservação digital e são concebidos no dicionário de dados PREMIS (um padrão internacional de fato para metadados de preservação) (CHEN; REILLY, 2011; DAPPERT *et al*., 2013; SAYÃO, 2010). O *Premis Editorial Committee* (2015, p. 2, tradução nossa) define os metadados de preservação (*preservation metadata)* "[...] como a informação que um repositório usa para apoiar o processo de preservação digital". Segundo Dappert e Enders (2010) e Caplan (2017), tratam-se das informações que descrevem um recurso digital no repositório para garantir o seu acesso e seu uso a longo prazo. Para Márdero Arellano (2008), são aqueles alusivos ao conteúdo do recurso, ao seu contexto e à estrutura de criação, além das alterações feitas em seu ciclo de vida. Definidos assim, nota-se que os metadados de preservação são construídos para cumprir uma ampla série de funções diferentes, todavia relacionadas (SAYÃO, 2010). Esses metadados suportam os distintos requisitos da preservação digital que, consoante Lavoie e Gartner (c2013) e *Premis Editorial Committee* (2015), se propõem a manter a disponibilidade; a renderização (tornar o objeto perceptível para um usuário via reprodução – para materiais visuais –, exibição – para materiais de áudio –, ou por outros meios próprios ao seu formato); a compreensibilidade; a identidade; a persistência; a autenticidade (qualidade de que o objeto é o que ele pretende ser, as quais a integridade do seu conteúdo e a origem podem ser verificadas); e a viabilidade (propriedade de ser legível pela mídia de armazenamento) de objetos digitais por longos períodos de tempo.

Isto posto, *Digital Preservation Coalition* ([201-?]) e Gilliland (c2016) deduzem certas razões dos metadados serem importantes para a preservação digital que, junto com as considerações dos autores supracitados, podem ser descritas da seguinte forma:

- Tomada de decisões – informações vinculadas a um objeto digital, como o *software* para abri-lo, o tempo que ele precisa ser mantido ou o histórico das alterações feitas nele, ajudam os profissionais a tomarem decisões sobre como e porque preservá-lo.
- Questões legais – os metadados permitem que os sistemas rastreiem níveis de direitos, licenças e informações de reprodução existentes para os itens originais, os seus objetos associados e as múltiplas versões destes.
- Persistência – a documentação por metadados de como o objeto de informação foi criado e mantido, como se comporta e como se liga com outros objetos será crucial à sua existência, independente do sistema atual usado para armazená-lo e recuperá-lo.
- Contexto para significado – os metadados fornecem informações de contexto requeridas para que futuros usuários entendam o significado do conteúdo de um registro, exercendo um papel vital na documentação de relações e na indicação da autenticidade, da integridade estrutural/processual e do grau de completude dos objetos.

Estas justificações expressam determinadas informações descritivas, administrativas e estruturais a serem incorporadas pelos metadados de preservação. Neste sentido, agrupando as ponderações de Caplan (2017), de Dappert *et al*. (2013), de Dappert e Enders (2010), de Formenton *et al*. (2017), de Lavoie e Gartner (c2013), da *National Library of New Zealand* (2003) e de Sayão (2010), identificamos um conjunto de informações e funções inter-relacionadas, que apoiam a gestão da preservação digital, abrangidas na captura, na criação e na manutenção de metadados de preservação:

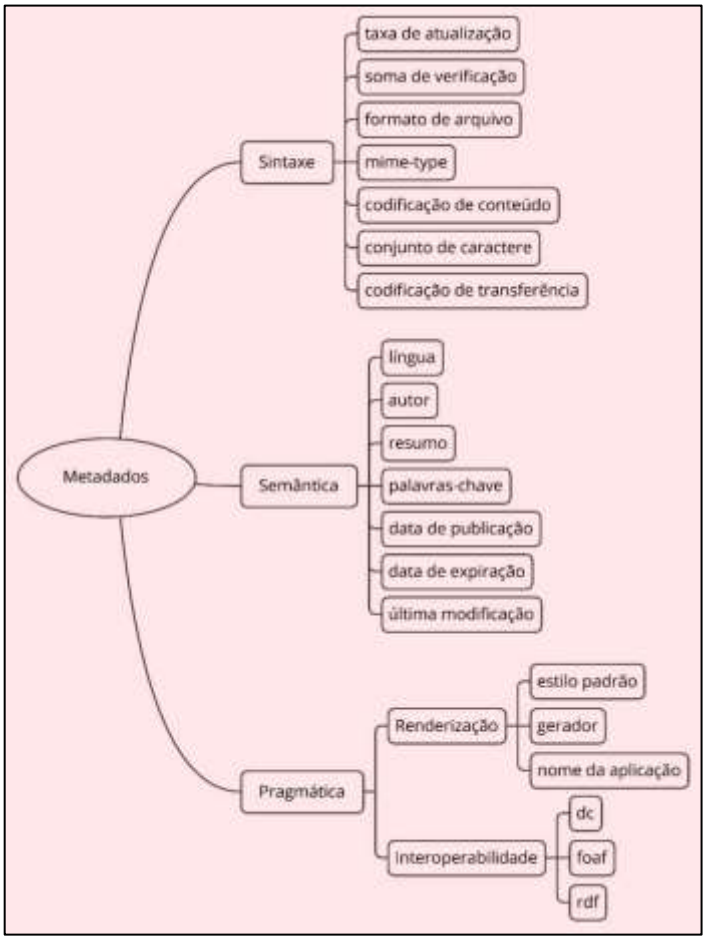
- Registro de informações sobre agentes – pessoas, organizações, *software* e *hardware* – com funções nos direitos, nos ambientes computacionais de renderização e nas ações – de preservação, disseminação, acesso, uso etc. – que afetam o objeto.
- Registro das dependências técnicas necessárias para acessar, renderizar – ou apresentar, executar etc. – e usar o objeto.
- Registro de informações que estabeleçam as propriedades significativas do objeto, isto é, características do objeto original e do ambiente que devem ser mantidas por ações de preservação para uma comunidade de usuários (por exemplo, as imagens em uma página da *Web*), orientando decisões sobre quais ações devem ser selecionadas.
- Registro das relações estruturais físicas e lógicas do objeto (por exemplo, qual imagem está integrada em qual *website* e qual página segue qual em um livro digitalizado), além de informações sobre seu meio de armazenamento.

Apesar de haver pouco trabalho que reúna e sintetize experiências de implementação de metadados de preservação para o acúmulo e a consolidação de melhores práticas na preservação digital ou, ainda, que avalie os custos inclusos na coleta e na gerência de metadados de preservação e os benefícios práticos de recair nestes custos (LAVOIE; GARTNER, c2013), os metadados de preservação são um componente-chave de todo arquivamento digital. Tais metadados documentam informações de conteúdo e de proveniência, autenticidade, fixidez, referência, contexto, direitos etc. alinhadas ao modelo de informação do *Open Archival Information System* (OAIS) e a seus três pacotes de informação (isto é, Pacote de Submissão de Informação – PSI, Pacote de Arquivamento de Informação – PAI e Pacote de Disseminação de

Informação – PDI), os quais garantem que recursos/objetos digitais sejam mantidos, retidos, identificados, acessados, decifrados, renderizados e usados de forma coesa e precisa no tempo.

Como indicado por Banos *et al*. (c2013) e por Melo e Rockembach (2020), o uso de metadados também constitui uma das principais facetas de arquivabilidade (*archivability facets*) de *websites*, ou melhor, fatores que devem ser levados em conta para calcular a extensão em que o *site* satisfaz às condições para a transferência segura de seu conteúdo a um arquivo da *Web* para intuitos de preservação. Com base em um modelo geral de perspectiva compartilhada em diversas disciplinas de informação – Filosofia, Linguística, Ciências da Computação etc. –, os autores consideram os metadados em três níveis (resumidos e demonstrados na Figura 1) para medição da capacidade de arquivamento de um *site* (arquivabilidade), a saber: sintaxe (por exemplo, como isso é expresso); semântica (por exemplo, sobre o que é isso); e pragmática (por exemplo, o que você pode fazer com isso).

Figura 1. Facetas de arquivabilidade: Metadados.



Fonte: adaptado de Banos *et al*. (c2003).

Banos *et al*. (c2013) explicam que metadados de codificação de conteúdo e de transferência podem ser inseridos pelo servidor em cabeçalhos *Hypertext Transfer Protocol* (HTTP); além disso, metadados de renderização, como o nome do aplicativo, a linguagem do usuário final para compreender o conteúdo; e, ainda, as informações descritivas, como autor e palavras-chave, que ajudam a entender como o conteúdo é classificado, podem ser incluídos no atributo e nos valores do elemento *HyperText Markup Language* (HTML). Para promover uma melhor interoperabilidade, os autores indicam o uso de metadados e de esquemas de descrição conhecidos, como, por exemplo, o DC, o *Friend of a Friend* (FOAF) e o *Resource Description Framework* (RDF); ademais, a existência de elementos de metadados selecionados é verificada

para elevar a possibilidade de implementar a extração automatizada e o refinamento dos metadados na coleta e ingestão do conteúdo *Web* ou, logo após, na fase de gestão do repositório.

A julgar que o arquivamento da *Web* é um processo relativamente novo para as instituições de patrimônio cultural, há poucos padrões; assim, as práticas de metadados variam muito dentre as distintas iniciativas na área, seja entre as bibliotecas nacionais ou, até, por diferenças nas abordagens de descrição entre as duas tradições de descrição bibliográfica e arquivística de recursos que não promovem a interoperabilidade dos metadados, pois, por exemplo: na catalogação em bibliotecas a natureza do conteúdo é revelada principalmente pelo título (se ele for descritivo) e pelos termos de assunto; de outra forma, os arquivistas rotineiramente utilizam notas extensas de texto livre para descrever tanto o conteúdo quanto o contexto do material (DI PRETORO; GEERAERT, 2019; DOOLEY; BOWERS, c2018). Destas práticas de metadados nas iniciativas de arquivamento da *Web*, mencionamos:

- O Arquivo.pt, serviço da Fundação para a Ciência e a Tecnologia (FCT) do Ministério da Educação e Ciência de Portugal, que permite a pesquisa e o acesso a páginas da *Web* portuguesa arquivadas, elucida certos metadados sobre os conteúdos de um *site* para a sua preservação, como: Descrição (*description*), texto breve descrevendo o conteúdo da página; Palavras-chave (*keywords*), expressões representativas dos principais temas da página; e *Dublin Core*, metadados do DC (ARQUIVO.PT, 2018).
- O *Internet Archive*, fundação que fornece acesso gratuito e universal a uma biblioteca digital com mais de 500 bilhões de páginas da *Web* e outros conteúdos arquivados, indica metadados que têm significado especial na descrição do conteúdo dos itens do arquivo, como Patrocinador (*sponsor*); *Scanner*; Data de Digitalização (*scandate*); Contagem de Imagens (*imagecount*); e Tipo de Mídia (*mediatype*) (INTERNET ARCHIVE, 2018). Aliás, como citado por Samouelian e Dooley (c2018), o seu serviço de arquivamento da *Web* dispõe de dezesseis campos DC, os quais os usuários podem escolher, e a capacidade de adicionar campos personalizados manualmente.
- No arquivo da *Web* da Biblioteca do Congresso americano, programa que gerencia, preserva e cede acesso a conteúdo da *Web* arquivado, codificam-se subelementos de metadados MODS no registro de *sites* para coleções temáticas e por eventos, como: Texto (<*text*>) para os escopos (domínio), no elemento Parte (<*part*>); Identificador (<*identifier*>) para o URL da fonte, no elemento Item Relacionado (<*relatedItem*>); e Lugar (<*place*>) no elemento Informação de Origem (<*originInfo*>) (LIBRARY OF CONGRESS, [2021]).

Diante disso, consoante Dooley *et al*. (2017), a OCLC *Research* estabeleceu o grupo de trabalho WAM em face do desafio da falta de uma abordagem comum para criar metadados na comunidade de arquivamento da *Web*. A fim de criar recomendações para metadados descritivos e facilitar a descoberta de conteúdo da *Web* arquivado, provendo uma ponte entre as abordagens bibliográficas e arquivísticas para a descrição, o grupo publicou três relatórios que incluem: uma revisão de literatura das necessidades de metadados descritivos dos usuários finais de arquivos da *Web* e dos profissionais que criam e gerem tais metadados (VENLET *et al*., c2018); uma análise de ferramentas de coleta da *Web*, com vista à sua funcionalidade para extração de metadados descritivos dos arquivos rastreados (SAMOUELIAN; DOOLEY, c2018); e diretrizes para ajudar instituições e pessoas a melhorarem a consistência e a eficiência de suas práticas de criação de metadados nessa área emergente (DOOLEY; BOWERS, c2018).

Neste último relatório, o grupo WAM indica um conjunto enxuto de elementos de dados, com definições de conteúdo e notas de utilização (ou seja, um dicionário de dados) adequadas às características únicas dos *websites* arquivados e relevantes para a descrição de materiais em bibliotecas e arquivos, como nos níveis de item e coleção, os quais podem ser usados isoladamente ou junto com outros padrões de conteúdo e de estrutura de dados mais

granulares (DOOLEY *et al*., 2017; DOOLEY; BOWERS, c2018). Além disto, conforme Di Pretoro e Geeraert (2019), cada elemento de dados WAN contém a vantagem de breves mapeamentos (*crosswalks*) para o DC, o EAD, o MARC 21, o MODS e o schema.org, que se destinam a facilitar tais conversões. Com base em Dooley e Bowers (c2018), o conjunto de quatorze elementos de dados do dicionário de dados WAM, para a descrição de *websites* ou de coleções de *sites* arquivados, é composto por:

1. Colecionador (*collector*) – a instituição incumbida da curadoria e da gestão de um *site* ou coleção arquivada.
2. Contribuidor (*contributor*) – a entidade (organização ou pessoa) que fez contribuições significativas, mas secundárias, ao conteúdo de um *site* ou coleção arquivada.
3. Criador (*creator*) – uma organização ou pessoa com a responsabilidade principal de ter criado o conteúdo intelectual de um *site* ou coleção arquivada.
4. Data (*date*) – uma única data ou intervalo de datas ligado a um evento no ciclo de vida de um *site* ou coleção arquivada.
5. Descrição (description) – as notas que explicam o conteúdo, o contexto e os aspectos de um *site* ou coleção arquivada.
6. Extensão (extent) – uma indicação do tamanho de um *website* ou coleção arquivada.
7. Gênero/Forma (*genre/form*) – um termo que determina o tipo de conteúdo de um *site* ou coleção arquivada.
8. Língua (*language*) – o(s) idioma(s) do conteúdo arquivado, incluindo os recursos visuais e de áudio com componentes linguísticos.
9. Relação (*relation*) – as relações todo/parte entre um único *website* arquivado e qualquer coleção a qual pertença.
10. Direitos (*rights*) – declarações de direitos e permissões legais outorgados pelo direito de propriedade intelectual ou demais acordos jurídicos.
11. Fonte de descrição (*source of description*) – informações sobre extração/criação dos metadados em si, como fontes de dados e data de obtenção dos dados das fontes.
12. Assunto (*subject*) – o(s) tópico(s) principal(s) que descreve(m) o conteúdo de um *site* ou coleção arquivada.
13. Título (*title*) – o nome pelo qual um *site* ou coleção arquivada é conhecida.
14. URL (url) – o endereço na *Internet* de um *site* ou coleção arquivada.

De outro modo, mediante a análise dos metadados de vários projetos de arquivamento da *Web*, como da Biblioteca Nacional da Austrália e dos *Smithsonian Institution Archives* nos Estados Unidos, Kim e Lee (2007) sugerem metadados descritivos e administrativos para o arquivamento intensivo da *Web*. Julgando que a maioria dos metadados dos projetos revisados se baseiam em DC e que o arquivamento intensivo da *Web* exige elementos de metadados mais detalhados, devido à seletividade voltada à qualidade, os autores adotaram, além dos elementos básicos do DC simples, outros elementos administrativos comuns nestes projetos, como:

- Disponibilidade (*availability*) – como o conteúdo *Web* pode ser obtido ou informação de contato.
- Público (*audience*) – o grupo esperado para utilizar o conteúdo da *Web*.
- Data da captura (*date captured*) – a data associada com a captura do *site* no arquivo.
- Data de validação (*date validated*) – a data em que a página *Web* foi validada, como sendo de fato codificada, usando o W3C *Markup Validation Service* ou outros serviços.

- Método de coleta (*collecting method*) – o método de coleta de conteúdos da *Web*, como, por exemplo, automático, manual ou transferido.
- Ferramenta de coleta (*collecting tool*) – os *softwares* necessários no processo de coleta de conteúdos *Web*.

Com efeito, o esquema DC demonstra ser notável para a descrição de conteúdos da *Web* arquivados, visto as semelhanças substanciais do padrão com o conjunto de elementos de dados WAN (DOOLEY; BOWERS, c2018), a adoção e a adaptação dos seus elementos essenciais nos metadados de iniciativas de arquivamento intensivo da *Web* ou, ainda, a simplicidade do padrão que motiva o seu uso geral no arquivamento extensivo da *Web* (KIM; LEE, 2007). Apesar da ambiguidade envolvida no escopo dos metadados de preservação, que, consoante Dappert e Enders (2010) e Lavoie e Gartner (c2013), é retratada pelas dificuldades em categorizá-los com precisão, podendo estes se estender por todas as classes de metadados, o trabalho centrou nos metadados descritivos e de preservação que apoiam a descoberta, a identificação, a apresentação, a interoperabilidade e a preservação digital a longo prazo de coleções de *sites* arquivados.

Nesse sentido, a definição e, talvez, a adaptação de padrões e esquemas de metadados tornam necessária uma ação em políticas de preservação digital no arquivamento da *Web*. Deve-se considerar as etapas do processo de arquivamento (seleção, captura etc.), as tecnologias (robôs rastreadores etc.) e os métodos de arquivamento adotados (domínio, temático etc.) e os tipos de conteúdo *Web* coletados, mantidos e disponibilizados (página *Web*, rede social etc.), bem como o atendimento às necessidades dos usuários finais, a série de informações a serem registradas e as decisões tomadas ante questões de direitos autorais, privacidade, custos, qualidade etc. e um futuro de imprevisões inerentes à preservação digital das informações publicadas na *Web*.

4 IDENTIFICAÇÃO DE PADRÕES E ESQUEMAS DE METADADOS PARA ARQUIVAMENTO DA *WEB*

A utilidade dos metadados dá-se da sua compreensibilidade por aplicações de *software* e por pessoas que os usam. Conhecidos como vocabulários de metadados, conjuntos de elementos ou, também, como formatos, os esquemas (*schemas*) podem ser formalmente padronizados através de organizações de normalização (ISO, NISO e W3C, por exemplo) e, em acréscimo, hospedados e mantidos por órgãos líderes da indústria ou da comunidade, como a US *Library of Congress*, que os endossa para uso em suas comunidades-alvo (RILEY, c2017). Os padrões de metadados (*metadata standards*) ajudam a tornar os metadados o mais úteis possível, pois, conforme *Digital Preservation Coalition* [201-?], cedem diretrizes para uma formatação uniforme à medida que os esquemas são diretrizes para formatos uniformes de metadados, assim, os padrões e os esquemas garantem que os metadados para registros digitais sejam interoperáveis.

Isto posto, chamados também de esquemas, os esquemas de metadados (*metadata schema*) são o conjunto de elementos de metadados (e regras para o seu uso) de um padrão criados para um propósito, como descrever um tipo de recurso informacional (CHAN; ZENG, c2006; NATIONAL INFORMATION STANDARDS ORGANIZATION, c2004). Em Zeng e Qin (2008, p. 323, tradução nossa), os esquemas de metadados constituem:

> Uma especificação processável por máquina que define a estrutura, codificação de sintaxe, regras, e formatos para o conjunto de elementos de metadados em uma linguagem formal num esquema. Na literatura o termo esquema de metadados refere-se usualmente ao conjunto de elementos na sua totalidade, bem como a codificação dos elementos e a estrutura com uma linguagem de marcação.

De fato, através de Castro (2012), Chan e Zeng (c2006), *National Information Standards Organization* (c2004) e Vellucci (2000), constata-se que o esquema (*schema*) é uma entidade no todo, incluindo os componentes semânticos e de conteúdo (chamados de conjunto de elementos de metadados), como a codificação dos metadados com uma sintaxe ou linguagem de marcação (o formato MARC e uma XML/SGML DTD, por exemplo), que têm três partes ou características básicas:

1. Estrutura – o modelo de dados ou arquitetura utilizada para comportar os metadados e a maneira como as declarações (*statements*) dos metadados são expressas. Como exemplos, pode-se citar a arquitetura de metadados RDF e o esquema XML METS.
2. Semântica – os nomes e significados dos elementos e seus refinamentos.
3. Conteúdo – as declarações ou instruções de como e quais valores devem ser atribuídos aos elementos.

Deste modo, o esquema de metadados (*schema*) define atributos e regras, sob os aspectos semânticos e estruturais, consistindo de outros tipos de esquemas (os *schemes*), que determinam a sintaxe de codificação dos dados, que por sua vez auxilia no estabelecimento da estrutura e da semântica (significado) dos atributos e valores em um padrão de metadados (ALVES, 2010; ZENG; QIN, 2008). A partir da revisão e da análise da literatura, identificamos vários padrões e esquemas de metadados usados para descrição de recursos em distintos domínios. A maioria dos padrões recorrentes teve as suas origens no momento em que a *Web* estava em seu começo. Na segunda metade dos anos 90 e início dos anos 2000, houve um rápido desenvolvimento de formatos para as necessidades de comunidades específicas e a codificação de objetos digitais complexos, os quais são delimitados por seus próprios conjuntos de elementos de metadados, por suas particularidades e pelos domínios de aplicação. A seguir são discutidos alguns dos principais padrões de metadados vigentes e indicados na literatura especializada para o arquivamento da *Web*. No entanto, não serão retratados todos os elementos de cada esquema e sim será realizado o apontamento apenas daqueles que fazem parte da análise dos resultados, ocasião em que são expostos o mapeamento e a indicação de elementos para o arquivamento da *Web*.

4.1 Padrão Dublin Core

O DC tem seu início em Chicago, na 2ª *International World Wide Web Conference*, em 1994, num debate sobre semântica e a *Web* diante da dificuldade da descoberta de recursos. Tal fato fez a OCLC e o *National Center for Supercomputing Applications* (NCSA) realizarem o OCLC/NCSA *Metadata Workshop* na cidade norte-americana de Dublin, Ohio, em 1995, em que se discutiu como um conjunto semântico básico seria útil para a busca e a recuperação de recursos baseados na *Web*. O resultado foi chamado de "metadados *Dublin Core*" com base na localização do *workshop* (DUBLIN CORE METADATA INITIATIVE, c2020a). Consoante Harper (2010) e Sayão (2010), o conjunto de elementos DC é pequeno e simples, de modo que é compreensível semanticamente; ademais, o DC é representado por diversas sintaxes, como codificado em HTML ou em XML e estruturado em RDF, propiciando o intercâmbio e o reuso.

Hoje, na versão 1.1, o DC é um vocabulário de dois níveis: simples e qualificado. Assim, o DC simples abrange quinze propriedades ou elementos essenciais (o *core*) e o DC qualificado contém elementos adicionais, além de qualificadores que especificam o significado do elemento (refinamento de elemento) ou identificam esquemas na interpretação do seu valor (esquema de codificação) (DUBLIN CORE METADATA INITIATIVE, 2012, 2020b). Para o escopo da preservação digital, Formenton *et al*. (2017) destaca alguns elementos DC qualificado, como, por exemplo, Formato (*format*), Identificador (*identifier*), Direitos (*rights*), Detentor de Direitos (*rightsHolder*) e Proveniência (*provenance*), que, embora mais voltados ao acesso do que para preservação, registram informações previstas nos metadados de

preservação PREMIS.

Independentemente das críticas à estrutura e ao conjunto muito simplista e genérico dos elementos DC (sobretudo, defronte aos demais formatos, como o MARC), o DC é um norteador da interoperabilidade semântica e do consenso entre diversas comunidades no mundo, inclusive executa um papel de liderança na criação de metadados descritivos de arquivamento da *Web* (DOOLEY *et al*., 2017; DOORN; TJALSMA, 2007; HARPER, 2010). Como apontado pelo grupo de trabalho WAM da OCLC *Research*, conforme Dooley e Bowers (c2018), Samouelian e Dooley (c2018) e Venlet *et al*. (c2018), o esquema de metadados descritivos do DC na versão 1.1 é amplamente usado na descrição de *websites* arquivados pelos usuários do *Archive-It*, um serviço de arquivamento da *Web* por assinatura do *Internet Archive*.

4.2 Padrão MODS

Projetado pela Biblioteca do Congresso dos Estados Unidos em 2002, o esquema MODS pode ser adotado em particular para aplicações de bibliotecas. Expresso em XML, este padrão de metadados descritivos inclui um subconjunto de campos MARC 21 e usa etiquetas baseadas em palavras e não numéricas, permitindo uma fácil compreensão (LIBRARY OF CONGRESS, 2016, 2018). Como vantagens do MODS, segundo Guenther (2003) e Mccallum (2004), nota-se que o MODS é mais simples que o MARC completo e enseja uma descrição mais rica frente ao DC qualificado; ademais, no MODS há o reagrupamento de certos elementos MARC e, em alguns casos, o que está em vários elementos MARC é reunido num único elemento MODS. Por exemplo, campos e subcampos MARC relativos à entrada principal e secundária de Nome (100 e 700) estão reagrupados no elemento Nome (<*name*>) MODS.

Atualmente na versão 3.7, o esquema MODS possui um conjunto de vinte elementos de metadados descritivos de nível superior, por meio do qual concede informações bibliográficas que integram demais esquemas XML, tais como o METS e o PREMIS. Sob o enfoque da preservação digital, Formenton *et al*. (2017) observam três elementos MODS: Informação de Origem (<*titleInfo*>), Item Relacionado (<*relatedItem*>) e Condição de Acesso (<*accessCondition*>). Para os autores, estes elementos documentam informações úteis que auxiliam os metadados de preservação, seja na comprovação da autenticidade, integridade e proveniência dos objetos seja na identificação dos direitos do recurso eletrônico que intervêm na preservação, no acesso e no uso dos seus conteúdos.

Embora os elementos MODS herdem a semântica dos elementos MARC, a conversão de um registro MARC original para MODS e depois o retorno para MARC resulta em perda de dados ou em alguma perda de especificidade na marcação. Em certos casos, se reconvertidos em MARC 21, os dados podem não ser inseridos exatamente no mesmo campo em que iniciaram, pois um campo MARC pode ter sido mapeado para um mais geral no MODS e, à vista disto, os dados em si não serão perdidos, somente a identificação detalhada do tipo de elemento que eles representam. Já em outros casos, um elemento MARC (por exemplo, o campo 340 Meio Físico) pode não ter um elemento equivalente MODS e, em seguida, os dados específicos podem ser perdidos ao se converterem para MODS. Logo, o MARC XML deve ser usado antes para uma troca sem perdas (GUENTHER, 2003; LIBRARY OF CONGRESS, 2016). Sobre exemplos de uso do MODS no arquivamento da *Web*, Guenther e Myrick (2007) indicam o Arquivo da *Web* da Biblioteca do Congresso americano, criado originalmente no projeto "*Mapping the Internet Electronic Resources Virtual Archive*" (MINERVA) em parceira com o *Internet Archive*, o qual é composto por coleções de *sites* arquivados que são catalogados com o MODS.

4.3 Padrão EAD

O esquema EAD originou-se em um projeto da biblioteca da Universidade da Califórnia, Berkeley, em 1993. Dirigido por Daniel Pitti, o projeto Berkeley visou desenvolver um padrão de codificação não-proprietário para instrumentos de pesquisa legíveis por

computador, como inventários, índices, registros, guias e documentos criados por arquivos, bibliotecas, museus e repositórios, para apoiar o uso de suas coleções (LIBRARY OF CONGRESS, 2013). Conforme Allison-Bunnell (2016) e Pala (2017), a versão EAD3 centra-se na simplificação do padrão e no aumento de clareza e de consistência semântica ante as versões EAD 1.0 e EAD 2002, promovendo a interoperabilidade e a melhora da funcionalidade em ambientes internacionais e multilíngues.

Hoje, na versão 1.1.1 do EAD3, este padrão XML tem um conjunto de cento e sessenta e cinco elementos descritivos e oitenta e cinco atributos, que fornece informações bibliográficas que se alinham a outros esquemas XML, tais como o *Encoded Archival Context – Corporate Bodies, Persons, and Families* (EAC-CPF) (SOCIETY OF AMERICAN ARCHIVISTS, 2019). No propósito da preservação digital, Formenton *et al*. (2017) observam certos elementos EAD 2002 mantidos na versão 1.1.1 do EAD3, como a Descrição Arquivística (<*archdesc*>). De acordo com os autores, os padrões DC, MODS e EAD, mesmo que sejam mais aplicáveis à descoberta, à busca, à recuperação ou à localização de recursos do que à preservação, são esquemas úteis para o registro de metadados descritivos de amparo ao PREMIS e ao METS.

Ainda que a falta de recursos e de conhecimento/*expertise* disponível numa instituição influencie a sua adoção, nos últimos vinte anos, como levantado por Eidson e Zamon (2019), o EAD mantêm-se relevante pelo grande número de arquivos que o adotaram e continuam a usá-lo atualmente para publicarem seus instrumentos de pesquisa *online*. Dos exemplos de uso do EAD no arquivamento da *Web*, há o Arquivo *Online* da Califórnia, que fornece acesso público e gratuito a descrições detalhadas de coleções de fontes primárias mantidas por instituições em todo o estado da Califórnia, como o arquivo da *Web* da Universidade da Califórnia em Irvine (https://oac.cdlib.org/findaid/ark:/13030/c8q81jn9/), através de instrumentos de pesquisa EAD.

4.4 Padrão VRA Core

Desenvolvido pela VRA em 1996, o VRA *Core* é um esquema para a descrição de obras culturais visuais – incluindo pinturas, desenhos, esculturas, arquitetura, fotografias etc. – , e de imagens que as documentam. É usado como um formato independente e como um esquema de extensão do METS[2] para objetos que contêm recursos de patrimônio cultural (VISUAL RESOURCES ASSOCIATION, 2014). Consoante Lima, Santos e Santarém Segundo (2016) e Lubas, Jackson e Schneider (2013), este padrão tem versões, sendo a VRA 1.0 (1996) baseada no CDWA, a VRA 2.0 (1996), que indicou a busca de padrões CCO e a VRA 3.0 (2000), que semelha ao DC em simplicidade, número de elementos e em qualificadores.

Atualmente, na versão 4.0, lançada em 2007, o esquema VRA *Core* em XML suporta a interoperabilidade e a troca de registros. O VRA *Core* 4.0 dispõe de dezenove elementos descritivos e de nove atributos globais, no qual o elemento *wrapper* de nível superior – Obra (*work*), Coleção (*collection*) ou Imagem (*image*) – inclui os demais dezoito elementos em registros individuais (VISUAL RESOURCES ASSOCIATION, 2007, 2014). Para fins de preservação digital, notamos elementos, como Localização (*location*), Direitos (*rights*) e Fonte (*source*), que podem apoiar o PREMIS e o METS na identificação e na definição da fidedignidade, da autenticidade, da integridade, da proveniência e do contexto de obras culturais e de suas representações.

A despeito de sua especificidade e da imposição de certas restrições à criação de *links* para registros não VRA *Core*, acrescido ao fato de ser menos comum diante de outros formatos, como colocado por Eíto-Brun (2015) e Senander III (2013), é possível criar *links* no esquema para instrumentos de pesquisa e o processo de conversão de registros VRA *Core* para registros MARC é bastante simples, direto e eficiente. Quanto aos exemplos de uso do VRA *Core* 4.0 no arquivamento da *Web*, indiretamente há o arquivo da *Web* da biblioteca da Universidade de

[2] *External schemas for use with METS*. Disponível em: https://www.loc.gov/standards/mets/mets-extenders.html. Acesso em: 23 jul. 2021.

Cornell que tem *sites* de coleções catalogadas com o VRA *Core*, como *Mysteries at Eleusi Images of Inscriptions* e *Billie Jean Isbell Andean Collection: Images from the Andes*.

4.5 Padrão PREMIS

O PREMIS remete ao nome de um grupo de trabalho patrocinado pela OCLC e pela *Research Libraries Group* (RLG) nos Estados Unidos de 2003 a 2005. Esse grupo criou um relatório final em 2005 chamado *Data Dictionary for Preservation Metadata*, que define um conjunto básico de unidades semânticas distribuídas em quatro tipos de entidades relacionadas entre si em seu modelo de dados (Objetos, Eventos, Agentes e Direitos), sendo implementável e de larga aplicação, a fim de apoiar a preservação digital em sistemas de repositórios. No dicionário de dados PREMIS, 'unidade semântica' corresponde a um pedaço de informação ou conhecimento e são as propriedades que descrevem entidades importantes com papéis quanto às atividades de preservação digital, isto é, os objetos digitais e seus contextos, eventos no ciclo de vida, agentes envolvidos na preservação e direitos. Isto posto, 'entidade' seria uma abstração para um conjunto de "coisas" (ambientes, eventos etc.) descritas pelas mesmas propriedades (CAPLAN, 2017; DAPPERT; ENDERS, 2010; PREMIS EDITORIAL COMMITTEE, 2015).

O dicionário de dados PREMIS não tem como alvo certas classes de metadados que já estão bem atendidas/supridas pelos padrões existentes, como é o caso de metadados descritivos e metadados técnicos de formato específico, combinando-se assim frequentemente com outros diferentes padrões (METS, *Metadata Authority Description Schema* – MADS, Z39.87/NISO *Metadata for Images in XML Schema* – MIX, por exemplo) para cobrir funcionalidades complementares suportadas por eles. Ademais, embora muito influenciado pelo modelo OAIS, que é amplamente aceito como um dos principais padrões a serem seguidos para normalizar repositórios de preservação digital, o PREMIS provê informações-chave, que abrangem todo o ciclo de vida dos objetos digitais e que vão além do âmbito do repositório, pois: fornece informações específicas para preservar os objetos digitais, enquanto o OAIS cede categorias mais amplas destas informações; e permite o registro de informações sobre objetos digitais que ocorrem antes de serem inseridos no sistema, o que não é coberto pelo OAIS (GUENTHER; DAPPERT; PEYRARD, c2016; LAVOIE; GARTNER, c2013; SAYÃO, 2010).

Hoje, na versão 3.0, emitida em 2015, o dicionário de dados PREMIS apresenta orientações para organização e concepção sobre metadados de preservação. Como citado anteriormente, o dicionário de dados PREMIS está estruturado em torno de um modelo de dados e, também, de implementações, como o esquema XML padrão associado[3], que define quais "coisas" precisam ser descritas (as entidades do PREMIS) e quais informações necessitam ser conhecidas pelo repositório de preservação para serem ditas sobre elas (as unidades semânticas do PREMIS) (GUENTHER; DAPPERT; PEYRARD, c2016; PREMIS EDITORIAL COMMITTEE, 2015). Conforme Caplan (2017) este esquema XML cedido pela PREMIS *Maintenance Activity* condiz diretamente com o dicionário de dados PREMIS, permitindo a descrição de Objetos, Eventos, Agentes e Direitos, como o uso do PREMIS representado em XML para a troca de metadados entre sistemas de preservação. Para Formenton *et al*. (2017), devido ao PREMIS aplicar o modelo de informação OAIS e os requisitos de preservação de objetos digitais (autenticidade, proveniência etc.), todas as suas entidades/unidades semânticas são vitais à preservação digital.

Embora a falta de treinamento/*expertise* e de integração com o sistema existente possam trazer barreiras à sua adoção em instituições de patrimônio cultural (ALEMNEH; HASTINGS, 2010), o dicionário de dados PREMIS fornece uma estrutura notável para descrever e preservar ambientes computacionais (*hardware*, *software* etc.) que suportam a renderização ou a execução dos objetos digitais e a sua utilização em longo prazo. Como exemplo, o PREMIS é adotado na descrição de ambientes de renderização para conteúdos *Web* da Biblioteca Nacional da França, que hospeda o arquivo da *Web* francesa (DAPPERT *et al*.,

[3] Disponível em: https://www.loc.gov/standards/premis/v3/premis-v3-0.xsd. Acesso em: 26 jul. 2021.

2013). Outro exemplo de uso do PREMIS, no arquivamento da *Web*, incluem Bailey e LaCalle (2015) e Rowell e Krewer (2016), que apresentam a visão do *Internet Archive* sobre como os metadados de preservação PREMIS interagem com o formato *Web ARChive* (WARC), um padrão de arquivo para conteúdo *Web*.

4.6 Padrão METS

Criado pela *Digital Library Federation* (DLF), o METS teve como precursor o projeto *Making of America* II (MOA2), de 1997, que elaborou um formato de documento XML para codificar metadados descritivos, administrativos e estruturais para obras textuais e baseadas em imagens. Expresso em XML, o METS possibilita codificar os metadados necessários na gestão de objetos de bibliotecas digitais em um repositório e na troca destes objetos entre repositórios (ou entre repositórios e seus usuários) (LIBRARY OF CONGRESS, 2017). Segundo Cantara (2005), Mcdonough (2006) e Sayão (2010), o METS é um mecanismo flexível para organizar todos os metadados associados ao objeto digital, exprimir as ligações complexas entre múltiplas classes de metadados e, adicionalmente, associar um objeto com comportamentos ou serviços, dando suporte à interoperabilidade, à escalabilidade e à preservação digital a longo prazo.

Atualmente na versão 1.12.1, de 2019, o esquema METS está direcionado a codificar objetos complexos em bibliotecas digitais. Para compartilhar documentos XML consoante o METS e estabelecer práticas comuns, o padrão define os componentes de um perfil METS e o esquema XML para codificá-lo. Estes perfis descrevem em detalhes uma classe de documentos METS para criar e processar documentos METS de acordo com um perfil específico, sendo que um documento METS constitui sete seções principais: Cabeçalho METS (<*metsHdr*>), Metadados descritivos (<*dmdSec*>), Metadados administrativos (<*amdSec*>), Arquivo (<f*ileSec*>), Mapa estrutural (<*structMap*>), *Links* estruturais (<*structLink*>) e Comportamento (<*behaviorSec*>) (DIGITAL LIBRARY FEDERATION, 2010). Na preservação digital, Formenton *et al*. (2017) salientam que um documento METS pode atuar na execução dos pacotes de informação OAIS.

Apesar de ser possível o uso do METS com o PREMIS, isto não é totalmente simples por alguns motivos. Há uma imperfeita correlação entre as duas estruturas, visto que o primeiro divide as informações em seções distintas, dependendo se são metadados técnicos, de direitos etc., e o segundo possui seções para Objetos, Eventos etc. Também METS e PREMIS possuem alguma duplicação (por exemplo, cada um define uma *tag* para armazenamento de *checksums*, impondo a decisão de se registrar esses elementos duplicados nas seções METS, nas seções PREMIS, ou em ambas), o que implica a necessidade de adoção de melhores práticas para usá-los juntos [4] em apoio à não variação de representação dos dados e à promoção da interoperabilidade (CAPLAN, 2017). Para mais, como citado por Lavoie e Gartner (c2013), a flexibilidade incluída no METS pode causar problemas de interoperabilidade, pois, quando um conteúdo tão diverso, tratado de várias maneiras, é permitido dentro das seções de um documento METS, torna-se mais difícil o intercâmbio de registros METS, contudo, isto é mitigado em certa medida pelos perfis METS registrados[5]. Por exemplo, o perfil METS para capturas de *sites* do projeto *ECHO DEPository*, na Universidade de Illinois em Urbana-Champaign (HABING, 2006), visa a transferência e a preservação digital do conteúdo de captura da *Web* entre repositórios. Outro exemplo dentro do arquivamento da *Web* inclui Truman (2016) e Veikkolainen e Lager (2016), que expõem o arquivo *Web* finlandês mantido pela Biblioteca Nacional da Finlândia, em que o conteúdo compreende arquivos em formato WARC nos pacotes de informação METS.

[4] *Using PREMIS with METS*. Disponível em: https://www.loc.gov/standards/premis/premis-mets.html. Acesso em: 23 jul. 2021.

[5] *Registered profiles*. Disponível em: https://www.loc.gov/standards/mets/mets-registered-profiles.html. Acesso em: 23 jul. 2021.

5 ANÁLISE DOS PADRÕES DE METADADOS À LUZ DA PRESERVAÇÃO DIGITAL NO ARQUIVAMENTO DA *WEB*

Os padrões e esquemas de metadados DC, MODS, EAD, VRA *Core*, PREMIS e METS possuem características em comum e algumas singularidades. As ponderações realizadas aqui basearam-se nos princípios da preservação digital de longo prazo, nas definições do modelo de informação OAIS, nas informações expressas pelos metadados de preservação e nos metadados descritivos e administrativos para o arquivamento da *Web*, sobretudo com os elementos WAN de Dooley e Bowers (c2018), os elementos de Kim e Lee (2007) e os metadados de Banos *et al*. (c2013) e de iniciativas na área (incluindo aqueles exibidos em registros de coleções de arquivos da *Web*, como das bibliotecas da Universidade de Cornell e do Congresso americano), os quais foram descritos no trabalho. De forma não exaustiva, o Quadro 1 sintetiza os aspectos básicos dos padrões aludidos e os elementos de metadados (ou as unidades semânticas para o PREMIS) tidos, para esta pesquisa, como importantes na preservação de conteúdos *Web*.

Quadro 1. Padrões e elementos de metadados de apoio à preservação digital no arquivamento da *Web*.

PADRÃO	CARACTERÍSTICAS	ELEMENTOS DE METADADOS ÚTEIS PARA A PRESERVAÇÃO DIGITAL NO ARQUIVAMENTO DA *WEB*	
DC qualificado (versão 1.1)	- Esquema XML ou em outras sintaxes tido como flexível, extensível, simples e interoperável; e - Aplicável à descoberta de recursos *Web* e para o arquivamento da *Web* por seu uso geral pelos usuários do *Archive-It*.	▪ Título (*title*) ▪ Criador (*creator*) ▪ Assunto (subject) ▪ Descrição (*description*) ▪ Colaborador (*contributor*) ▪ Data (*date*) ▪ Tipo (*type*) ▪ Formato (*format*)	▪ Identificador (*identifier*) ▪ Fonte (*source*) ▪ Língua (*language*) ▪ Relação (*relation*) ▪ Cobertura (*coverage*) ▪ Direitos (*rights*) ▪ Detentor de Direitos (*rightsHolder*) ▪ Proveniência (*provenance*)
MODS (versão 3.7)	- Esquema XML derivado do MARC 21 tido como mais rico que o DC e mais simples que o MARC completo; - Pode ser usado junto com o MADS e como um esquema de extensão do METS; e - Aplicável aos objetos de bibliotecas digitais e para *sites* arquivados, tais como os das coleções do arquivo da *Web* da Biblioteca do Congresso americano.	▪ Informação de Título (<*titleInfo*>) ▪ Nome (<*name*>) ▪ Tipo de Recurso (<*typeOfResource*>) ▪ Gênero (<*genre*>) ▪ Informação de Origem (<*originInfo*>) ▪ Língua (<*language*>) ▪ Descrição Física (<*physicalDescription*>) ▪ Resumo (<*abstract*>) ▪ Índice (<*tableOfContents*>) ▪ Nota (<*note*>) ▪ Assunto (<*subject*>)	▪ Item Relacionado (<*relatedItem*>) ▪ Identificador (<*identifier*>) ▪ Localização (<*location*>) ▪ Condição de Acesso (<*accessCondition*>) ▪ Parte (<*part*>) ▪ Extensão (<*extension*>) ▪ Informação de Registro (<*recordInfo*>)
EAD3	- Esquema XML e DTD minucioso, que é compatível com a norma de descrição arquivística ISAD(G); - Pode ser usado junto com o EAC-CPF e inclui tanto a indicação de elementos correspondentes no MARC, no MODS, na ISAD(G) e na HTML como mapeamentos (*crosswalks*) para o MARC 21, o MODS e a ISAD(G); e - Aplicável à codificação de	▪ Título da Unidade (<*unittitle*>) ▪ Origem (<*origination*>) ▪ Nome Pessoal (<*persname*>) ▪ Nome da Organização (<*corpname*>) ▪ Nome de Família (<*famname*>) ▪ Cabeçalhos de Acesso Controlado (<*controlaccess*>) ▪ Resumo (<*abstract*>) ▪ Acréscimos (<*accruals*>) ▪ Informação de Aquisição (<*acqinfo*>)	▪ Gênero/Característica Física (<*genreform*>) ▪ Descrição Física (<*physdesc*>) ▪ Objeto de Arquivo Digital (<*dao*>) ▪ Identificação da Unidade (<*unitid*>) ▪ Língua do Material (<*langmaterial*>) ▪ Língua (<*language*>) ▪ Físico Estruturado (<*physdescstructured*>)

(versão 1.1.1)	instrumentos de pesquisa de arquivos, como, por exemplo, os instrumentos de pesquisa do Arquivo *Online* da Califórnia que cedem descrições detalhadas das coleções do arquivo da *Web* da Universidade da Califórnia em Irvine nos Estados Unidos.	▪ Biografia ou História (<*bioghist*>) ▪ Escopo e Conteúdo (<*scopecontent*>) ▪ Histórico de Custódia (<*custodhist*>) ▪ Nota Descritiva de Identificação (<*didnote*> ▪ Outros Dados Descritivos (<*odd*>) ▪ Data da Unidade (<*unitdate*>)	▪ Material Relacionado (<*relatedmaterial*>) ▪ Condições de Acesso (<*accessrestrict*>) ▪ Condições de Uso (<*userestrict*>) ▪ Localização Física (<*physloc*>) ▪ Informação de Processo (<*processinfo*>) ▪ Repositório (<*repository*>) ▪ Agência de Manutenção (<*maintenanceagency*>) ▪ Histórico de Manutenção (<*maintenancehistory*>)

Quadro 1. Padrões e elementos de metadados de apoio à preservação digital no arquivamento da *Web*. (conclusão)

PADRÃO	CARACTERÍSTICAS	ELEMENTOS DE METADADOS ÚTEIS PARA A PRESERVAÇÃO DIGITAL NO ARQUIVAMENTO DA *WEB*	
VRA *Core* (versão 4.0)	- Esquema XML simples que pode ser usado junto com o CCO, tendo a indicação de elementos equivalentes no CCO, no CDWA e no DC; e - Aplicável à descrição de obras culturais originais e as suas reproduções, como certas coleções da biblioteca da Universidade de Cornell, que têm um arquivo da *Web* com os *websites* destas.	▪ Obra, Coleção ou Imagem (<*work*>, <*collection*>, <*image*>) ▪ Agente (<*agent*>) ▪ Contexto Cultural (<*culturalContext*>) ▪ Data (<*date*>) ▪ Descrição (<*description*>) ▪ Inscrição (<*inscription*>) ▪ Localização (<*location*>) ▪ Material (<*material*>) ▪ Medidas (<*measurements*>)	▪ Relação (<*relation*>) ▪ Direitos (<*rights*>) ▪ Fonte (<*source*>) ▪ Estado/Edição (<*stateEdition*>) ▪ Período/Estilo (<*stylePeriod*>) ▪ Assunto (<*subject*>) ▪ Técnica (<*technique*>) ▪ Referência Textual (<*textref*>) ▪ Título (<*title*>) ▪ Tipo de Obra (<*worktype*>)
PREMIS (versão 3.0)	- Esquema XML que enfoca o repositório de preservação e a sua gestão; - Pode unir-se a outros padrões, como MODS, DC, EAD, METS etc., para cobrir metadados fora do seu escopo e funções adicionais; e - Aplicável ao apoio da preservação de objetos digitais, tal como na descrição de ambientes de renderização para conteúdos da *Web*.	▪ *objectIdentifier/Category* ▪ *preservationLevel* ▪ *significantProperties* ▪ *objectCharacteristics* ▪ *originalName* ▪ *storage* ▪ *signatureInformation* ▪ *environmentFunction/Designation/Registry/Extension* ▪ *relationship* ▪ *linkingEventIdentifier/RightsStatementIdentifier*	▪ *eventIdentifier/Type/DateTime* ▪ *eventDetailInformation/OutcomeInformation* ▪ *linkingAgentIdentifier/ObjectIdentifier* ▪ *agentIdentifier/Name/Type/Version/Note/Extension* ▪ *linkingEventIdentifier/RightsStatementIdentifier/EnvironmentIdentifier* ▪ *rightsStatement/Extension*
	- Esquema XML flexível que organiza e vincula formas de metadados aos objetos num sistema; - Pode estruturar os pacotes PSI, PAI ou PDI do OAIS e incluir padrões na seção de Metadados Descritivos, como DC, e ter o PREMIS na seção de	▪ Agente (<*agent*>) ▪ Identificador Alternativo (<*altRecordID*>) ▪ Referência de Metadados (<*mdRef*>) ▪ Invólucro (*wrapper*) de Metadados (<*mdWrap*>) ▪ Metadados Técnicos (<*techMD*>)	▪ Localização de Arquivo (<*FLocat*>) ▪ Conteúdo de Arquivo (<*FContent*>) ▪ Fluxo de Bytes (*byte stream*) de Componente (<*stream*>) ▪ Arquivo de Transformação (<*transformFile*>) ▪ Divisão (<*div*>) ▪ Indicador (*pointer*) de Arquivo (<*fptr*>)

METS (versão 1.12.1)	Metadados Administrativos; e - Aplicável à transferência e à preservação digital do conteúdo de captura da *Web* (*websites*) entre repositórios através do perfil METS do projeto *ECHO DEPository*.	▪ Metadados de Direitos de Propriedade Intelectual (<*rightsMD*>) ▪ Metadados de Fonte (<*sourceMD*>) ▪ Metadados de Proveniência Digital (<*digiprovMD*>) ▪ Grupo de Arquivo (<*fileGrp*>) ▪ Arquivo (<*file*>)	▪ Indicador (*pointer*) METS (<*mptr*>) ▪ Link do Mapa Estrutural (<*smLink*>) ▪ Comportamento (<*behavior*>) ▪ Definição de Interface (<*interfaceDef*>) ▪ Mecanismo (<*mechanism*>)

Fonte: os autores.

Primeiramente, é importante ressaltar que todos os padrões de metadados analisados no Quadro 1 são expressos na sintaxe XML e, em certo ponto, são flexíveis e/ou extensíveis. Como padrão aberto e legível para computadores e humanos, o XML atende as necessidades de preservação digital e permite a descrição e o intercâmbio de diversos tipos de dados na *Web* e em outros ambientes; ademais, facilita a integração e o uso combinado de vários esquemas baseados nesta mesma linguagem, como os esquemas externos para uso conjunto com o METS, que incluem o EAC-CPF, MARC, MIX etc. Aliás, há a *Dublin Core Metadata Initiative* – DCMI, no caso do DC, e o *Network Development and MARC Standards Office* da Biblioteca do Congresso americano, para os outros padrões analisados (salvo o EAD e o VRA *Core* que são mantidos, nessa ordem, pelo *Technical Subcommittee for Encoded Archival Standards* – TS-EAS da *Society of American Archivists* – SAA e pelo VRA *Core Oversight Committee*), que uniformizam a descrição e a representação das informações através de esquemas de codificação de valores.

Em segundo lugar, o uso conjunto de vários padrões de metadados, impelido por indicações de metadados externos, por elementos equivalentes/correspondentes e por mapeamentos (*crosswalks*) ou pela adoção comum de sintaxes, normas e vocabulários, que retratam a flexibilidade e a extensibilidade dos esquemas e aumentam a interoperabilidade de dados, se faz aceitável, visto a alta complexidade dos tipos de recursos a serem descritos e as diversas etapas dos processos de preservação digital de longo prazo e de arquivamento da *Web*. Baseando-se na ambiguidade do escopo dos metadados de preservação, inferimos ainda que todas as classes de metadados – descritivos, linguagens de marcação (*markup languages*) etc. – são vitais ao alcance da preservação de conteúdos *Web*. Por hora, é verossímil que não tenhamos como fixar qual é o único padrão que garanta plenamente a preservação digital, mas os padrões existentes podem se completar para documentar as informações exigidas na gerência da preservação e do acesso utilizável de objetos digitais complexos, como os *websites*.

Entre os metadados assíduos nos padrões analisados estão os identificadores, que podem ser contidos na unidade *objectIdentifier* PREMIS e nos elementos Identificador e Relação DC; Item Relacionado, Identificador e Localização MODS; Material Relacionado, Identificação da Unidade e Objeto de Arquivo Digital EAD; Referência Textual VRA; Localização de Arquivo, Referência de Metadados, Definição de Interface, Mecanismo e Indicador METS. Julgando as relações de um único *website* que está sendo descrito com qualquer coleção à qual pertença ou com demais recursos, os identificadores provêm a identificação exclusiva e distintiva do recurso ao qual os metadados se referem, como a sua localização eletrônica. Logo, o registro de um URL do *site* arquivado (de acesso, captura etc.) e do URL para o recurso relacionado refletem os metadados definidos na Informação de Descrição de Preservação (referência e contexto) do modelo de informação OAIS e os princípios da preservação digital de manter o contexto e de identificar e localizar os objetos.

De fato, o DC, o MODS, o EAD e o VRA *Core* são mais cabíveis à descrição de recursos digitais para fins de sua descoberta, recuperação, apresentação e interoperabilidade. Mesmo que os escopos destes padrões de metadados estejam inerentemente voltados à etapa de acesso ao invés de exatamente para a preservação por longo prazo, alguns dos seus elementos descritivos são úteis em apoiar os metadados de preservação PREMIS. Assim, as informações

cedidas por eles permeiam aspectos de representação e de preservação, como características e dependências técnicas, alterações feitas, cadeia de custódia ou propriedade, procedência, relações estruturais físicas e lógicas, direitos etc., os quais são pertinentes na gerência de objetos digitais arquivados e que, em algum grau, traduzem parte dos contornos dos metadados de preservação OAIS e os princípios da preservação digital de garantir a fidedignidade, a autenticidade e a integridade dos objetos e de manter o contexto, a proveniência e a recuperação dos mesmos ao longo do tempo.

Sob à luz da preservação digital no arquivamento da *Web* e pautando-se nos exemplos de descrições de conteúdos *Web* arquivados de Dappert *et al*. (2013), *Digital Library Federation* (2010), Dooley e Bowers (c2018), Habing (2006) e *Library of Congress* (2018), distribuímos os elementos de metadados indicados no Quadro 1 dos padrões DC, MODS, EAD e VRA *Core* segundo as informações que eles poderão registrar para *websites* e coleções de *sites* arquivados:

- Título, Criador, Assunto, Colaborador e Língua DC; Informação de Título, Localização, Nome, Língua, Índice, Assunto e Parte MODS; Título da Unidade, Origem, Objeto de Arquivo Digital, Nome Pessoal, Nome da Organização, Nome de Família, Cabeçalhos de Acesso Controlado, Língua do Material, Língua e Repositório EAD; e Período/Estilo, Agente, Localização, Assunto e Título VRA, que exprimem o nome atribuído ao recurso descrito, o tópico, o idioma, a pessoa ou a organização responsável por criar o seu conteúdo intelectual ou fazer contribuições a ele e a instituição ou repositório que detém o recurso. Por exemplo, o nome e o idioma do *site* ou da coleção arquivada, a entidade que criou seu conteúdo ou fez contribuições secundárias e a instituição responsável pela sua seleção, curadoria ou gestão, além dos assuntos temáticos, nomes de lugares geográficos e de entidades usados para tópico principal que descreve o conteúdo arquivado ou *site*.
- Descrição, Data e Cobertura DC; Informação de Origem, Resumo e Nota MODS; Data da Unidade, Nota Descritiva de Identificação, Outros Dados Descritivos, Informação de Processo, Resumo, Acréscimos, Biografia ou História, Escopo e Conteúdo, Histórico de Custódia e Informação de Aquisição EAD; e Contexto Cultural, Localização, Descrição, Data e Período/Estilo VRA, que apontam um relato do conteúdo, do escopo e do contexto do recurso descrito, um período de tempo ligado a um evento no ciclo de vida do recurso e a cadeia de custódia, a procedência e o tópico ou o âmbito espaço-temporal do recurso. Por exemplo, a indicação de datas de direitos autorais ou de quando o *site* foi iniciado/inativado, foi/começou a ser arquivado e o URL foi capturado (com a sua frequência), além de proveniência (se um *site* faz parte de uma coleção temática mais ampla etc.) e de um decreto legal ou outra razão para selecionar o *site* ou conteúdo para arquivamento.
- Fonte e Proveniência DC; Informação de Origem MODS; Data da Unidade EAD; e Fonte VRA, que expressam uma referência à fonte das informações registradas sobre o recurso descrito e sobre um outro recurso do qual ele é derivado, as mudanças na custódia e a propriedade do recurso desde a sua criação; e a origem do recurso, incluindo local de origem/publicação, publicador e datas associadas, como a data e o local em que o *site* arquivado foi criado/emitido e a data de sua captura.
- Relação DC; Item Relacionado MODS; Material Relacionado EAD; e Relação VRA, que informam uma referência para outro recurso relacionado ao recurso que está sendo descrito. Por exemplo, as relações todo/parte entre um único *site* arquivado e qualquer coleção de *sites* arquivados a qual pertença (com a inclusão do seu título), entre um *site* arquivado e uma coleção de arquivos analógicos ou

outros materiais digitais, como as páginas *Web* constituintes do *site* e as imagens e vídeos que integram o *site*.

- Direitos e Detentor de Direitos DC; Condição de Acesso MODS; Condições de Acesso e Condições de Uso EAD; e Direitos VRA, que registram os direitos do recurso descrito, a pessoa/organização que tem ou administra tais direitos, as restrições (ou a sua falta) e as condições que afetam o acesso, a renderização e o uso do recurso. A título de exemplo, a indicação para uso no local e de um período em que o conteúdo arquivado ou *site* é restrito, se o acesso ao conteúdo está aberto e se os titulares de direitos permitem reuso após acesso.
- Tipo e Formato DC; Tipo de Recurso, Gênero e Descrição Física MODS; Cabeçalhos de Acesso Controlado, Gênero/Característica Física, Físico Estruturado e Descrição Física EAD; e Obra, Coleção ou Imagem, Material, Medidas, Tipo de Obra e Técnica VRA, que expõem a natureza do recurso descrito e o seu formato, dimensões, técnica e estilo. Por exemplo, a indicação se o conteúdo arquivado é *website*, arquivos *Web*, mídia social etc. e de que se trata de uma coleção com um número particular de *sites* arquivados.
- Extensão e Informação de Registro MODS; Localização Física, Agência de Manutenção e Histórico de Manutenção EAD; e Inscrição e Estado/Edição VRA, que documentam informações sobre o recurso com o uso de mais de um esquema, a localização física do recurso e a instituição/serviço responsável por sua criação, manutenção e divulgação, a identificação da edição do recurso e o seu histórico de criação, revisões, atualizações e outras alterações, além de informações para a gestão e a interpretação do registro de metadados, como, por exemplo, a origem do registro do *site* ou da coleção arquivada (gerado por máquina ou não etc.) e o seu idioma, data em que foi criado pela primeira vez, organização que o criou ou alterou sua versão original e regras usadas para o conteúdo da descrição (isto é, vocabulários controlados, normas de catalogação, etc.).

Apesar de estar fora dos seus propósitos, os metadados técnicos MIX, *Technical Metadata for Text* (TextMD), *Multimedia Content Description Interface* (MPEG-7), *Audio/Video Technical Metadata Extension Schema* (*Audio/Video*MD) e os dados de autoridade MADS e EAC-CPF podem ainda ser usados com o PREMIS junto ao METS no amparo ao registro das circunstâncias de criação (data de criação, nome do dispositivo de criação etc.), do histórico de mudanças feitas (documentadas, autorizadas etc.), das características e dependências técnicas (tamanho, *hardware* etc.) e dos demais aspectos de materiais audiovisuais, de texto e em mais formatos integrados nos *sites* arquivados, bem como ao registro de dados sobre agentes com funções na criação e contribuição, na seleção, curadoria ou gestão, nos direitos, na renderização e nas ações que afetam esses materiais. Sendo assim, estes padrões apoiam a interoperabilidade, a gerência e a preservação de objetos digitais complexos, como *sites* que incluem vários formatos e tipos de conteúdo, devendo ponderá-los para a definição e a validação da procedência, a autenticidade e a integridade dos seus conteúdos.

Por sua vez, o PREMIS retrata o uso prático dos conceitos de metadados de preservação delineados no modelo de informação OAIS e, em seguimento, reflete os requisitos e os princípios da preservação digital, o que faz com que todas as suas unidades semânticas sejam importantes para a preservação a longo prazo de *websites* arquivados. Por isso, inspirando-nos em Guenther, Dappert e Peyrard (c2016), que ilustram relações entre as unidades semânticas do dicionário de dados PREMIS e as categorias de informação OAIS, salientamos as unidades *significantProperties, environmentFunction/Designation/Registry/Extension* e *relationship* (informação de contexto e proveniência, informações estruturais e outras representações OAIS) que podem detalhar, por exemplo, que apenas o conteúdo precisa ser

mantido para uma página *Web*, contendo animações que não foram tidas como vitais; o ambiente que suporta a renderização e a execução de um *site*; e as relações envolvendo ambiente técnico e relações estruturais entre partes integrantes de um *site*.

Finalmente, num repositório OAIS, o METS serve como um esquema central na gestão de *websites* arquivados e na transferência destes objetos entre sistemas (ou entre sistemas e seus usuários), incluindo o DC, MODS, EAD, VRA *Core*, MARC XML etc. na seção de Metadados Descritivos como o PREMIS, MIX, *Text*MD, *Audio*MD e *Video*MD etc. na seção de Metadados Administrativos do documento METS. Na segunda seção, as unidades PREMIS (por exemplo, *eventIdentifier/Type/DateTime* ou *eventDetailInformation/OutcomeInformation*) registram, no elemento Metadados de Proveniência Digital (<*digiprovMD*>), quaisquer ações relacionadas à preservação realizadas nos vários arquivos que compõem um *site* ou quais modificações foram feitas em um objeto digital (*website*) e/ou em suas partes constituintes durante seu ciclo de vida que, segundo *Digital Library Federation* (2010), podem ser usadas para julgar como esses processos alteraram ou corromperam a capacidade do objeto de representar com precisão o item original.

Assim, os metadados descritivos e administrativos (<*mdRef*> e <*mdWrap*>; <*techMD*>, <*rightsMD*>, <*sourceMD*> e <*digiprovMD*>) podem ser externos ao documento METS, sendo que estes últimos registram relações de original/derivado entre arquivos, como arquivos foram criados e armazenados etc. Úteis aos requisitos da preservação digital, as seções de Cabeçalho METS e Arquivo (<*agent*> e <*altRecordID*>; <*transformFile*>, <*fileGrp*>, <*file*>, <*FLocat*>, <*FContent*> e <*stream*>) incluem metadados sobre o documento METS em si e listam (por formato etc.) arquivos que formam o conteúdo de *websites*. Outras seções são Comportamento (<*behavior*>, <*interfaceDef*> e <*mechanism*>) para renderizar ou exibir o *site* e Mapa Estrutural e Ligações Estruturais (<*div*>, <*fptr*> e <*mptr*>; <*smLink*>), que ordenam os *hiperlinks* entre arquivos que compõem os objetos ou entre outros objetos, como uma página *Web* com imagem hiperligada à outra página *Web*, registrando a estrutura de hipertexto dos *sites* arquivados separada dos arquivos HTML do próprio *site* e que pode ser mostrada aos usuários para sua compreensão e para a navegação do conteúdo (DIGITAL LIBRARY FEDERATION, 2010; LIBRARY OF CONGRESS, 2017).

6 CONSIDERAÇÕES FINAIS

Na prática, os principais problemas da preservação digital derivam das particularidades dos objetos aos quais se pretende manter o acesso, a recuperação e o uso no tempo. Um dos exemplos de objetos digitais complexos são os *sites* que contêm tanto uma ampla gama de *links* de hipertexto para permitir a navegação de uma página *Web* para outra, como vários arquivos e formatos com alta dependência de tecnologias para o seu acesso, interpretação, renderização e uso, que se tornam obsoletas com o tempo; aliás, estão sujeitos à dinâmica e à efemeridade da *Web*, onde os seus conteúdos são criados e publicados e, por isto, são perdidos ou sofrem rapidamente alterações em sua forma original. Logo, essas facetas obrigam a refletir sobre as questões de autenticidade, de integridade e de contexto dos *websites* arquivados e, ainda, a elucidar as distinções entre *sites* ao vivo e suas versões fixas arquivadas, como a utilidade de abordagens mistas de descrição para um único *site* ou uma coleção arquivada devido a sua heterogeneidade.

Como um dos aspectos sobre a garantia de preservação digital, a adoção de metadados para a preservação por longo prazo auxilia nas tomadas de decisões e no controle de requisitos legais, de versões, da continuidade de acesso, uso e interpretação e de outras questões atreladas ao arquivamento de objetos em sistemas. Os esquemas de metadados podem prover a interoperabilidade de objetos entre repositórios/serviços, incentivar o uso comum de vocabulários, tesauros e listas controladas (padrões de valor de dados) – como LCSH, *Internet MIME types*, ISO 639. –, ou normas, regras e códigos de catalogação bibliográfica e arquivística (padrões de conteúdo de dados) – tal como o RDA – e permitir a descrição conjunta ou a

inclusão de metadados de outros esquemas XML com indicações para metadados externos, como no elemento Extensão MODS.

Resumindo, a pesquisa feita identificou, sistematizou e analisou padrões e esquemas de metadados para o arquivamento da *Web*, debatidos na Ciência da Informação e em áreas afins. Além disso, indicou que os metadados descritivos e técnicos DC, MODS, EAD, VRA *Core*, MIX etc. e os dados de autoridade MADS e EAC-CPF têm uma aplicação mais voltada a apoiar o PREMIS e o METS, seja em permitir a identificação e a localização seja em ceder dados técnicos, de renderização, integridade e fixidez, direitos e agentes com funções nas ações que afetam *sites* arquivados. Também concluiu que, incorporando metadados descritivos, estruturais e administrativos (e de preservação, como o PREMIS), o METS é útil em simplificar a ordenação e a gerência das partes constituintes dos *sites* e de seus metadados, vincular de forma hierárquica os distintos arquivos (textos, imagens etc.) que compõem os *sites* e, em adição, gerir tais objetos complexos, atuando como um PSI, PAI e PDI num OAIS.

Por outro lado, através da literatura referenciada, constatamos algumas desvantagens dos padrões de metadados identificados neste trabalho que podem ser adaptados e/ou aplicados à preservação digital e ao arquivamento da *Web*: o DC sofre críticas à sua estrutura e ao conjunto muito simplista e genérico de elementos (sobretudo, frente a outros formatos, como o MARC); no MODS, as conversões de registros MARC original para este padrão e depois o retorno para MARC podem resultar em perda de dados ou em alguma perda de especificidade na marcação; no EAD, a ausência de recursos e de conhecimento numa instituição podem influenciar a sua utilização; o VRA *Core* detém especificidade, impõe certas restrições à criação de *links* para registros não VRA *Core* e é menos comum em comparação com os demais formatos; no PREMIS, a falta de treinamento/*expertise* e de integração com o sistema existente podem trazer barreiras à sua adoção; e o METS possui uma flexibilidade que causa problemas de interoperabilidade e, também, uma imperfeita correlação com o PREMIS, incluindo duplicações entre os elementos destes dois esquemas de metadados.

Ainda, como necessidades de investigação na questão de metadados dentro do arquivamento da *Web*, Dooley e Bowers (c2018) citam, por exemplo, os limites indefinidos entre metadados descritivos e outras categorias de metadados, como é o caso das datas de rastreamento, que são claramente tanto descritivas quanto técnicas; e quais os tipos de metadados a serem capturados, incluindo como eles são extraídos, mesclados com metadados descritivos e tornados inteligíveis aos usuários finais. Reforçando o argumento das autoras, propomos, além de maiores estudos sobre este tema recente na literatura científica nacional e internacional, que as novas pesquisas explanem os dilemas e as soluções tomadas na implementação de cada padrão para o escopo de conteúdos arquivados da *Web*. Ademais, as pesquisas deveriam examinar de que forma os metadados dos padrões identificados no trabalho podem ser melhor harmonizados, evitando problemas de duplicações e de redundâncias, visto que a análise dos resultados indicou que DC, MODS, EAD e VRA *Core* ampararam METS e PREMIS na descoberta e na documentação de aspectos técnicos dos *websites* arquivados e na comprovação de sua autenticidade, contexto e proveniência.

Enfim, diferentes tipos de metadados são importantes no arquivamento da *Web*, mas este trabalho focou os metadados descritivos e administrativos (sobretudo de preservação). Certos elementos de metadados ou unidades semânticas dos padrões identificados puderam ser sinalizados nesta pesquisa como sendo úteis à preservação de *websites* em sistemas de arquivamento digital. Por exemplo, no DC, os elementos indicados no Quadro 1 incluem informações definidas nas unidades do dicionário de dados PREMIS, como os direitos autorais e seus titulares, a identificação única e persistente, as relações todo/parte e de derivação e as dependências técnicas do objeto digital. Aliás, o DC mostrou ser um expoente para o arquivamento da *Web* por suas semelhanças com os elementos WAN de Dooley e Bowers (c2018) e, no uso dos elementos de Kim e Lee (2007), no *Internet Archive*, no Arquivo.pt e em outras iniciativas notáveis da área.

Assim, os resultados do trabalho proporcionam um respaldo teórico, técnico e estruturado de padrões e de esquemas de metadados, que podem ser usados em arquivos da *Web* concebidos para atender a preservação e fornecer o acesso duradouro de conteúdos *Web* arquivados. Tanto os elementos de metadados como as unidades semânticas apontadas na pesquisa para preservação digital no arquivamento da *Web* colaborarão para a escolha dos padrões de metadados de acordo com as necessidades das organizações públicas, privadas, sem fins lucrativos, de pesquisa e patrimônio cultural que estão interessadas e/ou envolvidas em iniciativas nacionais e internacionais na área ou, ainda, para a percepção das informações a serem previstas e exigidas para assegurar a descrição, a preservação e a gestão consistente dos *sites* arquivados num sistema que foram selecionados e coletados a partir de um domínio eletrônico, evento, local ou tópico (ciência e tecnologia etc.).

Portanto, evidencia-se que a garantia de preservação digital no arquivamento da *Web* só será factível com a adoção efetiva de padrões de metadados em suporte à administração do arquivamento e da manutenção do acesso permanente e utilizável dos conteúdos *Web* no tempo. Estas estruturas de descrição definirão a identidade e a persistência, a coerência e a compreensibilidade, o acesso e a representação, as funcionalidades, a autenticidade, a integridade e a confiabilidade, o contexto e a proveniência de *websites* selecionados, coletados e armazenados em sistemas de informação para preservação, além de determinarem a descoberta, a recuperação, a apresentação, a navegação e a arquivabilidade de *websites*, como a interoperabilidade semântica entre sistemas

REFERÊNCIAS

ALEMNEH, Daniel Gelaw; HASTINGS, Samantha Kelly. Exploration of adoption of preservation metadata in cultural heritage institutions: case of PREMIS. **Proceedings of the American Society for Information Science and Technology**, v. 47, n. 1, p. 1-8, Nov./Dec. 2010.

ALLISON-BUNNELL, Jodi. Review of Encoded Archival Description Tag Library: version EAD3. **Journal of Western Archives**, v. 7, n. 1, p. 1-4, 2016.

ALVES, Rachel Cristina Vesú. **Metadados como elementos do processo de catalogação**. 2010. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, 2010.

ALVES, Rachel Cristina Vesú. Metadados para representação e recuperação da informação em ambiente Web. *In*: MARINGELLI, Isabel Cristina Ayres da Silva. (org.). **IV Seminário Serviços de Informação em Museus**: informação digital como patrimônio cultural. São Paulo: Pinacoteca de São Paulo, 2017. p. 95-106.

ARQUIVO.PT. **Metadados acerca dos conteúdos**. [*S. l.*], ago. 2018.

BAILEY, Jefferson; LACALLE, Maria. Don't warc away: preservation metadata and web archives. *In*: AMERICAN LIBRARY ASSOCIATION (ALA) ANNUAL CONFERENCE, 16., June 2015, San Francisco, California. **Proceedings** […]. San Francisco, California: ALA, 2015. p. 1-46.

BANOS, Vangelis *et al*. CLEAR: a credible method to evaluate website archivability. *In*: INTERNATIONAL CONFERENCE ON PRESERVATION OF DIGITAL OBJECTS (iPRES), 10, May c2013, Lisboa, Portugal. **Proceedings** […]. Lisboa, Portugal: iPRES, 2010. p. 9-18.

CANTARA, Linda. METS: the metadata encoding and transmission standard. **Cataloging & Classification Quarterly**, Philadelphia, v. 40, n. 3/4, p. 237-253, 2005.

CAPLAN, Priscilla. **Understanding PREMIS**. [Washington, DC]: Library of Congress Network Development and MARC Standards Office, 2017. 22 p.

CASTRO, Fabiano Ferreira de Castro. **Elementos de interoperabilidade na catalogação descritiva**: configurações contemporâneas para a modelagem de ambientes informacionais digitais. 2012. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, SP, 2012.

CHAN, Lois Mai; ZENG, Marcia Lei. Metadata interoperability and standardization: a study of methodology part i: achieving interoperability at the schema level. **D-Lib Magazine**, [*S. l.*], v. 12, n. 6, June c2006.

CHEN, Mingyu; REILLY, Michele. Implementing METS, MIX, and DC for sustaining digital preservation at the University of Houston Libraries. **Journal of Library Metadata**, [*S. l.*], v. 11, n. 2, p. 83-99, May 2011.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. The evolution of web archiving. **International Journal on Digital Libraries**, v. 18, n. 3, p. 191-205, Sept. 2017.

DAPPERT, Angela *et al*. Describing and preserving digital object environments. **New Review of Information Networking**, Philadelphia, v. 18, n. 2, p. 106-173, Oct. 2013.

DAPPERT, Angela; ENDERS, Markus. Digital preservation metadata standards. **Information Standards Quarterly (ISQ)**, v. 22, n. 2, p. 4-13, spring 2010.

DI PRETORO, Emmanuel; GEERAERT; Friedel. Behind the scenes of web archiving: metadata of harvested websites. Archives et Bibliothèques de Belgique – Archief – En Bibliotheekwezen in Belgie; Archief, in press, trust an Undertanding: The value of metadata en a digitally joined-up world. 2019.

DIGITAL LIBRARY FEDERATION. **<METS> Metadata Encoding and Transmission Standard**: primer and reference manual. Version 1.6. [Washington, DC], 2010. 144 p.

DIGITAL PRESERVATION COALITION. **Metadata**. [Glasgow, Scotland], [201-?]. 2 p. (Digital Preservation Topical Notes, 5).

DOOLEY, Jackie M. *et al*. Developing web archiving metadata best practices to meet user needs. **Journal of Western Archives**, [Provo], v. 8, n. 2, p. 1-14, 2017.

DOOLEY, Jackie; BOWERS, Kate. **Descriptive metadata for web archiving**: recommendations of the oclc research library partnership web archiving metadata working group. Dublin, Ohio: Online Computer Library Center (OCLC) Research, Feb. c2018. 53 p.

DOORN, Peter; TJALSMA, Heiko. Introduction: archiving research data. **Arch Sci**, v. 7, p. 1-20, Sept. 2007.

DUBLIN CORE METADATA INITIATIVE. About DCMI. **DCMI History**. [*S. l.*], June c2020a.

DUBLIN CORE METADATA INITIATIVE. DCMI Usage Board. Specifications. **DCMI Metadata Terms**. [*S. l.*], Jan. 2020b.

DUBLIN CORE METADATA INITIATIVE. DCMI Usage Board. Specifications. **Dublin Core Metadata Element Set, Version 1.1***:* reference description. [*S. l.*], June 2012.

EIDSON, Jennifer G.; ZAMON, Christina J. EAD twenty years later: a retrospective of adoption in the early twenty-first century and the future of ead. **The American Archivist**, v. 82, n. 2, p. 303-330, 2019.

EÍTO-BRUN, Ricardo. A metadata infrastructure for a repository of civil engineering records: eac-cpf as a cornerstone for content publishing. **Journal of Archival Organization**, v. 12, n. 1-2, p. 62-76, 2015.

FORMENTON, Danilo *et al*. Os padrões de metadados como recursos tecnológicos para a garantia da preservação digital. **Biblios**, Pittsburgh, n. 68, p. 82-95, jul. 2017.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 184 p.

GILLILAND, Anne J. Setting the stage. *In*: BACA, Murtha. (ed.). **Introduction to metadata**. 3rd ed. Los Angeles, California: Getty Publications, c2016. 92 p.

GUENTHER, Rebecca S. MODS: the metadata object description schema. **Portal**: Libraries and the Academy, v. 3, n. 1, p. 137-150, Jan. 2003.

GUENTHER, Rebecca Squire; DAPPERT, Angela; PEYRARD, Sébastien. An introduction to the PREMIS data dictionary for digital preservation metadata. *In*: GUENTHER, Rebecca Squire; DAPPERT, Angela; PEYRARD, Sébastien. **Digital preservation metadata for practitioners**. Cham, Switzerland: Springer, Dec. c2016. p. 23-36.

GUENTHER, Rebecca; MYRICK, Leslie. Archiving web sites for preservation and access: MODS, METS and MINERVA. **Journal of Archival Organization**, v. 4, n. 1/2, p. 141-166, 2007.

HABING, Thomas G. **ECHO Dep METS Profile for Web Site Captures**. [*S. l.*], 2006.

HARPER, Corey A. Dublin Core Metadata Initiative: beyond the element set. **Information Standards Quarterly (ISQ)**, v. 22, n. 1, p. 19-28, winter 2010. Acesso em: 2 jun. 2020.

INTERNET ARCHIVE. Internet Archive APIs. About Archive.org metadata. **Internet archive metadata**. [San Francisco, California], Dec. 2018.

KIM, Heejung; LEE, Hyewon. Development of metadata elements for intensive web archiving. **Journal of the Korean Society for Information Management**, Songdo, South Korea, v. 24, n. 2, p. 143-160, June 2007.

LAVOIE, Brian; GARTNER, Richard. Preservation metadata. 2nd edition. **DPC Technology Watch Report**, v. 13, n. 3, p. 1-36, May c2013.

LIBRARY OF CONGRESS. **Development of the Encoded Archival Description DTD**. Dec. 2013.

LIBRARY OF CONGRESS. **METS**: an overview & tutorial. [Washington, DC], Mar. 2017.

LIBRARY OF CONGRESS. **MODS user guidelines**. MODS elements and attributes. Version 3. [Washington, DC], Aug. 2018.

LIBRARY OF CONGRESS. **MODS**: uses and features. [Washington, DC], Feb. 2016.

LIBRARY OF CONGRESS. Programs. Web archiving. About this program. Web archives. **Collections with web archives**. [Washington, DC], [2021].

LIMA, Fábio Rogério Batista; SANTOS, Plácida Leopoldina V. A. C.; SANTARÉM SEGUNDO, José Eduardo. Padrão de metadados no domínio museológico. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 3, p. 50-69, jul./set. 2016.

LUBAS, Rebecca L.; JACKSON, Amy S.; SCHNEIDER, Ingrid. Using VRA Core 4.0. *In*: LUBAS, Rebecca L.; JACKSON, Amy S.; SCHNEIDER, Ingrid. **The metadata manual**: a practical workbook. Oxford, UK: Chandos Publishing, 2013. p. 135-164.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 8. ed. atual. São Paulo: Atlas, 2017. 368 p.

MÁRDERO ARELLANO, Miguel Ángel. **Critérios para a preservação digital da informação científica**. 2008. Tese (Doutorado em Ciência da Informação) - Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2008.

MASANÈS, Julien. **Web Archiving**. Berlin: Springer, c2006. 234 p.

MCCALLUM, Sally H. An introduction to the metadata object description schema (MODS). **Library Hi Tech**, v. 22, n. 1, p. 82-88, 2004.

MCDONOUGH, Jerome P. METS: standardized encoding for digital library objects. **International Journal on Digital Libraries**, v. 6, n. 2, p. 148-158, April 2006.

MELO, Jonas Ferrigolo; ROCKEMBACH, Moisés. Arquivabilidade de websites para preservação digital: estudo a partir da área da saúde. **Reciis** – Rev Eletron Comun Inf Inov Saúde, v. 14, n. 3, p. 529-545, jul./set. 2020.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding metadata**. Bethesda, Maryland: NISO Press, c2004. 16 p.

NATIONAL LIBRARY OF NEW ZEALAND. **Metadata standards framework**: preservation metadata (revised). Wellington, New Zealand: National Library of New Zealand, June 2003. 50 p.

PALA, Francesca. Lo standard EAD3 per la codifica dei dati archivistici: qualche novità e molte conferme. **JLIS.it**, Macerata, v. 8, n. 3, p. 148-176, Sept. 2017.

PENNOCK, Maureen. Web-Archiving. **DPC Technology Watch Report**, v. 13, n. 1, p. 1-45, Mar. c2013.

PREMIS EDITORIAL COMMITTEE. **PREMIS data dictionary for preservation metadata**. Version 3.0. [*S. l.*: *s. n.*], Nov. 2015. 273 p.

RILEY, Jenn. **Understanding metadata**: what is metadata, and what is it for? Baltimore, Maryland: National Information Standards Organization (NISO), c2017. 45 p.

ROCKEMBACH, Moises; PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de preservação digital no arquivamento da web. **RICI**: R.Ibero-amer. Ci. Inf., Brasília, v. 11, n. 1, p. 168-182, jan./abr. 2018.

ROWELL, Chelcie Juliet; KREWER, Drew. Preservation metadata for complex digital objects. A Report of the ALCTS PARS Preservation Metadata Interest Group Meeting. American Library Association Annual Conference, San Francisco, June 2015. **Technical Services Quarterly**, v. 33, n. 2, p. 179-183, Mar. 2016.

SAMOUELIAN, Mary; DOOLEY, Jackie. **Descriptive metadata for web archiving**: review of harvesting tools. Dublin, Ohio: Online Computer Library Center (OCLC) Research, Feb. c2018. 23 p.

SAYÃO, Luís Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. **Enc. Bibli**: R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, v. 15, n. 30, p. 1-31, 2010.

SENANDER III, Mathew. Converting vra core records to marc records: a study in crosswalking. **Library Philosophy and Practice**, Lincoln, Dec. 2013.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 24. ed. rev. e atual. São Paulo: Cortez, 2016. 320 p.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. rev. e atual. Florianópolis: Universidade Federal de Santa Catarina (UFSC), 2005. 139 p.

SOCIETY OF AMERICAN ARCHIVISTS. Technical Subcommittee for Encoded Archival Standards. **Encoded Archival Description Tag Library**: version EAD3 1.1.1. Chicago, Dec. 2019. 422 p.

TRUMAN, Gail. **Web archiving environmental scan**. Harvard Library Report. [Cambridge, Massachusetts]: Harvard University, Jan. 2016. 83 p.

VEIKKOLAINEN, Petteri; LAGER, Lassi. Long-term preservation of the finnish web archive. *In*: INTERNATIONAL INTERNET PRESERVATION CONSORTIUM (IIPC) GENERAL ASSEMBLY, 10., April 2016, Reykjavik, Iceland. **Proceedings** […]. Reykjavik, Iceland: IIPC, 2016. p. 195-203.

VELLUCCI, Sherry L. Metadata and authority control. **Library Resources & Technical Services (LRTS)**, [Chicago], v. 44, n. 1, p. 33-43, Jan. 2000.

VENLET, Jessica *et al*. **Descriptive metadata for web archiving**: literature review of user needs. Dublin, Ohio: Online Computer Library Center (OCLC) Research, Feb. c2018. 48 p.

VISUAL RESOURCES ASSOCIATION. **An introduction to VRA Core**. VRA Core 4.0 introduction. [*S. l.*], Oct. 2014. 2 p.

VISUAL RESOURCES ASSOCIATION. **VRA Core 4.0 element description**. [*S. l.*], May 2007. 37 p.

ZENG, Marcia Lei.; QIN, Jian. **Metadata**. New York, United States: Neal-Schuman Publishers, June 2008. 365 p.