

Universidade Federal de São Carlos
Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade

**Identificação de padrões de metadados para preservação
digital**

DANILO FORMENTON

São Carlos – SP
2015

DANILO FORMENTON

Identificação de padrões de metadados para preservação digital

Dissertação apresentada ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, do Centro de Educação e Ciências Humanas, da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência, Tecnologia e Sociedade.

Orientadora: Profa. Dra. Luciana de Souza Gracioso

Coorientador: Prof. Dr. Fabiano Ferreira de Castro

São Carlos – SP
2015

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

F725ip Formenton, Danilo
 Identificação de padrões de metadados para
 preservação digital / Danilo Formenton. -- São Carlos
 : UFSCar, 2016.
 102 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2015.

1. Preservação digital. 2. Padrão de metadados. 3.
Metadados de preservação. 4. Ciência da Informação. 5.
Informação e tecnologia. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Danilo Formenton, realizada em 27/11/2015:

A handwritten signature in blue ink, appearing to read "G", is positioned above a horizontal line.

Profa. Dra. Luciana de Souza Gracioso
UFSCar

A handwritten signature in blue ink, appearing to read "C", is positioned above a horizontal line.

Profa. Dra. Ariadne Chloe Mary Furnival
UFSCar

A handwritten signature in blue ink, appearing to read "Tonic de log de meu fin", is positioned above a horizontal line.

Profa. Dra. Maria da Graça de Melo Simões
UC

A handwritten signature in blue ink, appearing to read "F", is positioned above a horizontal line.

Prof. Dr. Fabiano Ferreira de Castro
UFSCar

Dedico este trabalho à minha família e ao meu futuro.

AGRADECIMENTOS

Primeiramente, agradeço à Deus por iluminar as minhas decisões e por me dar forças ao longo de todo o trajeto percorrido para a consecução de mais este trabalho.

À minha família, principalmente aos meus pais, Noé e Rosemary, pelo apoio e carinho, pelos valores e ensinamentos recebidos e por tudo que fizeram por mim. Ao meu querido irmão Alessandro, pelos conselhos, incentivos e apoio incondicional. À minha cunhada Yaisa, pelas conversas e apoio, e ao meu sobrinho Felipe, por me trazer alegrias nos momentos difíceis.

À Profa. Dra. Luciana de Souza Gracioso, pela relevante orientação e colaboração desde a minha graduação, pela amizade, confiança, paciência e grande dedicação, pelo aprendizado obtido, incentivo e pelas valiosas oportunidades de trabalho e de apresentação de estudos.

Ao Prof. Dr. Fabiano Ferreira de Castro, pela considerável coorientação e parceria, pelo interesse, perseverança e pelas várias sugestões pertinentes durante a elaboração do trabalho.

À Profa. Dra. Maria da Graça Melo Simões da Universidade de Coimbra, pelo interesse e disposição, pelas sugestões e presteza em participar da Banca Examinadora de Defesa.

À Profa. Dra. Ariadne Chloe Mary Furnival da Universidade Federal de São Carlos (UFSCar), pela sua atenção e permanente auxílio na obtenção de fontes de informação notáveis para a pesquisa, pelas recomendações e sua prontidão em aceitar compor a Banca Examinadora de Defesa deste trabalho.

Ao Prof. Dr. Roniberto Morato do Amaral da UFSCar, pela amizade e enorme astúcia em propiciar e ocasionar o início da coorientação do Prof. Fabiano nesta dissertação.

A todos os colegas de turma e professores do Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade da UFSCar, pela amizade e convívio harmônico, pelos trabalhos feitos e pelos valores, conhecimentos e ensinamentos que levarei para toda a minha vida.

Enfim, a todos que contribuíram de diversas formas para a realização deste trabalho.

"A vida é um processo fluente e em alguns lugares do caminho coisas desagradáveis ocorrerão. Podem deixar cicatrizes, mas a vida continua a fluir. É como a água fluente, que ao estagnar-se, torna-se podre; não pare! Continue bravamente... porque cada experiência nos ensina uma lição".

Bruce Lee

RESUMO

Com o expressivo aumento na produção e aquisição de conteúdos digitais, ações de recuperação e de preservação de dados a longo prazo estão sendo criadas e estudadas defronte às mudanças tecnológicas de armazenamento e de acesso aos recursos informacionais. Nesse contexto, objetiva-se verificar em que medida padrões ou esquemas de metadados no âmbito da preservação digital têm sido discutidos pelo campo da Ciência da Informação e áreas afins (Arquivologia, Biblioteconomia, Museologia etc.), apontando de que maneira estes esquemas poderiam atender as demandas de estruturação de repositórios institucionais de maneira mais adequada para a preservação de documentos digitais. Como metodologia destaca-se a realização de um estudo exploratório, pautado em pesquisa bibliográfica e documental referente aos padrões ou esquemas de metadados utilizados no escopo da preservação digital. Aplicou-se a análise comparativa dos padrões ou esquemas de metadados identificados na literatura, baseada em elementos de metadados considerados essenciais para a preservação digital. Como resultado apresenta-se um quadro referencial teórico, técnico e sistematizado de padrões ou esquemas de metadados aplicáveis à preservação digital, tais como DC, MODS, EAD, ANSI/NISO Z39.87 ou MIX, METS e PREMIS e, ao mesmo tempo, que podem ser relevantes para a modelagem de repositórios institucionais digitais. Constatou-se que a garantia de preservação digital a longo prazo só será possível com a adoção efetiva de padrões ou esquemas de metadados, pois são eles que definirão a descrição, a representação, a consistência e a persistência do recurso/objeto digital no ambiente informacional, além de determinar a interoperabilidade entre sistemas. Não há como afirmar qual é o único esquema de metadados existente que assegure totalmente a preservação digital e, sim, que diferentes padrões podem trabalhar juntos, de modo a se auto complementarem para o registro eficaz das informações requeridas na gestão da preservação e do acesso utilizável ao longo do tempo de objetos/recursos digitais.

Palavras-chave: Preservação digital. Padrões de metadados. Metadados de preservação. Ciência da Informação. Informação e Tecnologia.

ABSTRACT

With the significant increase in the production and acquisition of digital content, initiatives for the recovery and long-term data preservation are being created and studied in the light of changes in technological storage and access to information resources. In this context, the aim of the research was to verify to what extent standards or metadata schemas for digital preservation have been discussed by the field of Information Science and related fields (Archival Science, Library Science, Museology etc.), pointing out how these schemas could meet the structuring demands of institutional repositories more adequately for the preservation of digital documents. In terms of method, an exploratory study, based on bibliographic and documentary research relating to standards or metadata schemes used in the scope of digital preservation, was conducted. A comparative analysis of standards or metadata schemes identified in the literature was carried out, based on core metadata elements considered for digital preservation. Regarding the results, a theoretical technical framework is presented, that seeks to systematize standards or metadata schemes applicable to digital preservation, such as DC, MODS, EAD, ANSI/NISO Z39.87 or MIX, METS and PREMIS and at the same time, that may be relevant for modeling digital institutional repositories. It was found that the guarantee of long-term digital preservation will only be possible with the effective implementation of standards and metadata schemas, because it is they that define the description, representation, consistency and persistence of the digital object/resource in the information environment, as well as determining interoperability between systems. There is no way to say which of the existing metadata schema fully ensures digital preservation but that it is indeed possible to have different standards working together in order to complement the effective registration of the information required in managing preservation and usable access over time of digital objects/resources.

Keywords: *Digital preservation. Metadata standards. Preservation metadata. Information Science. Information and Technology.*

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 - Ambiente SAAI..... | 42 |
| Figura 2 - Modelo de informação SAAI..... | 43 |
| Figura 3 - Modelo funcional SAAI..... | 46 |
| Figura 4 - Exemplo de um registro DC em XML..... | 66 |
| Figura 5 - Exemplo de um documento MODS em XML..... | 70 |
| Figura 6 - Exemplo de um documento EAD em XML..... | 73 |
| Figura 7 - Exemplo de um documento MIX em XML..... | 77 |
| Figura 8 - Exemplo de um documento METS contendo metadados MODS em XML..... | 80 |
| Figura 9 - Exemplo de um documento METS contendo metadados MODS, MIX e PREMIS em XML..... | 85 |

LISTA DE QUADROS

| | |
|--|----|
| Quadro 1 - Estratégias de preservação digital..... | 30 |
| Quadro 2 - Padrões ou esquemas de metadados e seus escopos..... | 62 |
| Quadro 3 - Padrões e elementos de metadados de apoio a preservação digital..... | 87 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|---------|---|
| AACR2 | Anglo-American Cataloguing Rules Second Edition |
| ABNT | Associação Brasileira de Normas Técnicas |
| AIIM | Association for Information and Image Management |
| ANSI | American National Standards Institute |
| ASCII | American Standard Code for Information Interchange |
| BND | Biblioteca Nacional Digital |
| C&T | Ciência e Tecnologia |
| CCSDS | Consultative Committee for Space Data Systems |
| CDCC | Centro de Divulgação Científica e Cultural |
| CDPI | Canadian Digital Preservation Initiative |
| CD-ROM | Compact Disc Read-Only Memory |
| CEDARS | CURL Exemplars in Digital Archives |
| CLIR | Council on Library and Information Resources |
| CNRI | Corporation for National Research Initiatives |
| CONARQ | Conselho Nacional de Arquivos |
| CTS | Ciência, Tecnologia e Sociedade |
| DC | Dublin Core |
| DCMI | Dublin Core Metadata Initiative |
| DCTA | Departamento de Ciência e Tecnologia Aeroespacial |
| DDA | Digital Data Archive |
| DDC | Dewey Decimal Classification |
| DGLAB | Direção-Geral do Livro, dos Arquivos e das Bibliotecas |
| DLF | Digital Library Federation |
| DOI | Digital Object Identifier |
| DPC | Digital Preservation Coalition |
| DPOE | Digital Preservation Outreach and Education |
| DTD | Document Type Definition |
| EAD | Encoded Archival Description |
| FEDORA | Flexible Extensible Digital Object Repository Architecture |
| HTML | HyperText Markup Language |
| IAE | Instituto de Aeronáutica e Espaço |
| IBICT | Instituto Brasileiro de Informação em Ciência e Tecnologia |
| ID | Informação Descritiva |
| IDF | International DOI Foundation |
| IEC | International Electrotechnical Commission |
| IES | Instituições de Ensino Superior |
| IETF | Internet Engineering Task Force |
| IFLA | International Federation of Library Associations and Institutions |
| IIPC | International Internet Preservation Consortium |
| INPE | Instituto Nacional de Pesquisas Espaciais |
| IR | Informação de Representação |
| ISAD(G) | General International Standard Archival Description |

| | |
|---------|---|
| ISBN | International Standard Book Number |
| ISO | International Organization for Standardization |
| JISC | Joint Information Systems Committee |
| JPEG | Joint Photographic Experts Group |
| LAC | Library and Archives Canada |
| LC | Library of Congress |
| LCC | Library of Congress Classification |
| LCSH | Library of Congress Subject Headings |
| LCWA | Library of Congress Web Archives |
| LOM | Learning Object Metadata |
| MARC | Machine Readable Cataloging |
| MeSH | Medical Subject Headings |
| METS | Metadata Encoding and Transmission Standard |
| MIT | Massachusetts Institute of Technology |
| MIX | Metadata for Images in XML |
| MOA2 | Making of America II |
| MODS | Metadata Object Description Schema |
| MPEG | Moving Picture Experts Group |
| MS-DOS | MicroSoft Disk Operating System |
| NASA | National Aeronautics and Space Administration |
| NCSA | National Center for Supercomputing Applications |
| NDIIPP | National Digital Information Infrastructure and Preservation Program |
| NDNP | National Digital Newspaper Program |
| NDSA | National Digital Stewardship Alliance |
| NEDLIB | Networked European Deposit Library |
| NISO | National Information Standards Organization |
| NLM | National Library of Medicine |
| NPES | Association for Suppliers of Printing, Publishing and Converting Technologies |
| OAIS | Open Archival Information System |
| OCLC | Online Computer Library Center |
| ODL | Oxford Digital Library |
| PAI | Pacote de Arquivamento de Informação |
| PANDORA | Preserving and Accessing Networked Documentary Resources of Australia |
| PDF | Portable Document Format |
| PDF/A | Portable Document Format/Archive |
| PDI | Pacote de Disseminação de Informação |
| PREMIS | PREservation Metadata: Implementation Strategies |
| PSI | Pacote de Submissão de Informação |
| PURL | Persistent Uniform Resource Locators |
| RDA | Resource Description and Access |
| RDF | Resource Description Framework |
| REACH | Record Export for Art and Cultural Heritage |
| RFC | Request for Comments |
| RFSA | Russian Federal Space Agency |

| | |
|--------|--|
| RLG | Research Libraries Group |
| RLUK | Research Libraries United Kingdom |
| RODA | Repositório de Objectos Digitais Autênticos |
| SAA | Society of American Archivists |
| SAAI | Sistema Aberto de Arquivamento de Informação |
| SATIS | Science and Technology in Society |
| SGML | Standard Generalized Markup Language |
| SICI | Serial Item and Contribution Identifier |
| SISCON | Science in a Social Context |
| SPAR | Scalable Preservation and Archiving Repository |
| TEL | The European Library |
| TGN | Thesaurus of Geographic Names |
| TIC | Tecnologias de Informação e Comunicação |
| TIFF | Tagged Image File Format |
| UDC | Universal Decimal Classification |
| UFSCar | Universidade Federal de São Carlos |
| UHDL | University of Houston Digital Library |
| UMinho | Universidade do Minho |
| UPF | Universal Preservation Format |
| URL | Uniform Resource Locator |
| URN | Uniform Resource Name |
| US | United States |
| USGS | United States Geological Survey |
| USP | Universidade de São Paulo |
| W3C | World Wide Web Consortium |
| XML | Extensible Markup Language |

SUMÁRIO

| | |
|--|-----------|
| 1 INTRODUÇÃO | 15 |
| 2 A PRESERVAÇÃO DIGITAL NO CONTEXTO CTS..... | 19 |
| 3 PRESERVAÇÃO DIGITAL: revisitando conceito, problemas, requisitos, estratégias, modelo de referência SAAI e políticas..... | 23 |
| 3.1 Problemas da preservação digital | 24 |
| 3.2 Requisitos para a preservação digital | 26 |
| 3.3 Estratégias para a preservação digital..... | 29 |
| 3.3.1 Estratégias estruturais | 30 |
| 3.3.2 Estratégias operacionais | 35 |
| 3.4 Modelo de referência SAAI..... | 40 |
| 3.4.1 Esquema SAAI | 41 |
| 3.4.2 Ambiente SAAI | 41 |
| 3.4.3 Modelo de informação SAAI | 42 |
| 3.4.4 Modelo funcional SAAI | 46 |
| 3.5 Subsídios para construção de políticas de preservação digital | 48 |
| 4 METADADOS: definições, classificações, metadados de preservação e esquemas | 54 |
| 4.1 Definição e classificação para metadados | 55 |
| 4.2 Metadados para a preservação digital..... | 57 |
| 4.3 Esquemas de metadados | 60 |
| 4.3.1 Padrão <i>Dublin Core</i> | 63 |
| 4.3.2 Padrão MODS | 67 |
| 4.3.3 Padrão EAD | 71 |
| 4.3.4 Padrão ANSI/NISO Z39.87 | 74 |
| 4.3.5 Padrão METS | 78 |
| 4.3.6 Padrão PREMIS..... | 81 |
| 4.4 Análise dos padrões de metadados à luz da preservação digital | 86 |
| 5 CONSIDERAÇÕES FINAIS..... | 90 |
| REFERÊNCIAS | 93 |

1 INTRODUÇÃO

Com o advento e a expansão do acesso à *Internet* em consonância com o uso intensivo das Tecnologias de Informação e Comunicação (TIC), no desenvolvimento de ambientes informacionais tornou-se notório o expressivo aumento gradativo na produção de recursos informacionais, principalmente em formato digital. Garantir a autenticidade e a integridade dos recursos digitais, é apontado pela comunidade científica como um dos grandes desafios do século XXI. Soma-se a isto a necessidade de promover a sua organização, a disponibilização e a preservação a longo prazo de modo adequado, que permita assegurar o acesso ao conteúdo informacional, amenizando assim os possíveis riscos de sua perda permanente.

Ademais, os objetos/recursos digitais são dinâmicos, tal como o próprio ambiente onde são criados, transportados e armazenados caracterizando-se pela vulnerabilidade e vertiginosa obsolescência, tornando muito complexo o propósito de proteger estes objetos e preservar o seu acesso ao longo do tempo. Nesta perspectiva, a preservação digital vêm sendo alvo de pesquisas pela comunidade científica, conforme podemos constatar em Almeida, Cendón e Souza (2012), Barbedo, Corujo e Sant’ana (2011), Campos e Saramago (2007), Grácio (2012) e Grácio, Fadel e Valentim (2013). Consideramos que a preservação digital trata-se do armazenamento, da manutenção e do acesso aos recursos digitais por longo prazo, em virtude do uso de uma ou mais estratégias de preservação (RUSSELL; SERGEANT, 1999).

Uma das estratégias de preservação digital existentes refere-se à adoção de padrões de metadados, estruturas formais de descrição de recursos para comunidades específicas, de modo a apoiar a gestão e a preservação de objetos digitais em ambientes informacionais, tais como os repositórios institucionais. A título de exemplo o Repositório de Objectos Digitais Autênticos (RODA), da Universidade do Minho (UMinho) e da Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB) em Portugal, e *University of Houston Digital Library* (UHDL), nos Estados Unidos, são projetos onde há o uso combinado de vários esquemas, como *Dublin Core* (DC); *Metadata Object Description Schema* (MODS); *Encoded Archival Description* (EAD); *American National Standards Institute* (ANSI)/*National Information Standards Organization* (NISO) Z39.87 ou *Metadata for Images in XML* (MIX); *Metadata Encoding and Transmission Standard* (METS) e *PREServation Metadata: Implementation Strategies* (PREMIS).

Marcondes e Sayão (2009, p. 10) compreendem que os repositórios institucionais são “[...] elementos de uma rede ou infraestrutura informacional de um país ou de um domínio institucional destinados a garantir a guarda, preservação a longo prazo e, fundamentalmente, o livre acesso à produção científica de uma dada instituição”.

Na Ciência da Informação, campo que trata da informação e sua recuperação, as poucas publicações nacionais se pautam numa discussão mais conceitual relatando a preservação digital como o caminho mais apropriado para fins de assegurar a preservação e o acesso, por longo período, ao conteúdo dos documentos em formato digital.

Nesta acepção, o campo de estudos Ciência, Tecnologia e Sociedade (CTS) se relaciona ao tema tratado na dissertação, visto que dispor o acesso aos conteúdos científico-tecnológicos é um fator relevante para uma visão crítica-avaliativa sobre as interações entre CTS. Portanto, as iniciativas voltadas à preservação digital, em especial as desenvolvidas por Instituições de Ensino Superior (IES) e demais instituições públicas, se justificam tanto pela necessidade de assegurar a promoção de acesso aos objetos digitais a longo prazo como ainda, atualmente, pela necessidade de atender a legislação vigente que assegura o direito ao acesso à informação (Lei de acesso à informação n. 12.527/2011). Sendo assim, haja vista as demandas de ações de cooperação e da estratégia da adoção de metadados para a preservação digital, torna-se importante discutir detalhadamente os padrões de metadados que poderiam ser previstos, aplicados e mesclados para satisfazer as especificidades das instituições que estejam desenvolvendo seus repositórios institucionais.

A partir deste cenário, a questão de pesquisa a que esta investigação se propôs a responder é: quais padrões ou esquemas de metadados poderiam ser considerados pelas instituições que estão desenvolvendo seus repositórios institucionais, para que estas pudessem contemplar a preservação digital?

Parte-se da hipótese que há uma certa carência, no âmbito dos estudos nacionais de preservação digital na Ciência da Informação e áreas afins, que discutam profundamente e sistematizem os metadados e as características de esquemas de metadados aplicáveis à preservação digital. Teve-se como pressuposto que não seria possível identificar apenas um padrão de metadados que atenda todas as especificidades de uma particular instituição, sendo preciso propor medidas combinadas de diferentes esquemas para a modelagem de repositórios institucionais que visam contemplar a preservação digital, em razão da diversidade de materiais digitais que poderão ser tratados e das ações relativas aos mesmos que deverão ser feitas (descrição, gestão, preservação etc.).

Os metadados de preservação são de grande importância estratégica na busca pela preservação a longo prazo dos recursos digitais garantindo o acesso, a autenticidade e a integridade do seu conteúdo informacional, uma vez que o ambiente digital caracteriza-se pela ampla complexidade, propiciando dificuldades na pesquisa, recuperação e armazenamento por longo período das informações contidas em materiais neste domínio.

Todavia, os metadados de preservação ainda precisam ser mais investigados e discutidos pela Ciência da Informação no cenário nacional, o que torna esta dissertação fundamental, pois a mesma permite estabelecer um mapeamento das informações sobre o uso de padrões de metadados na preservação digital. Ademais, este estudo propicia o acesso não apenas às pesquisas relativas ao assunto, mas também um diagnóstico que auxiliará tanto na coordenação de iniciativas e medidas futuras nesse contexto, como nas diretrizes existentes no planejamento de instituições patrimoniais (bibliotecas, arquivos e museus, por exemplo) e de organizações que pretendam implementar um programa ou política de preservação digital.

Neste sentido, esta dissertação teve o objetivo geral de verificar em que medida padrões ou esquemas de metadados no âmbito da preservação digital têm sido discutidos pela Ciência da Informação e áreas afins (Arquivologia, Biblioteconomia, Museologia etc.), apontando de que modo estes esquemas poderiam atender as demandas de estruturação de repositórios institucionais de maneira mais adequada para a preservação de documentos digitais. O resultado deste mapeamento e análise vislumbrou contribuir para possíveis delimitações do conjunto de diretrizes e políticas a serem adotadas pelas instituições interessadas e/ou envolvidas com a preservação e o acesso ao longo do tempo de informações digitais.

Para tal fim foram determinados os seguintes objetivos específicos:

- a. Identificar na literatura nacional e internacional as pesquisas em preservação digital no campo da Ciência da Informação e áreas afins;
- b. Identificar, sistematizar e analisar os padrões ou esquemas de metadados articulados no âmbito da preservação digital de longo prazo; e
- c. Sinalizar, a partir dos padrões identificados, os metadados que são potencialmente aplicáveis à modelagem de repositórios institucionais.

Para atingir os objetivos propostos optou-se metodologicamente pelo desenvolvimento de uma investigação exploratória, pautada em pesquisa bibliográfica e documental concernente aos padrões ou esquemas de metadados utilizados no escopo da preservação digital. A partir de Sellitz et al. (1967 citado por GIL, 2010, p. 27), compreendemos que os estudos exploratórios buscam “[...] proporcionar maior familiaridade como o problema, com vistas a torná-lo mais explícito ou a construir hipóteses. Seu planejamento tende a ser bastante flexível, pois interessa considerar os mais variados aspectos relativos ao fato ou fenômeno estudado”. Por sua vez, as investigações bibliográficas são efetuadas através de materiais já publicados, como livros, teses, dissertações, artigos etc., utilizando-se de informações ou de categorias teóricas documentadas e manipuladas por outros pesquisadores; as pesquisas documentais abrangem como fonte uma

ampla gama de documentos criados para propósitos variados, tal como material institucional, jurídico, de divulgação, iconográfico, itens pessoais etc. (GIL, 2010; SEVERINO, 2007).

Como forma de análise de resultados dos materiais, aplicou-se a análise comparativa dos padrões ou esquemas de metadados identificados na literatura, baseada em elementos de metadados considerados essenciais para a preservação digital.

A dissertação está organizada do seguinte modo:

O primeiro capítulo faz uma introdução com relação ao tema de investigação, elucida a questão de pesquisa, a hipótese, os objetivos, a metodologia e as justificativas.

No segundo capítulo é discutido a relação entre a área de estudos CTS e a concepção de preservação digital na Ciência da Informação.

O terceiro capítulo aborda a preservação digital, no tocante aos problemas envolvidos, os requisitos necessários, as estratégias existentes, o modelo de referência *Open Archival Information System* (OAIS) ou Sistema Aberto de Arquivamento de Informação (SAAI) e os subsídios para a construção de políticas.

No quarto capítulo discute-se metadados no âmbito da preservação digital, incluindo definições e classes para metadados, os metadados de preservação e os principais padrões utilizados, tais como DC, MODS, EAD, ANSI/NISO Z39.87 ou MIX, METS e PREMIS, além disto, através de uma análise dos esquemas de metadados supracitados, sinaliza os metadados que podem ser relevantes no suporte à preservação digital.

No quinto capítulo temos as considerações finais sobre os resultados da pesquisa, assim como apontamentos acerca do conhecimento obtido e sugestões para futuros trabalhos.

2 A PRESERVAÇÃO DIGITAL NO CONTEXTO CTS

Até a década de 1970 as tomadas de decisões associadas à ciência e à tecnologia (C&T) refletiam fundamentalmente aos anseios e as vontades dos próprios cientistas ou especialistas, entretanto, após as sequelas deixadas pelas guerras e pela devastação do meio ambiente, o apoio incondicional e a perspectiva puramente romântica, positiva e otimista da qual se tinha acerca dos avanços em C&T acabou se modificando.

Parte das sociedades passaram a dispor então de uma visão mais crítica sobre o assunto de modo a introduzi-lo em movimentos sociais e debates políticos e, sobretudo, a questionar os progressos e as aplicações em C&T que não estavam propiciando a satisfação das demandas e o desenvolvimento do bem estar social de toda a população. Neste cenário, iniciou-se no âmbito acadêmico o chamado movimento ou enfoque Ciência, Tecnologia e Sociedade (CTS).

A partir de meados do século XX, nos países capitalistas centrais, foi crescendo o sentimento de que o desenvolvimento científico, tecnológico e econômico não estava conduzindo, linear e automaticamente, ao desenvolvimento do bem-estar social. Após uma euforia inicial com os resultados do avanço científico e tecnológico, nas décadas de 1960 e 1970, a degradação ambiental, bem como a vinculação do desenvolvimento científico e tecnológico à guerra (as bombas atômicas, a guerra do Vietnã com seu napalm desfolhante) fizeram com que a ciência e a tecnologia (C&T) se tornassem alvo de um olhar mais crítico. Além disso, a publicação das obras *A estrutura das revoluções científicas*, pelo físico e historiador da ciência Thomas Kuhn, e *Silent spring*, pela bióloga naturalista Rachel Carson, ambas em 1962, potencializaram as discussões sobre as interações entre ciência, tecnologia e sociedade (CTS). Dessa forma, C&T passaram a ser objeto de debate político. Nesse contexto, emerge o denominado movimento CTS. (AULER; BAZZO, 2001, p. 1, grifo do autor).

O CTS, campo de estudos e de trabalho acadêmico interdisciplinar surtido nas linhas de investigação advindas das áreas de filosofia e de sociologia da ciência, hoje consolidado institucionalmente, propõem-se a “[...] ressaltar a importância social da ciência e da tecnologia, de forma a enfatizar a necessidade de avaliações críticas e análises reflexivas sobre a relação científico-tecnológica e a sociedade.” (PINHEIRO; SILVEIRA; BAZZO, 2007, p. 74).

Como um dos valores ou princípios primordiais do movimento CTS, está a defesa do acesso das sociedades aos conhecimentos e às informações sobre o desenvolvimento científico-tecnológico, a contextualização das atividades científicas e tecnológicas como processos sociais resultantes de fatores culturais, políticos e econômicos e, segundo Auler e Bazzo (2001), a reivindicação por decisões mais democráticas e menos tecnocráticas, com a participação de um maior número de atores sociais nas tomadas de decisões referentes à C&T.

Ao retratar o campo de estudos das inter-relações entre CTS, Cerezo (2002, p. 6) citando Barnes (1987) e Latour (1992) faz uma interessante colocação:

O ponto-chave é a apresentação da ciência-tecnologia não como um processo ou atividade autônoma, que segue uma lógica interna de desenvolvimento em seu funcionamento ótimo, mas como um processo ou produto inerentemente social, em que os elementos não técnicos (por exemplo, valores morais, convicções religiosas, interesses profissionais, pressões econômicas, etc.) desempenham um papel decisivo em sua gênese e consolidação. A complexidade dos problemas abordados e sua flexibilidade interpretativa, a partir de distintos marcos teóricos, fazem necessária a presença desses elementos não técnicos, na forma de valores ou de interesses contextuais.

Existe uma diversidade de programas de colaboração multidisciplinar que constituem os denominados estudos CTS que, de acordo com Cerezo (2002, p. 9), enfatizam a dimensão social da C&T, uma vez que compartilham em comum “(a) a rejeição da imagem da ciência como uma atividade pura; (b) a crítica da concepção da tecnologia como ciência aplicada e neutra; e (c) a condenação da tecnocracia”. Para Santos e Ichikawa (2002, p. 240) as tradições teóricas em CTS podem ser definidas em duas, ou seja:

- A tradição europeia, que nasceu com os “Programas Fortes” de sociologia do conhecimento científico, e que centra seu estudo na análise dos antecedentes ou os condicionantes da ciência; e
- A tradição norte-americana, que centra seus estudos nas consequências sociais e ambientais do conhecimento científico.

A tradição europeia, originada nos “programas fortes” concluídos na década de 70 por autores da Universidade de Edimburgo como David Bloor, Barry Barnes ou Steven Shapin e marcada por uma interpretação radical da obra de Thomas Kuhn, é caracterizada mais pela tradição de investigação acadêmica do que pela tradição educativa ou divulgativa; já a tradição americana, o qual teve seu início associado ao movimento pragmático norte-americano e a obra de ativistas ambientais e sociais como Rachel Carson e Ernst Friedrich (E. F.) Schumacher, é demasiadamente mais ativista, consolidando-se institucionalmente por meio do ensino e da reflexão política (GONZÁLEZ GARCÍA et al., 1996, citado por CEREZO, 2002).

Os estudos e programas CTS estão sendo desenvolvidos desde seu início em três grandes direções: no campo da investigação, promovendo uma nova perspectiva nas áreas de filosofia e de sociologia da ciência, prevalecendo uma nova visão não essencialista e contextualizada da ciência como processo social; no campo das políticas públicas, promovendo processos de tomadas de decisão mais democráticos relativos a políticas técnico-científicas; e no campo da educação, introduzindo as discussões e reflexões CTS sobre a C&T aos programas e materiais de ensino secundário e superior (CEREZO, 2002; SANTOS; ICHIKAWA, 2002).

A partir destas perspectivas, entendemos que a preservação digital encontra aplicações nos objetivos pretendidos pelo movimento ou enfoque CTS, haja vista que a preservação digital procura em sua essência salvaguardar a memória e o acesso, por longos períodos, aos conteúdos informacionais em ambientes digitais. Desta forma, consideram-se suas relações recíprocas,

oportunizando aos indivíduos de hoje e as futuras gerações a preservação e a disposição do acesso utilizável à conteúdos informacionais sobre C&T.

Todavia, prover e assegurar o acesso às informações científico-tecnológicas não é o suficiente na perspectiva CTS pois, ao contextualizar a importância do enfoque CTS para o ensino médio como elemento impulsor de uma compreensão ampla e social do contexto científico-tecnológico, Pinheiro, Silveira e Bazzo (2007) destacam que é cada vez mais necessário que a população apresente também meios que lhe permita avaliar e participar das decisões que venham a interferir no âmbito do qual esteja presente.

No panorama nacional, dados recentes obtidos na enquete de percepção pública da C&T promovido em 2010 pelo Ministério da Ciência, Tecnologia e Inovação e pelo Museu da Vida, contraditam a hipótese de que um maior nível de instrução ou de informação desencadearia a atitudes genericamente mais positivas acerca do papel da C&T na sociedade, pois: I) Em torno de 60% dos brasileiros declararam ter um grande interesse por assuntos de C&T (e também uma postura positiva e otimista), porém, possuem uma carência de conhecimento destes assuntos e acessam pouca informação científica; e II) Ademais, com o aumento da informação, os indivíduos passam a valorizar a potência ligada ao conhecimento científico e tecnológico, entretanto, enfocando também os riscos e perigos (CASTELFRANCHI et al., 2013).

Sob a perspectiva de aderir aos objetivos do movimento CTS no contexto educacional brasileiro frente a nossa inexperiência democrática constatada ao longo da trajetória histórica do país, Auler e Bazzo (2001, p. 12) salientam que “[...] além de conhecimentos/informações, necessários para uma participação mais qualificada da sociedade, necessitamos, também, iniciar a construção de uma cultura de participação”. Isto posto, dois caminhos são essenciais para uma maior participação qualificada da população: a inserção do enfoque CTS na educação em C&T oferecida nas escolas e a educação não-formal proveniente dos museus e centros de ciência.

A introdução do enfoque CTS no âmbito educacional, área de investigação e ação social do movimento CTS, devendo ser legitimado através do sistema educativo, prevê: I) Uma reestruturação curricular dos conteúdos e o aperfeiçoamento dos docentes, rompendo-se assim com a concepção metodológica e atitudinal tradicional de ensino-aprendizagem da produção do saber; e II) Uma maior participação dos alunos, que estes exerçam a criatividade e recebam subsídios para que assumam no futuro, dentro das instituições de ensino e extramuros, uma postura questionadora e crítica-reflexiva acerca do contexto científico-tecnológico e social. Há vários trabalhos sobre o assunto que vêm sendo elaborados em instituições escolares como, por exemplo, os projetos britânicos *Science in a Social Context* (SISCON) e o *Science and Technology in Society* (SATIS). Ademais, duas associações se destacam por suas pesquisas

neste domínio: a britânica *Association for Science Education* e a norte-americana *National Science Teachers Association* (CEREZO, 2002; PINHEIRO; SILVEIRA; BAZZO, 2007).

Os museus e centros de ciência, espaços de divulgação e educação não-formal em C&T, têm colaborado na diminuição das desigualdades provindas dos níveis variados de acesso pelos visitantes à educação científica e técnica em escolas, porém, deve-se ressaltar: I) Para que estabeleçam um elo fidedigno com seu público real e potencial é necessário que estas instituições promovam experiências significativas que propiciem desde o incentivo ao aumento do número de visitantes como a extensão da sua função social; e II) Os programas de comunicação precisam investigar conceitos e técnicas cujo os aspectos sociais e culturais deste conhecimento estejam inseridos, possibilitando aos museus de ciência o papel de intercessores entre a sociedade e os organismos decisórios. Assim, a formação de um público mais culto sobre os fenômenos e conceitos científicos melhorará a sua condição para a discussão, o acompanhamento e a reivindicação de políticas públicas relativas a aspectos atuais e questionáveis da ciência. Dentre as instituições nacionais conhecidas nessa área, podemos citar: o Centro de Divulgação Científica e Cultural (CDCC) da Universidade de São Paulo (USP/ São Carlos - SP) e o Museu da Vida, da Fundação Oswaldo Cruz, no Rio de Janeiro (VALENTE; CAZELLI; ALVES, 2005).

Neste sentido, a preservação digital torna-se uma aliada aos estudos CTS. O acesso aos conteúdos científico-tecnológicos juntamente com a criação, a avaliação e o aprimoramento de políticas científicas abertas à participação da população e de modelos de projetos ou programas práticos de popularização, comunicação e educação em C&T, por exemplo, contribuirá numa aproximação da ciência com o público não especialista ou técnico e na formação de indivíduos menos passivos ou alienados intelectualmente que exerçam concretamente, numa democracia, os seus respectivos direitos e deveres como cidadãos nas sociedades.

Desta maneira, o público em geral e especialista poderá obter um posicionamento mais autônomo, pertinente e crítico-avaliativo bem como uma participação mais ativa, efetiva e de maior qualidade nas opiniões ou tomadas de decisões na área de C&T, proporcionando, além disso, a compreensão, o questionamento e a reflexão crítica acerca dos valores, dos interesses, das causas, das consequências sociais e ambientais e, também, das questões culturais, políticas, econômicas ou éticas inerentes nas interações entre CTS.

3 PRESERVAÇÃO DIGITAL: revisitando conceito, problemas, requisitos, estratégias, modelo de referência SAAI e políticas

A sociedade contemporânea vive um real e constante risco de perda permanente de sua própria história. Em consequência dos avanços em C&T e da rápida obsolescência das TIC, o acesso adequado e a interpretação de todas as partes dos conteúdos de registros de informações ou dados (especialmente, em formato digital e eletrônico) poderá ser comprometido, de forma parcial ou total, ao longo do tempo. Nesta perspectiva, com o expressivo aumento gradativo na produção e aquisição de conteúdos digitais, medidas de recuperação e preservação de dados a longo período estão sendo desenvolvidas e estudadas defronte às mudanças tecnológicas de armazenamento e de acesso aos recursos informacionais.

Contudo, manter a autenticidade e a integridade dos recursos¹, além da sua organização (incluindo indexação, classificação e catalogação), da disposição e da guarda por longo prazo efetivamente, de modo que permita assegurar o acesso contínuo ao seu conteúdo informacional, amenizando assim os possíveis riscos de sua perda definitiva é apontada pela comunidade científica como um dos grandes desafios do século XXI. A complexidade da execução e da gerência do processo de preservação digital advém das características dos recursos/objetos digitais, bem como da carência de infraestrutura, da adoção de formatos de arquivos de dados abertos e de padrões ou esquemas de metadados para a identificação, a administração e a salvaguarda dos objetos.

Apesar de não ser uma preocupação recente, os profissionais da Ciência da Informação (os bibliotecários, os arquivistas ou profissionais das áreas afins) vem realizando esforços para conscientizar as organizações e, em certa medida, os pesquisadores (por exemplo, pela falta de seu acesso aos dispositivos e programas descontinuados ou obsoletos que viabilizem a leitura das informações) de que os formatos digitais e os suportes físicos e digitais não são duradouros. Neste cenário, as iniciativas e pesquisas voltadas à preservação digital tornam-se essenciais para o estabelecimento de políticas de informação em empresas ou instituições e a realização de boas práticas por parte dos criadores/produtores de informação digital.

¹ Um documento digital autêntico constitui aquele que podemos comprovar ser o que significa ser, que tenha sido de fato criado pela entidade que alude tê-lo produzido, e que realmente foi criado no período alegado. Por sua vez, a integridade de um documento digital remete à condição de plenitude e fixidez do documento, onde o seu conteúdo produzido (incluindo as funcionalidades, a aparência, a estrutura e a forma) não suporta modificações imprevistas, maliciosas e/ou inadvertidas. A verificação da autenticidade e da integridade de um documento digital é realizada por meio de diversos mecanismos como, por exemplo, leiaute, tipos de fontes, vocabulários do período, recursos de assinatura digital e códigos de verificação (*checksums*) (BODÊ, 2008; CORRÊA, 2010; INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2001; SWISS FEDERAL ARCHIVES, 2009).

A definição de preservação digital é um elemento primordial para qualquer discussão relativa aos múltiplos aspectos envolvidos neste processo. Para Bullock (1999) as expressões “preservação digital” ou “arquivamento digital” significam a tomada de medidas para garantir a longevidade de documentos eletrônicos, podendo estas serem aplicadas tanto aos documentos que são “nascidos digitais” e armazenados *online* ou em *Compact Disc Read-Only Memory* (CD-ROM), disquetes ou outros suportes físicos, quanto também para os produtos de conversão do formato analógico para o formato digital, se o acesso ao longo do tempo é pretendido.

Tendo o conceito de autenticidade de um objeto digital associado à guarda do conteúdo informacional original de sua criação/produção, Grácio, Fadel e Valentim (2013, p. 113) entendem que a preservação digital é “[...] um processo de gestão organizacional que abrange várias atividades necessárias para garantir que um objeto digital possa ser acessado, recuperado e utilizado no futuro, a partir das TIC existentes na época e com garantias de autenticidade”.

O termo técnico preservação digital de longo prazo refere-se ainda à ação de proteger informação de modo adequado e independente por longo período, informação armazenada independente que dispõe de documentação satisfatória para sua compreensão e adoção, sem recursos específicos obsoletos, por uma comunidade. Exige certos métodos e técnicas propícias para cada formato e mídia, propondo-se assegurar a inalterabilidade dos registros digitais. Os objetos digitais são tipos de arquivos em meio digital, formados de conjuntos de cadeias de *bits*² acerca de conteúdos informacionais, metadados e identificadores (ARELLANO, 2008).

Embora cada definição possua a sua visão particular sobre o processo, percebe-se que a preservação digital é hoje uma importante área de investigação para salvaguardar a memória e o acesso por longo prazo aos conteúdos informacionais em ambientes digitais. Portanto, cabem aos profissionais da informação manterem o ideal e o desenvolvimento de métodos em prol da preservação e que garantam a recuperação efetiva dos registros digitais ao longo do tempo. Através da revisão da literatura acadêmica e especializada, são discutidos as principais questões articuladas nos estudos e que são inerentes neste domínio.

3.1 Problemas da preservação digital

As principais dificuldades da preservação digital provém das características do ambiente digital e da natureza dos próprios objetos que procura preservar. Os objetos digitais se diferem

² Bit (*binary bit*) ou dígito binário remete a mínima unidade de informação arquivada num computador, sendo que um *bit* possui o valor único de 0 ou 1. O *byte* equivale a um conjunto de dígitos binários armazenados e processados como uma unidade (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2015).

dos formatos tradicionais pois, segundo Thomaz (2004), são acessíveis meramente por meio de combinações específicas de componentes de *hardware*, de *software*, de mídia e de pessoal técnico. No mundo digital, ainda que nas mais apropriadas condições de armazenamento, a autora afirma que as mídias digitais podem ter a sua existência suspensa em decorrência da ausência ou inadequação de qualquer um dos outros componentes explicitados.

Há algumas divergências na preservação de meios digitais e não digitais. Os relatórios impressos, por exemplo, quando são mantidos em seu formato original, todos os seus aspectos são preservados em sua forma física, isto é, o formato, o leiaute e o conteúdo. Nesta situação, é impossível dissociar seus elementos individuais (conteúdo sem leiaute, por exemplo) já que estão intrinsecamente associados. Todavia, os objetos digitais podem ser decompostos em seus elementos individuais, sendo preciso um empenho mais árduo para preservá-los “inteiramente”. Como exemplo, pode-se manter o conteúdo de um documento eletrônico, mas perder seu leiaute ou, ainda, manter a presença física (o arquivo de dados, por exemplo) de um objeto, porém deixar de preservar sua capacidade de leitura (BULLOCK, 1999; THOMAZ, 2004).

Agregam-se a estes problemas fundamentais da preservação digital, outras dificuldades de ordem técnica, legal, política, econômica e social, dentre as quais podemos ressaltar:

- ✚ O volume rapidamente crescente de informações em formato digital, somado a necessidade de estabelecer por quem e como esta gigantesca quantidade de materiais deve ser coletada, selecionada e preservada para gerações futuras (LUSENET, 2002).
- ✚ A multiplicação e a variedade de padrões e formatos de arquivo de dados, tornando-se assim preciso identificar e descrever os formatos mais adequados para a preservação da informação digital por longo prazo (LIBRARY AND ARCHIVES CANADA, 2010).
- ✚ A rápida e contínua obsolescência tecnológica ao nível dos formatos de arquivo, do *software*, do *hardware* e dos suportes, bem como os danos físicos ao nível dos arquivos e dos materiais componentes do *hardware* e dos suportes (BARBEDO; CORUJO; SANT’ANA, 2011).
- ✚ Os direitos de propriedade intelectual (direitos autorais, por exemplo), que podem interferir na realização de cópia, armazenamento, modificação e utilização de conteúdos de recursos digitais para fins de preservação (NATIONAL MUSEUM OF AUSTRALIA, 2012).
- ✚ O grau real de custos envolvidos e como deve ser o financiamento, além da insuficiência na promoção e conscientização do público e das instituições sobre a preservação dos dados, as implicações da preservação de longo prazo e as suas demandas (CAMPOS, 2002).

A partir dos problemas expostos anteriormente, nota-se que gerir, preservar e manter o acesso por longo período dos materiais digitais é algo muito mais complexo, quando equiparado aos materiais em suportes convencionais, como o papel. Portanto, torna-se crucial identificar e

analisar quais são os conjuntos de requisitos, indicados na literatura especializada, considerados necessários para se planejar e estabelecer uma adequada preservação digital.

3.2 Requisitos para a preservação digital

Os requisitos representam o conjunto mínimo de objetivos e de restrições determinadas no âmbito das organizações pelas partes envolvidas. Estes requisitos dividem-se em dois tipos, a saber: os requisitos funcionais, que descrevem as funções, ações ou procedimentos a serem desenvolvidos para se alcançar os resultados almejados; e os requisitos não-funcionais, que descrevem as propriedades globais destes resultados, por meio de aspectos fundamentais, como qualidade inerente do produto, custo/preço do produto, entrega/atendimento do produto, moral dos indivíduos, segurança dos indivíduos e impacto ambiental (THOMAZ, 2004).

Bullock (1999) e Thomaz (2004), respaldados nos conceitos do modelo de referência OAIS ou SAAI identificam nove requisitos a serem analisados para a preservação de objetos digitais por longo prazo, que podem ser assim compreendidos:

1. Fixar os limites do objeto a ser preservado – para fins de preservação é preciso determinar, de forma clara, quais os elementos presentes no objeto digital que serão realmente mantidos ou preservados (âmbito da criação de uma política de seleção), visto que tal objeto apresenta vantagens sob a perspectiva da navegação devido a sua natureza multimídia e hipertextual (relação com a preservação das funcionalidades contidas nos objetos).
2. Preservar a presença física – consiste em proteger o(s) arquivo(s) físico(s), ou seja, a camada primitiva de suporte da informação a ser representada; refere-se, então, ao(s) arquivo(s) de computador, às sequências de zeros e uns que são a base para o significado ou interpretação de um objeto digital.
3. Preservar o conteúdo – refere-se a manter a capacidade de acessar o conteúdo do objeto em seu nível mais baixo, como um arquivo texto em formato padrão *American Standard Code for Information Interchange* (ASCII)³, independente da instauração de variações de fontes e características de leiaute (âmbito da apresentação do conteúdo dos objetos).

³ Os conjuntos de caracteres referem-se ao sistema de codificação de uma sequência de caracteres em um *byte* de oito *bits*, ou octeto. Geralmente, conjuntos de caracteres e codificação de caracteres são considerados sinônimos. Assim, a norma *International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 8859-1:1998* define o esquema de codificação ASCII, o qual atribui valores a um conjunto de 256 caracteres (incluindo algarismos, letras, caracteres de controle, dentre outros símbolos) onde cada caractere é determinado utilizando um *byte* de 8 *bits* (LIBRARY AND ARCHIVES CANADA, 2010).

4. Preservar a apresentação – o conteúdo dos objetos digitais é exibido visualmente por meio do emprego de fontes de diferentes formatos e tamanhos, uso de espaço em branco, colunas, margens com ou sem ilustrações, cabeçalhos, rodapés, paginação, marcadores, numeração e assim por diante. Em determinados tipos de documentos digitais, como formatos padrão *Standard Generalized Markup Language (SGML)*⁴, *Extensible Markup Language (XML)*⁵ e alguns formatos *Portable Document Format (PDF)*⁶, as especificações de apresentação se localizam separadas do conteúdo.
5. Preservar a funcionalidade – os objetos digitais podem apresentar componentes multimídia (texto, imagens, gráficos, áudio e vídeos integrados ao seu conteúdo, por exemplo), existir em formato hipertexto (possuem a capacidade de se deslocarem de maneira dinâmica para outros locais do próprio documento ou para outro documento, por exemplo), conter conteúdo dinâmico (produzido automaticamente através de bancos de dados, por exemplo) ou dispor de funções de navegação (barras de ferramentas, tabelas interativas de conteúdos ou pesquisa a palavra-chave, por exemplo).
6. Preservar a autenticidade – para o propósito de preservação é preciso haver confiança que o objeto acessado é precisamente aquele que se busca (relação com a qualidade do serviço ou produto oferecido pelas instituições aos seus clientes), sendo que as possíveis modificações

⁴ SGML é uma linguagem de marcação (*markup language*), determinada na norma ISO 8879:1986, utilizada para a descrição formal da estrutura e dos conteúdos dos documentos. O conjunto de “etiquetas” (*tags*) na SGML, que permitem a marcação de informações, são usadas para identificar, nomear e descrever relações entre dados, para fins de sua gestão e manipulação. Aplicações baseadas em SGML são independentes da plataforma operacional e podem ser adotadas para múltiplas funções. Um documento SGML contém alguns elementos como, por exemplo, a *Declaration* que descreve o ambiente de processamento que é requerido, e a *Document Type Definition (DTD)*, o qual é um delimitado conjunto de etiquetas que constituem um modelo para descrição da estrutura e do conteúdo de um tipo de documento. Dentre os exemplos de DTD SGML relevantes, estão: a *Encoded Archival Description (EAD)*, para material arquivístico; a *HyperText Markup Language (HTML)* e, a XML, uma linguagem flexível e simples, que viabiliza a formatação das etiquetas (extensíveis e estabelecidas pelo usuário), a documentação de suas definições e a determinação do seu esquema (LIBRARY AND ARCHIVES CANADA, 2010).

⁵ XML é uma linguagem, baseada em SGML, para descrição de documentos eletrônicos e esquema sintático para descrição do conteúdo de dados textuais entre aplicações de informática. Trata-se de um formato padrão para criação, publicação, pesquisa, recuperação, armazenamento e transferência de documentos por meio eletrônico. É um padrão aberto, independente das plataformas operacionais e dos fabricantes de *software*, comprehensível por diversas aplicações e autoexplicativo. No âmbito das estratégias de preservação à longo prazo, a linguagem XML pode vir a garantir o encapsulamento dos metadados e das informações requeridas para compreensão dos objetos digitais originais. Pode ser tida como um tipo particular de migração, enriquece informação acerca de estruturas e significado, além de apoiar a interoperabilidade entre recursos de distintas áreas (ARELLANO, 2008).

⁶ A *Association for Suppliers of Printing, Publishing and Converting Technologies (NPES)*, e a *Association for Information and Image Management (AIIM) International* criaram e adaptaram um padrão internacional que define o uso do PDF da companhia *Adobe Systems* para o arquivamento e a preservação por longo prazo de documentos eletrônicos. O formato de arquivo é o *Portable Document Format/Archive (PDF/A)* e tem sido adotado pela norma ISO 19005-1:2005. É um padrão aberto para documentos baseados em texto, preserva o conteúdo e a aparência visual original dos documentos no decorrer do tempo e, além disto, é independente das ferramentas e dos sistemas utilizados para criar, armazenar ou processar os arquivos. O PDF/A será útil quando os aspectos de processamento (interpretação e representação) do conteúdo original são de maior ou de igual importância para a capacidade de redirecionar ou reaproveitar o conteúdo (LIBRARY AND ARCHIVES CANADA, 2010; THOMAZ, 2004).

(aplicação de migrações e conversões de formatos, por exemplo) pelas quais foi submetido, para assegurar a sua proteção e acesso utilizável por longo período, garantiram a preservação de suas propriedades originais (conteúdo, apresentação e funcionalidades, por exemplo).

7. Localizar e rastrear o objeto digital ao longo do tempo – logo após a sua criação, os objetos digitais tornam-se passíveis de serem modificados, copiados ou deslocados. Em qualquer referência ao objeto digital, este deve ser localizado na edição ou versão correta (âmbito da preservação da autenticidade e integridade do objeto).
8. Preservar a proveniência – consiste na identificação da origem do objeto digital (por quem, quando, onde e como foi criado, por exemplo) e o detalhamento do seu histórico (processos, procedimentos ou técnicas as quais foi submetido, por exemplo), por meio dos metadados para preservação digital, como medidas para auxiliar na comprovação ou garantia da sua autenticidade e integridade.
9. Preservar o contexto – os objetos digitais são definidos, caracterizados ou descritos por suas dependências de *hardware* e de *software* (associação com a capacidade de acesso de forma utilizável ao conteúdo informacional e da utilização de metadados para preservação de longo prazo), seus modos de distribuição e relacionamentos com outros objetos digitais (âmbito do histórico, da presença de conteúdo dinâmico e de demais funcionalidades nos objetos).

Em síntese, o conjunto de nove requisitos indicados a serem analisados e refletidos pelas organizações que se comprometem a manter informação digital ao longo do tempo, objetivam, sobretudo, salvaguardar por longo período os objetos digitais e a capacidade de acesso contínuo e utilizável aos seus conteúdos, refletindo assim os próprios pressupostos da preservação digital. Embora cada requisito tenha o seu conjunto particular de características, propósitos e resultados, todos eles estão intrinsecamente ligados, seja de modo implícito ou explícito.

Devido ao caráter genérico, os requisitos de Bullock (1999) e Thomaz (2004) também são utilizados na pesquisa como critérios de referência para conceber a preservação digital. No entanto, vários outros elementos devem ser analisados para a implementação do processo de preservação digital, como a definição das políticas organizacionais, a disposição de recursos, a identificação das demandas dos usuários, a aplicação de estratégias para preservação do formato de apresentação, das funcionalidades existentes e das demais propriedades originais dos objetos digitais no qual sustentam a validade de sua autenticidade e integridade, assim por diante.

Um outro requisito essencial para garantir a preservação digital é a adoção efetiva de padrões ou esquemas de metadados, pois são eles que definirão a descrição, a representação, a consistência e a persistência do recurso/objeto digital no ambiente informacional, assegurando

efetivamente a sua autenticidade, integridade e confiabilidade bem como o seu acesso, busca e recuperação ao longo do tempo, além de determinar a interoperabilidade⁷ entre sistemas.

Para solucionar ou amenizar os atuais problemas e desafios impostos pelo ambiente digital à preservação por longo período e, simultaneamente, usufruir dos avanços em C&T e dos recentes recursos ou ferramentas tecnológicas que disponibilizam, um conjunto variado de estratégias vêm sendo propostas e estudadas para a preservação digital.

3.3 Estratégias para a preservação digital

Bullock (1999) agrupa as estratégias em dois tipos: as estratégias de preservação para tentar solucionar o problema da obsolescência tecnológica, incluindo a migração, a emulação, a impressão e a preservação da tecnologia; e as estratégias para “assumir o controle da situação”, abrangendo a adoção de padrões, o desenvolvimento de guias, a documentação dos recursos e o investimento em infraestrutura adequada e/ou construção de parcerias. Long (2009) divide as estratégias em termos de chances de prover pelo menos soluções parciais, isto é, as metodologias não sustentáveis a longo prazo, como a estratégia de museu tecnológico⁸; e as que são sustentáveis por longo prazo, tal como a migração e a emulação. Pearson e Del Pozo (2009), por sua vez, optam por ordenar as estratégias em: ações de preservação primárias, que alteram diretamente os materiais digitais ou dados a serem preservados, como a migração; e ações de preservação secundárias, o qual mudam o modo em que o material é acessado e como este acesso é preservado ao longo do tempo, abrangendo a emulação e museu tecnológico.

Segundo Arellano (2008) e Thomaz (2004) as diferentes estratégias para a preservação digital podem ser reunidas em dois grupos, ou seja: as estratégias estruturais, que consistem nos investimentos iniciais advindos das instituições, a fim de construir um ambiente adequado para implantação do processo de preservação digital; e as estratégias operacionais, que constituem as ações ou medidas reais de preservação digital a serem desenvolvidas pelas organizações, que visam preservar por longo prazo objetos digitais. Baseando-se nesta categorização genérica, o Quadro 1 enumera as estratégias de preservação mais adotadas e relatadas na literatura.

⁷ Interoperabilidade é “[...] a capacidade de vários sistemas com diferentes plataformas de hardware e software, estruturas de dados e interfaces, de trocar dados com perda mínima de conteúdo e funcionalidade.” (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004, p. 2, tradução nossa).

⁸ Embora a estratégia de museu tecnológico tenha uma designação dissemelhante da estratégia de conservação da tecnologia, consideraremos nesta pesquisa que ambas tratam-se de um mesmo método de preservação. Tanto uma como a outra se propõem basicamente a preservar o *hardware* e o *software* original, associado com a criação ou o acesso aos objetos digitais, com a finalidade de disponibilizá-los para uso. Através de Long (2009) e Pearson e Del Pozo (2009) podemos obter uma lista de prós e contras do método de conservação da tecnologia.

Quadro 1 – Estratégias de preservação digital.

| Estratégias estruturais | Estratégias operacionais |
|--|--|
| <ul style="list-style-type: none"> - Adoção de padrões - Elaboração de manuais ou guias - Metadados para preservação digital - Investimento e montagem de infraestrutura - Formação de redes de relações | <ul style="list-style-type: none"> - Escolha do meio de armazenamento - Migração - Emulação - Impressão em papel ou microfilme - Conservação da tecnologia - Arqueologia digital |

Fonte: Elaborado pelo autor.

A seguir discutiremos cada um dos métodos ou estratégias de preservação supracitadas e que são usualmente tratadas por variados autores na literatura especializada.

3.3.1 Estratégias estruturais

a) Adoção de padrões

A adoção de padrões busca facilitar a execução das outras estratégias de preservação e maximizar a sua eficiência, com base na utilização preferível de padrões e formatos de arquivos de dados abertos, com amplo acesso e assistência técnica, para os quais exista uma crescente tendência de estabilidade e suporte por longo período. O uso de padrões abertos⁹ possibilita o estudo dos documentos e a sua conversão para novos padrões, enquanto que a conversão dos documentos nos formatos livres (gratuitos, com um código-fonte-aberto e com garantias de liberdade para estudo, execução e modificações) permite que estes possam ser acessados mesmo ocorrendo à obsolescência dos equipamentos e programas de informática em que foram desenvolvidos (ARELLANO, 2008; THOMAZ, 2004).

Através dos trabalhos de Bullock (1999) e Thomaz (2004), podemos considerar que a estratégia de adoção de padrões baseia-se numa abordagem em quatro partes:

1. Definir um conjunto limitado de formatos para armazenar os dados e informações¹⁰ – refere-se à seleção de um número reduzido de padrões e de formatos de arquivos de dados frente à grande multiplicidade de formatos existentes, com o propósito de simplificar a sua gestão e

⁹ Padrões cujas especificações estejam disponíveis publicamente para acesso e implementação, contrapondo-se aos padrões proprietários que são de usufruto de um indivíduo, organização ou grupo (THOMAZ, 2004).

¹⁰ A partir das melhores práticas internacionais e das experiências da *Library and Archives Canada* (LAC) sobre a coleta e preservação de conteúdos digitais, o documento *File Format Guidelines for Preservation and Long-term Access*, de 2010, oferece importantes orientações relativas aos formatos de arquivos e padrões para os vários tipos de conteúdo, como: texto, áudio, vídeo digital, imagens fixas, arquivamento Web, dados estruturados armazenados em banco de dados relacionais ou ferramentas de análise estatística e qualitativa, dados geoespaciais, entre outros (LIBRARY AND ARCHIVES CANADA, 2010).

a realização dos outros processos, como o monitoramento da tendência de obsolescência dos atuais padrões e formatos adotados pela instituição.

2. Utilizar padrões atuais para criar objetos digitais – consiste em adotar padrões atuais, de preferência abertos, com amplo acesso (livre uso ou estudo por longo prazo, por exemplo) e assistência técnica (auxílio de especialistas e profissionais com experiência e capacitação para assegurar sua manutenção), além de possuírem crescente tendência de estabilidade (sem perspectiva de obsolescência) e suporte (componentes de *hardware* e *software* que permitem o acesso e o registro dos dados e informações) por longos períodos.
3. Monitorar os padrões conforme se modificam – concernente ao acompanhamento periódico por meio de pesquisas, tanto a estabilidade e suporte dos padrões e formatos de arquivos de dados usados atualmente pela instituição, quanto os novos padrões e formatos desenvolvidos que, uma vez estabelecidos, poderão substituir futuramente os atuais padrões e formatos utilizados pela organização.
4. Migrar para novos padrões uma vez estabelecidos – através dos diagnósticos e dos relatórios obtidos em pesquisas e estudos periódicos quanto à tendência de estabilidade e suporte dos novos padrões e formatos abertos, com amplo acesso e assistência técnica, será possível substituir de maneira consciente e racional os antigos padrões e formatos empregados pela organização para os novos padrões e formatos de arquivos de dados criados e com nível de obsolescência estável.

As instituições que desempenham a atividade de preservação digital podem também ser suscetíveis à normalização. A título de exemplo, o modelo de referência OAIS ou SAAI, do *Consultative Committee for Space Data Systems* (CCSDS)¹¹ e da *International Organization for Standardization* (ISO)¹², é uma importante e promissora proposta para padronização de um sistema de arquivamento orientado à preservação e manutenção do acesso à informação digital

¹¹ O CCSDS foi formado em 1982 pelas principais agências espaciais do mundo para proporcionar um fórum para a discussão de problemas comuns na elaboração e execução de sistemas de dados espaciais. É hoje composto por onze agências membros, vinte e oito agências de observação e mais de cento e quarenta associados industriais. Entre as importantes agências membros e de observação nacionais e internacionais que participam do CCSDS, estão: o Instituto Nacional de Pesquisas Espaciais (INPE), o Departamento de Ciência e Tecnologia Aeroespacial (DCTA) do Instituto de Aeronáutica e Espaço (IAE), a *National Aeronautics and Space Administration* (NASA) e a *Russian Federal Space Agency* (RFSA). Desde a sua fundação, o CCSDS vem ativamente desenvolvendo Recomendações para padrões de sistemas de informação e de dados para promover a interoperabilidade e suporte recíproco entre as agências espaciais cooperantes bem como possibilitar a colaboração (tanto planejada como imprevista) de várias agências voos espaciais e novos recursos para futuras missões. Ademais, a padronização CCSDS reduz a carga de custos de missões de voo espacial ao permitir partilha de custos entre agências e comercialização rentável.

¹² A ISO é uma rede de órgãos de padronização nacionais de 163 países (apenas um membro por país), tendo o Brasil a participação da Associação Brasileira de Normas Técnicas (ABNT), que possui uma Secretaria Central com sede na cidade de Genebra na Suíça. Foi instaurada, oficialmente, em fevereiro de 1947 a fim de facilitar a unificação e coordenação internacional dos padrões da indústria.

a longo prazo, e fornece ainda uma estrutura para descrição e comparação de modelos de dados e arquiteturas de arquivos. O SAAI é usado em várias iniciativas internacionais de preservação digital, sobretudo, advindas da Austrália e da Europa (BULLOCK, 1999; THOMAZ, 2004).

b) Elaboração de manuais ou guias

O conjunto de manuais e guias¹³ criados em várias partes do mundo, através do registro das descobertas e experiências independentes ou colaborativas feitas por órgãos de preservação, arquivos e bibliotecas, permitem que as instituições obtenham conhecimento e orientação sobre os fundamentos básicos para a gestão de conteúdos digitais. Estes manuais oportunizam ainda que as instituições implementem um processo de preservação digital bem sucedido, reduzindo esforços e custos tal como adquirindo maior flexibilidade e segurança nas tomadas de decisões com relação as contingências ou problemas surgidos no decorrer do processo, assegurando ou amenizando assim, a não perda de dados e de informações consideradas relevantes.

De acordo com Bullock (1999) e Thomaz (2004) os manuais ou guias apresentam em comum algumas recomendações, que podem ser assim compreendidas:

- Reconhecer a responsabilidade inicial do criador na preservação de seus documentos – trata-se do reconhecimento do papel e do compromisso dos criadores/produtores de informação digital na preservação por longo prazo. A não adoção, por exemplo, de padrões e de formatos de arquivos de dados abertos na elaboração dos recursos digitais e a sua influência no desenvolvimento dos procedimentos de preservação digital.
- Identificar as responsabilidades da instituição arquivística – relaciona-se ao estabelecimento e à descrição do conjunto de obrigações, funções e ações a serem realizadas pela instituição arquivística para obtenção de êxito no processo de preservação e manutenção do acesso aos recursos digitais por longo período.
- Adotar diretrizes adequadas para seleção dos objetos a serem preservados (que destacam a questão dos padrões quando existir possibilidade de escolha de formatos) – refere-se tanto ao âmbito do requisito “Fixar os limites do objeto a ser preservado” apontado pelos autores como necessário para o processo de preservação digital, quanto à limitação, quando possível,

¹³ Como exemplo destas iniciativas, podemos citar: Recomendações para a produção de planos de preservação digital (DGLAB), Portugal, 2011; *Digital Preservation Policies: Guidance for archives (The National Archives)*, Reino Unido, 2011; *An Approach to the Preservation of Digital Records (National Archives of Australia)*, Austrália, 2002; *File Format Guidelines for Preservation and Long-term Access (LAC)*, Canadá, 2010; *United States Geological Survey (USGS) Guidelines for the Preservation of Digital Scientific Data (USGS Fundamental Science Practices Advisory Committee – Data Preservation Subcommittee)*, Estados Unidos, 2014; e a Carta para a Preservação do Patrimônio Arquivístico Digital (Conselho Nacional de Arquivos – CONARQ), Brasil, 2005.

de formatos de arquivos de dados (de preferência padrões) frente à alta multiplicidade de formatos existentes, a fim de simplificar a sua administração.

- Proteger os itens arquivados de modificação intencional e não intencional – associa-se à preservação dos documentos arquivados quanto às alterações advindas, por exemplo, das discórdias entre os profissionais responsáveis pela execução das atividades de preservação, pela existência de preconceitos, de negligências e de resistências às mudanças, ou por descuidos advindos pela falta ou não de conhecimento destes mesmos profissionais.
- Fornecer descrição de contexto incluindo histórico de criação, transferência e uso, e registros de auditoria – relaciona-se à descrição das tecnologias aplicadas na criação e requeridas para o funcionamento e o acesso ao conteúdo dos objetos digitais, dos procedimentos ou métodos a que foi submetido, além dos responsáveis pela sua criação ou produção.
- Descrever de forma completa os objetos digitais – refere-se à descrição da proveniência, das especificações das tecnologias de *hardware* e/ou *software* utilizadas e do formato de arquivo de dados original e atual, da localização e dos demais itens dos objetos que assegurem a sua identificação, a localização e, principalmente, a sua autenticidade e integridade ao longo do tempo, através dos metadados para a preservação digital.

c) Metadados para preservação digital

Existem alguns fatores básicos para a utilização de metadados para a descrição detalhada dos objetos digitais: simplificar a investigação e a identificação de suas respectivas fontes de informação, administrar seu fluxo no interior dos processos e, representar suas estruturas para viabilizar o acesso. Os programas de preservação podem ainda ter que escolher entre empregar e, eventualmente, adaptar um dos padrões de metadados atualmente adotados, como *Machine Readable Cataloging* (MARC) e DC, ou mesmo definir seu próprio esquema de metadados como uma solução completa ou solução mínima temporária, até o momento em que surja um novo padrão (THOMAZ, 2004; THOMAZ; SOARES, 2004).

A partir de Bullock (1999) e de Thomaz (2004) pode também ser observado que para o desenvolvimento e a execução da estratégia de adoção de metadados para preservação digital, três questões importantes devem ser ressaltadas:

1. Os dados contextuais básicos à gestão a longo prazo das informações eletrônicas. Elementos de metadados úteis na preservação podem incluir sistema operacional e *software* requerido para acessar um documento, identificadores, detalhes físicos de publicações em formatos tangíveis (o CD-ROM ou disquete, por exemplo), informação de gestão de direitos, padrão de codificação (o ASCII, por exemplo) e versão, histórico de migração e seu sucesso, dados

de auxílio a definição de autenticidade, além de versões e datas. Os projetos de conversão poderiam ainda empregar elementos de metadados adicionais, como o dispositivo de captura, resolução, compressão, material de origem e produtor (do documento digital).

2. A definição e a seleção do identificador único e persistente¹⁴ que necessita ser atribuído ao objeto digital. Um identificador único, com ampla abrangência (de preferência universal), segue o objeto em todo o seu ciclo de vida, apoiando no estabelecimento de sua autenticidade e propiciando ao usuário a confiança de que está acessando o recurso desejado ou procurado.
3. As formas de vinculação dos metadados ao conteúdo dos objetos digitais. O armazenamento dos metadados pode ser como parte integrante do objeto que descreve (incluídos em um cabeçalho HTML¹⁵, por exemplo) ou, além disso, como parte de um arquivo de informação separado (um registro MARC, por exemplo). O modelo de referência OAIS ou SAAI sugere também um pacote de informação formado de Informação de Conteúdo e de Informação de Descrição de Preservação, unindo assim os metadados ao objeto em pacotes. Semelhante ao modelo SAAI, um grupo de trabalho da *Society of Motion Picture and Television Engineers* elaborou o *Universal Preservation Format* (UPF), que consiste num mecanismo de arquivo de dados que adota um contêiner (arquivo que contém outros objetos associados inseridos) para incorporação de metadados no interior de objetos de mídias digitais.

O uso de metadados para preservação possui crucial importância na realização de outros métodos de preservação digital. Por exemplo, uma eficiente migração depende da criação de metadados para o registro do histórico de migrações do objeto (e dos próprios metadados) e o registro e guarda de informação do contexto para a compreensão dos futuros usuários sobre o âmbito tecnológico de criação do objeto digital. De forma similar a emulação requere metadados para descrição da tecnologia usada na criação dos recursos, e de técnicas de encapsulamento do

¹⁴ Os identificadores são conjunto de caracteres indicados para identificar de maneira inequívoca um documento. Entende-se por identificador persistente como aquele único identificador ininterruptamente vinculado a um objeto digital que, quando gerenciado, continuamente fornecerá acesso permanente ao objeto independente de alterações de local. Exemplos de identificadores são: o *Uniform Resource Name* (URN) da *Internet Engineering Task Force* (IETF); o *International Standard Book Number* (ISBN) da ISO; o *Handle System* da *Corporation for National Research Initiatives* (CNRI); o *Serial Item and Contribution Identifier* (SICI) da ANSI/NISO; o *Digital Object Identifier* (DOI) da *International DOI Foundation* (IDF); e o *Persistent Uniform Resource Locators* (PURL) da *Online Computer Library Center* (OCLC). Além dos elementos efetivos que apontem para o objeto digital, os metadados podem ser combinados para atuarem como um conjunto de dados de identificação, diferenciando um objeto de outro para fins de validação (ARELLANO, 2008; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004).

¹⁵ HTML é uma linguagem de marcação (*markup language*) simples, derivada do SGML, que se tornou o formato padrão para a produção de documentos destinados à *World Wide Web*. É usada para a criação de documentos de hipertexto que são portáteis de uma plataforma de computador para outra. Cada versão HTML contém um definido conjunto não extensível de etiquetas (*tags*), que permitem a marcação de dados, usadas para especificar a aparência ou os aspectos do documento que está sendo elaborado (LIBRARY AND ARCHIVES CANADA, 2010).

recurso/objeto (juntamente com o sistema operacional e seus aplicativos) e das especificações do programa emulador a ser adotado e o histórico do recurso digital (ARELLANO, 2008).

d) Investimento e montagem de infraestrutura

Para que as instituições efetivem a preservação e o acesso à informação digital por longo período é preciso uma infraestrutura de *hardware*, *software* e pessoas, isto é, um sistema de arquivamento digital. O ambiente operacional deste tipo possui mecanismos indispensáveis, como: sistema de cópia de segurança; sistema de armazenamento redundante; sistema de detecção e recuperação automática de falhas; sistema de segurança de acesso físico e lógico; e sistema hierárquico de armazenamento. Ademais, deve-se criar um programa de capacitação e de aperfeiçoamento, direcionado a distintas funções realizadas e com variados níveis de especialização, para o pessoal que executará e acessará o sistema (THOMAZ, 2004).

e) Formação de redes de relações

A concretização de um espaço dirigido para a preservação digital requer a formação de corporações, federações e consórcios¹⁶ de natureza informal (associações, alianças e parcerias de cooperação) ou formal (acordos oficiais entre contratantes e contratados), ou seja, uma rede distribuída de relações. Esta medida impõe significativas mudanças e adaptações às estruturas das instituições arquivísticas. As formações poderiam se estruturar em limites a nível regional ou nacional, e além disso, especializarem-se no arquivamento de certos tipos de informação digital, por intermédio de interesses particulares na disseminação de informação em fronteiras nacionais ou da idoneidade das partes envolvidas referente a alguma disciplina, tipos de informação e função arquivística (BULLOCK, 1999; THOMAZ, 2004).

3.3.2 Estratégias operacionais

a) Escolha do meio de armazenamento

A implantação bem sucedida dos processos de preservação digital depende da escolha apropriada do meio de armazenamento da informação. Ao retratar os múltiplos fatores que

¹⁶ Podem-se indicar, como exemplo, *The European Library* (TEL), na Europa; *Research Libraries United Kingdom* (RLUK) e *Digital Preservation Coalition* (DPC), no Reino Unido; *National Digital Stewardship Alliance* (NDSA), *National Digital Information Infrastructure and Preservation Program* (NDIIPP), *Digital Preservation Outreach and Education* (DPOE), *Digital Library Federation* (DLF) e *Council on Library and Information Resources* (CLIR), nos Estados Unidos; *Preserving and Accessing Networked Documentary Resources of Australia* (PANDORA), na Austrália; *Canadiana.org* e *Canadian Digital Preservation Initiative* (CDPI), no Canadá; e *International Internet Preservation Consortium* (IIPC), com abrangência mundial.

influenciam nesta escolha, Thomaz e Soares (2004) afirmam que as condições necessárias para o acesso às informações definirão, sobretudo, a espécie de disponibilidade (*online*, *near-line* e *off-line*) e a velocidade do dispositivo de leitura. No cenário governamental do Reino Unido, através de Brown (2008) e Thomaz (2004), as diretrizes do *The National Archives* recomendam seis critérios técnicos para seleção da mídia de armazenamento físico e removível:

1. Longevidade – a mídia deve possuir durabilidade mínima comprovada de dez anos. Não há vantagem para uma maior durabilidade, pois a obsolescência da tecnologia do dispositivo de leitura/gravação, usualmente, será algo mais expressivo que a degradação física da mídia.
2. Capacidade – a mídia deve fornecer uma capacidade conveniente ao volume de dados a serem armazenados e ao tamanho físico das instalações presentes. A diminuição do número de mídias a serem geridas, normalmente, propiciará eficiência e economia de recursos.
3. Viabilidade – a mídia e o dispositivo de leitura/gravação devem possuir métodos robustos para identificação de erros tanto para a leitura como para a gravação dos dados. É preferível a função de teste de integridade de mídia depois da gravação e a existência de técnicas comprovadas de recuperação de dados em situações de perda de dados. A mídia deve ser apenas-uma-gravação (*write-once*), ou ter um mecanismo de proteção de gravação confiável para prever contra exclusões acidentais e assegurar a integridade requerida dos dados.
4. Obsolescência – a mídia e o seu *hardware* e *software* de apoio devem, de preferência, ser baseados em tecnologia madura, ao contrário de inovadora, e devem estar bem estabelecidos no mercado e abundantemente disponíveis. As tecnologias de mídia devem, em especial, ser baseadas em padrões abertos, tanto para a mídia como para o dispositivo de leitura/gravação, ao invés de proprietárias de um único fabricante.
5. Custo – para avaliar os custos relativos as mídias de armazenamento, dois elementos devem ser considerados, ou seja, o custo da mídia e o custo de propriedade. O valor por *Megabyte* ou *Gigabyte* deve ser a base para comparações válidas de custos de mídia. O custo total de propriedade englobará custos de aquisição e manutenção do *hardware* e do *software* precisos e de qualquer equipamento de armazenamento requerido. Outros elementos a serem tratados são os custos de suporte e o período médio entre falhas do dispositivo de leitura/gravação.
6. Susceptibilidade – a mídia deve ter baixa susceptibilidade a danos físicos e ser tolerante a extensa variedade de condições ambientais sem perda de dados. A mídia magnética deve possuir elevado valor de *coercivity* (preferencialmente até 1.000 *Oersteds*), com o intuito de amenizar as chances de eliminação acidental de dados. *Coercivity* é a intensidade do campo magnético requerida para reduzir a zero a magnetização de materiais saturados de forma magnética, isto é, apagar as mídias; e *oersted* é a unidade de medida para a intensidade do

campo magnético. Qualquer medida necessária contra susceptibilidades, como requisitos de embalagem ou armazenamento, devem ser de natureza acessível e exequível.

b) Migração

A migração é uma das estratégias mais usadas pelas organizações. Consiste em transferir periodicamente dados e informações digitais, presentes numa certa configuração de *hardware* e *software* para outra, ou de uma geração de tecnologia para as gerações futuras. Este método protege a presença física e o conteúdo dos objetos, contudo, pode não preservar a apresentação, a funcionalidade e o contexto. Um planejamento inadequado associado a migrações sucessivas, pode refletir em perda inaceitável de dados e de elementos da apresentação, da funcionalidade e do contexto envolvidos nas ligações entre as entradas de banco de dados, caso essas ligações parem de funcionar. Os arquivos possuem um histórico de práticas eficientes para informações simples e homogêneas, porém, quanto aos objetos complexos¹⁷ ainda há a necessidade de testes com resultados satisfatórios (ARELLANO, 2008; BULLOCK, 1999; THOMAZ, 2004).

Para Bullock (1999) e Thomaz (2004) a migração objetiva limitar as perdas e preservar de forma utilizável o conteúdo dos objetos digitais. Esta estratégia baseia-se de ações aplicadas isoladamente ou em conjunto, não necessariamente de modo sequencial, com o propósito de:

- Copiar dados e informações digitais (os arquivos de computador, por exemplo) armazenados num determinado suporte ou meio de armazenamento com tendência de se tornar obsoleto ou fisicamente deteriorado para um suporte mais novo. A cópia, por exemplo, dos dados e das informações registradas em um disco flexível para um CD-ROM.
- Converter um formato de arquivo de dados visto como ultrapassado, em estado de desuso ou obsolescência para um mais atual, independente e de preferência padrão (relação com a estratégia estrutural de adoção de padrões). A conversão, por exemplo, do formato proprietário de arquivo de texto *Microsoft Word* para o formato padrão e aberto ASCII puro.
- Transferir documentos ou arquivos de computador de uma plataforma de *hardware* e de *software* (o contexto tecnológico) em processo de descontinuidade, ou seja, sem perspectiva de futuras versões compatíveis possuindo novas aplicações, para outra com estabilidade e suporte por longo prazo. A transferência, por exemplo, de arquivos criados ou presentes no sistema operacional *Windows NT* para um sistema atual, como *Windows* ou *Linux*.

¹⁷ Como exemplo, podemos salientar o *software* proprietário que, de acordo com Thomaz (2004), caracteriza-se pela proteção de direitos de propriedade ao seu desenvolvedor e pelo código-fonte-fechado ou indisponível, ou seja, a estrutura interna de instruções de programa a serem traduzidas ou interpretadas para ser possível o seu processamento por computador é mantida em sigilo pelo fabricante, inviabilizando mudanças na sua aplicação segundo as necessidades do adquirente.

c) Emulação

A emulação propõe à criação de novo *software* que imite o funcionamento do antigo *hardware* e/ou *software* no intuito de reproduzir o seu comportamento. Esta medida não apenas mantém a presença física e o conteúdo mas, também, a apresentação original e a funcionalidade disponível no *software* anterior. O seu uso é útil para materiais muito dependentes de específico *hardware* e *software*, que não se submeteriam à migração ou, ainda, quando houver a relevância da aparência do recurso original, onde não seja oportuno investir em tecnologias da informação de elevado custo. Todavia, a emulação para fins de preservação de objetos digitais a longo prazo precisa de mais testes ou avaliações sobre o custo/preço final envolvido, tendo sido um método adotado especialmente para prover compatibilidade retrospectiva de jogos eletrônicos e para modelar futuros sistemas¹⁸ (ARELLANO, 2008; BULLOCK, 1999; THOMAZ, 2004).

d) Impressão em papel ou microfilme

A geração de cópia impressa do documento digital representa uma alternativa barata e simples para preservá-lo de modo bem padronizado por longos períodos de anos. Este método permitiria fixar todo o objeto digital bem como preservar o conteúdo e leiaute, porém, a dinamicidade e heterogeneidade dos objetos (os componentes multimídia e o formato hipertexto, por exemplo) dificultam a sua prática. Outro procedimento interessante baseia-se na criação de cópias digitais e do uso de cópias em microfilme como substitutos arquivísticos para beneficiar o acesso, a funcionalidade e, também, a preservação de recursos produzidos originalmente em papel (BULLOCK, 1999; THOMAZ, 2004).

e) Conservação da tecnologia

A conservação da tecnologia propõem-se a salvaguardar a tecnologia de *hardware* e de *software*, envolvida na elaboração dos objetos digitais, a fim de disponibilizá-la para uso. Este procedimento preserva o conteúdo e assegura a visualização dos objetos em seu formato nativo com o leiaute e a funcionalidade original. Entretanto, a criação de “museus” de *hardware* ou

¹⁸ A compatibilidade retrospectiva ou retroativa (*backward compatibility*) pode estar relacionada com a execução de jogos eletrônicos antigos de diferentes plataformas de *hardware* e/ou *software* em computadores modernos, enquanto que a modelagem de futuros sistemas refere-se à construção de modelos de sistemas para visualização e identificação dos requisitos e das funções que o *software* deverá apresentar. Neste cenário, torna-se pertinente mencionar o *Internet Archive* nos Estados Unidos, uma biblioteca digital sem fins lucrativos que arquiva, preserva e oferece acesso público e gratuito a vários tipos de materiais digitais (textos, vídeos, áudios, imagens, *softwares*, páginas *Web* etc.). Como exemplo, esta organização proporciona gratuitamente a todos, através de uma versão ainda em desenvolvimento do emulador livre *DOSBox*, o funcionamento de inúmeros títulos de jogos clássicos do obsoleto *MicroSoft Disk Operating System* (MS-DOS) em nossos *softwares* de navegação (*web browsers*).

software requere condições de custo, de espaço e de suporte técnico de difícil prática à longo período, tornando-se uma medida provisória e em estado de declínio. A sua adoção é válida para objetos digitais relevantes desenvolvidos em formatos proprietários, como o PDF, e em *softwares* obsoletos (ARELLANO, 2008; BULLOCK, 1999; THOMAZ, 2004).

f) Arqueologia digital

A arqueologia digital trata-se de uma medida parcial de preservação. Consiste no uso de materiais digitais inacessíveis, seja pela obsolescência tecnológica e/ou pela deterioração física da mídia de armazenamento, para fins de investigação da memória, por exemplo, dos princípios das ferramentas e sistemas de gerenciamento eletrônico de documentos. Este método é indicado somente para situações em que a importância das informações legitime os elevados custos do procedimento, visto que não existem garantias de se recuperar e restaurar a plenitude dos dados que, dessa maneira, comprometem a definição da identidade, da autenticidade, da integridade e do contexto dos materiais (ARELLANO, 2008; BAGGIO, FLORES, 2013).

Na presente pesquisa, embora cada estratégia tenha diretrizes, intuitos e efeitos distintos, a maioria busca em sua essência salvaguardar a memória e o acesso a longo prazo aos conteúdos informacionais em ambientes digitais. Não há como afirmar, hoje, qual o método que deve ser aderido como padrão único por todas as instituições, posto que o volume crescente de diferentes métodos propostos não são plenamente satisfatórios e destinam-se a certos tipos de formatos de informação digital, de circunstâncias e de âmbitos organizacionais.

Tal especificidade permite, por sua vez, desde a disposição de várias opções aos desafios da preservação digital como a aplicação conjunta de múltiplas ações dependendo do contexto, dos objetivos e do planejamento institucional. Neste cenário, tornam-se primordiais o contínuo monitoramento tecnológico e o uso aliado de ferramentas robustas geradas pela C&T.

3.4 Modelo de referência SAAI

O modelo de referência¹⁹ OAIS ou SAAI, do CCSDS e da ISO, é uma relevante iniciativa com favorável recepção internacional acerca da problemática da preservação digital²⁰. O SAAI pode ser entendido, com base em Thomaz (2004) e Thomaz e Soares (2004), como um esquema conceitual, de alto nível (compreensão e determinação dos elementos de maneira simplificada e genérica impelindo maior aprofundamento e complexidade), que regulariza e coordena um sistema de arquivamento direcionado para a preservação e manutenção do acesso à informação digital por longo período, propondo-se basicamente a:

- a. Expandir tanto a concordância dos componentes e processos relativos à preservação e acesso à informação digital, como a consciência e o entendimento dos conceitos significantes para preservação de objetos digitais, sobremaneira entre as comunidades não arquivísticas;
- b. Estabelecer terminologias e conceitos comuns para descrição e comparação de modelos de dados e arquiteturas de arquivos; e
- c. Construir uma estrutura para orientar a identificação e a elaboração de padrões de suporte ao ambiente de arquivos.

O surgimento do SAAI ocorreu por meio de debate aberto com a cooperação de várias comunidades, tal como o aval de respeitáveis instituições de amplitude internacional com ampla experiência satisfatória na criação de padrões e recomendações (CCSDS e ISO), envolvidas ou interessadas com as volumosas pesquisas na área de preservação digital nos anos noventa. Estes esforços culminaram na publicação da primeira e da segunda versão do *Reference Model for an OAIS - Red Book* em 1999 e 2001, da primeira versão do *Blue Book* em 2002 e, posteriormente, da aprovação da norma ISO 14721:2003 *Space data and information transfer systems - OAIS - Reference model* (THOMAZ, 2004; THOMAZ; SOARES, 2004).

Nesta parte do trabalho as discussões basearam-se na revisão de estudos com conteúdos úteis sobre o modelo OAIS e na consulta da norma nacional em vigor ABNT NBR 15472:2007 Sistemas espaciais de dados e informações - Modelo de referência para um SAAI, que traduz e

¹⁹ Compreende-se modelo de referência (*reference model*) como “esquema usado para entender os relacionamentos típicos entre entidades de determinado ambiente e para desenvolver padrões ou especificações consistentes com esse ambiente.” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2007, p. 4).

²⁰ Iniciativas em prol da preservação de informações digitais por longo período, advindas de diversos países, tem adotado o modelo SAAI como estrutura fundamental para a preservação digital ou para o aprimoramento do entrosamento e dos serviços e processos aplicados entre comunidades variadas (bibliotecas, arquivos, repositórios digitais ou institucionais, dentre outras). É ainda possível obter informações das iniciativas e implementações do modelo por intermédio do projeto já em situação encerrada “OAIS Activities” da *Research Libraries Group* (RLG), mantida e disponibilizada somente para propósitos históricos em uma página Web pela OCLC. Disponível em: <<http://www.oclc.org/research/activities/oaisactivities.html>>. Acesso em: 28 fev. 2015.

adapta as especificações do modelo OAIS descritas na norma internacional ISO 14721:2003 e na Recomendação CCSDS 650.0-B-1:2002 *Reference Model for an OAIS - Blue Book*.

3.4.1 Esquema SAAI

O SAAI representa, para o propósito do modelo de referência, um tipo particular de arquivo que trata-se de uma organização de pessoas e sistemas, que aderiu o papel de manter por longo período informação acessível a um grupo de usuários retratados como Comunidade Alvo. O adjetivo “aberto” se remete à questão do modelo SAAI e as futuras diretrizes serem elaboradas em fóruns abertos, não tendo ligação quanto ao nível de obtenção à informação arquivada. Um arquivo tipo SAAI tem alguns deveres a cumprir que o diverge do emprego puro da palavra “arquivo”, dentre os quais estão (THOMAZ, 2004; THOMAZ; SOARES, 2004):

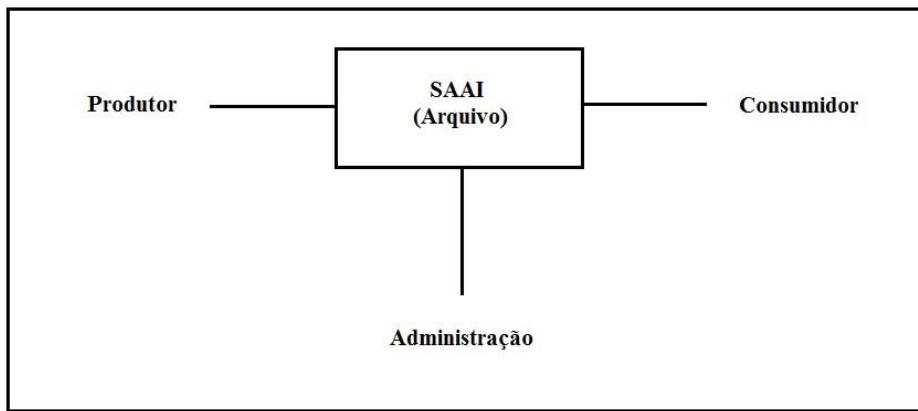
- ✚ Mediar e admitir informação apropriada de criadores/produtores de informação.
- ✚ Reter o correto controle da informação a fim de assegurar a sua preservação por longo prazo.
- ✚ A partir de si próprio ou associado aos demais parceiros, estabelecer quais as comunidades potenciais que devem integrar a Comunidade Alvo (ou seja, uma subclasse da entidade Consumidor no ambiente SAAI) e, assim sendo, devem dotar da capacidade de compreensão da informação oferecida.
- ✚ Assegurar informação comprehensível para a Comunidade Alvo desprovida da assistência dos criadores/produtores de informação.
- ✚ Adotar políticas e métodos documentados que assegurem a preservação da informação frente às contingências oportunas e, ainda, que permitam a disseminação desta mesma informação como cópias fidedignas do original ou que seja rastreada até o original.
- ✚ Permitir que a informação preservada esteja acessível para a Comunidade Alvo.

Para Thomaz (2004) e Thomaz e Soares (2004) o modelo SAAI especifica a arquitetura conceitual de um arquivo, com relação a seu ambiente, suas informações e suas funções requeridas para auxiliar ou apoiar as responsabilidades explicitadas anteriormente.

3.4.2 Ambiente SAAI

Devemos compreender a princípio o cenário de um arquivo tipo SAAI, isto é, as relações deste arquivo com o ambiente em que está presente. A Figura 1 reproduz este ambiente.

FIGURA 1 – Ambiente SAAI.



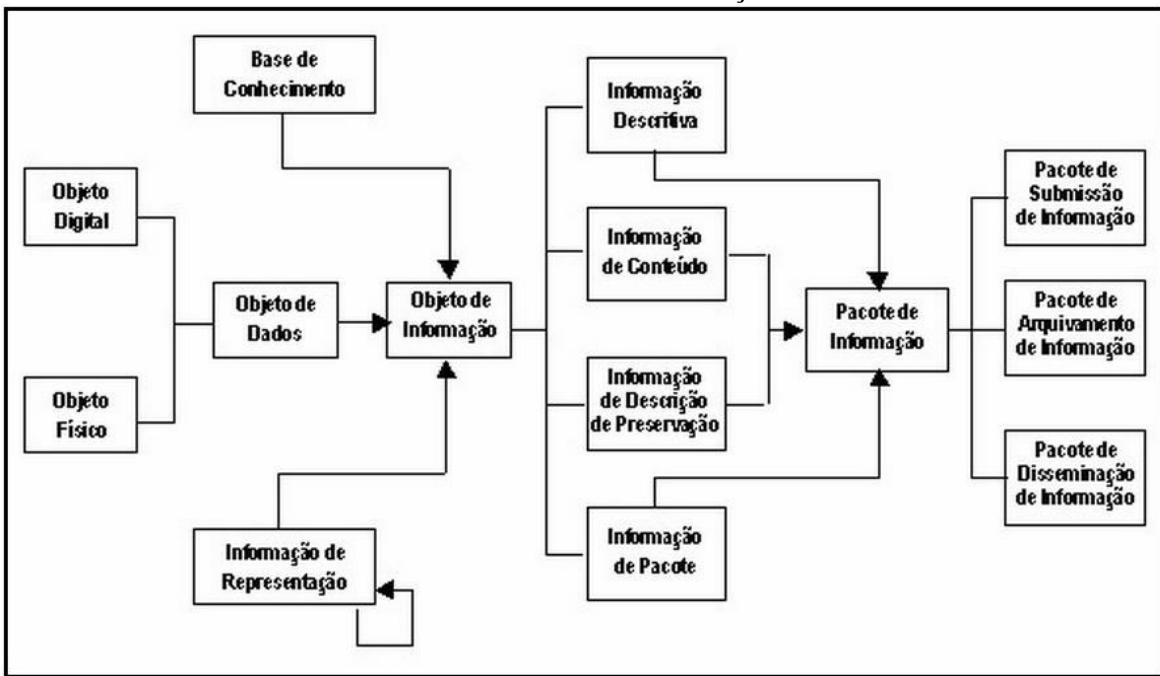
Fonte: Thomaz e Soares (2004).

Há quatro entidades que compõem o ambiente SAAI. Os Produtores são os indivíduos ou sistemas clientes que concedem a informação a ser preservada. Os Consumidores são os indivíduos ou sistemas clientes que possuem interesse em utilizar a informação mantida, sendo que a Comunidade Alvo é uma subclasse de Consumidores potenciais, a serem reconhecidos para a preservação de longo prazo e que requerem entendimento desta informação. Além disto, a Administração encarrega-se de definir políticas genéricas do arquivo, pois, não se relaciona com as práticas comuns da entidade Arquivo, sendo esta ação efetuada por uma função inerente ao próprio arquivo (ARELLANO, 2008; THOMAZ, 2004; THOMAZ; SOARES, 2004).

3.4.3 Modelo de informação SAAI

O SAAI define um modelo de informação para incorporação dos metadados requeridos na preservação e acesso de informação digital a longo prazo, conforme ilustrado na Figura 2.

FIGURA 2 – Modelo de informação SAAI.



Fonte: Thomaz e Soares (2004).

Baseando-se em Arellano (2008), Campos e Saramago (2007), Saramago (2004), Sayão (2010), Thomaz (2004) e Thomaz e Soares (2004) identificam-se um grupo de aspectos que são notáveis para a conceituação do modelo de informação SAAI, entre os quais se destacam:

- ✚ A criação, a manutenção e o aprimoramento de metadados especificados do arquivo é um fator crucial para a gerência satisfatória da preservação digital. Os metadados de preservação possibilitam, por exemplo, documentar os processos técnicos acerca da preservação, detalhar direitos de acesso, definir a autenticidade do conteúdo digital, registrar histórico de custódia do objeto digital, identificar o objeto de maneira única, dentro ou externamente, quanto ao seu arquivo, além de auxiliar a execução das estratégias de preservação digital.
- ✚ Os Objetos de Dados tratam-se da menção, de maneira coletiva, das duas formas existentes de informação (ou seja, qualquer conhecimento suscetível à transferência via dados), a saber: a) Objetos Físicos, com propriedades fisicamente observáveis (por exemplo, documentos em papel e amostras de solo); e b) Objetos Digitais, compostos por sequências, fluxos ou cadeias de *bits* (por exemplo, arquivos em formato PDF ou *Tagged Image File Format*²¹). Podem ser tidos como “informação significativa” pela Comunidade Alvo do arquivo, resultantes por

²¹ O *Tagged Image File Format* (TIFF) é um formato de arquivo indicado para imagens fixas (*still images*). Foi desenvolvido pela *Aldus* (atualmente parte da companhia *Adobe Systems*) em 1986. Os arquivos TIFF podem ou não ser compactados, além disso, são extensíveis e portáteis. Não são adeptos de um específico sistema operacional de computador, compilador ou processador. É considerado um formato de arquivo padrão para a descrição e o armazenamento de informações de imagens de *scanners*, *fax* e aplicações de fotografia digital (LIBRARY AND ARCHIVES CANADA, 2010).

intermédio da conciliação da Base de Conhecimento da Comunidade Alvo e a Informação de Representação (IR) agregada ao Objeto de Dados.

- ⊕ A Base de Conhecimento é um conjunto de informações incorporadas por cada indivíduo ou classe de indivíduos (Comunidade Alvo) para o entendimento da informação recebida. Por exemplo, uma Base de Conhecimento torna-se necessária para interpretar a informação em formato de código-fonte na linguagem de programação Java para os casos de Comunidades Alvo constituídas meramente de programadores Java.
- ⊕ A IR auxilia a Base de Conhecimento de modo a permitir que Comunidade Alvo comprehenda plenamente a informação preservada. Assim, visto que a Comunidade Alvo envolva todos os gêneros de programadores, é pertinente a informação sobre a construção e as combinações de programação Java para o entendimento integral, por parte desta classe de Consumidores, do Objeto de Dados arquivado (informação em formato de código-fonte Java). Além disto, para a sua própria compreensão a IR pode requerer mais IR. Por exemplo, a IR manifestada em ASCII precisa de IR suplementar para ASCII (um documento físico de explicação do formato padrão ASCII etc.). Logo, uma Rede de Representação decorre de objetos de IR.
- ⊕ Tido como um instrumento de reconstituição do sentido da informação, a IR pode ser de dois tipos: I) Informação estrutural, interpreta a sequência ou fluxo de *bits* dos Objetos de Dados incluindo especificação do formato dos dados e, ainda, descrição do ambiente de *hardware* e de *software* onde os dados foram criados, que são requeridos para o acesso vindouro; e II) Informação semântica, adiciona significado à estrutura de dados identificada por meio da Informação estrutural (por exemplo, na identificação de que a cadeia de *bits* é um texto em ASCII, a Informação semântica designa que o texto está escrito em língua inglesa).
- ⊕ No SAAI são observados três pacotes de informação, o qual são reservatórios ou contêineres que encapsulam a Informação de Conteúdo (recurso/conteúdo a ser arquivado e preservado) e sua Informação de Descrição de Preservação (múltiplos metadados de apoio a gerência da preservação de longo prazo), e que também propiciam os fluxos de informação, a saber:
 - Pacote de Submissão de Informação (PSI) – pacote expedido ou entregue do Produtor para o Arquivo. Baseia-se no acordo de submissão no esquema conceitual SAAI.
 - Pacote de Arquivamento de Informação (PAI) – pacote armazenado, administrado e protegido de modo efetivo no interior do Arquivo. Inclui o material que será mantido e a informação necessária para sua representação e preservação de longo prazo.
 - Pacote de Disseminação de Informação (PDI) – pacote transportado do Arquivo para um Consumidor em resposta a uma solicitação ou pesquisa de acesso. Inclui partes de um PAI a serem distribuídas ao respectivo cliente/usuário.

 Os Objetos de Informação constituem uma “informação significativa” para a Comunidade Alvo (ou seja, uma classe particular de Consumidores estabelecida pelo próprio arquivo), resultantes da conciliação entre o Objeto de Dados, a Base de Conhecimento da Comunidade Alvo e a IR. Podem ser de quatro tipos:

- 1) Informação de Conteúdo – informação primordial alusiva ao conteúdo que o sistema de repositório deve preservar junto com a IR (isto é, informações requeridas para a exibição e compreensão da cadeia de *bits* componente do Objeto de Dados arquivado como informação relevante para uma específica Comunidade Alvo).
- 2) Informação de Descrição de Preservação – informação para gestão da preservação da Informação de Conteúdo com que está associada, dividindo-se em quatro subtipos:
 - Informação de referência – elenca e descreve mecanismos usados para indicar identificadores à Informação de Conteúdo, propiciando que seja identificada e localizada durante o tempo de forma inequívoca, interna e externamente ao repositório. Por exemplo, identificadores locais (um número de chamada) ou persistentes, além da inclusão de resumos para descrição do conteúdo.
 - Informação de contexto – registra os relacionamentos entre a Informação de Conteúdo e seu ambiente, incluindo informações sobre motivações para a sua criação, relações com outras informações de conteúdo, dependências técnicas (*hardware, software* etc.), e forma de distribuição (através de rede etc.).
 - Informação de proveniência – registra a história da Informação de Conteúdo, incluindo informações sobre origem, cadeia de custódia (os direitos autorais), além de medidas ou ações de preservação realizadas e seus resultados.
 - Informação de fixidade – registra mecanismos de autenticação adotados para validar que o conteúdo da informação não teve alterações não registradas com danos a sua integridade. Por exemplo, assinaturas digitais e *checksums*.
- 3) Informação de Pacote ou de Empacotamento – congrega a Informação de Conteúdo (recurso/conteúdo mantido) e a Informação de Descrição de Preservação (metadados associados) em uma unidade ou pacote de informação identificável.
- 4) Informação Descritiva (ID) – apoia e simplifica o acesso à Informação de Pacote ou de Empacotamento com base em ferramentas de busca e recuperação.

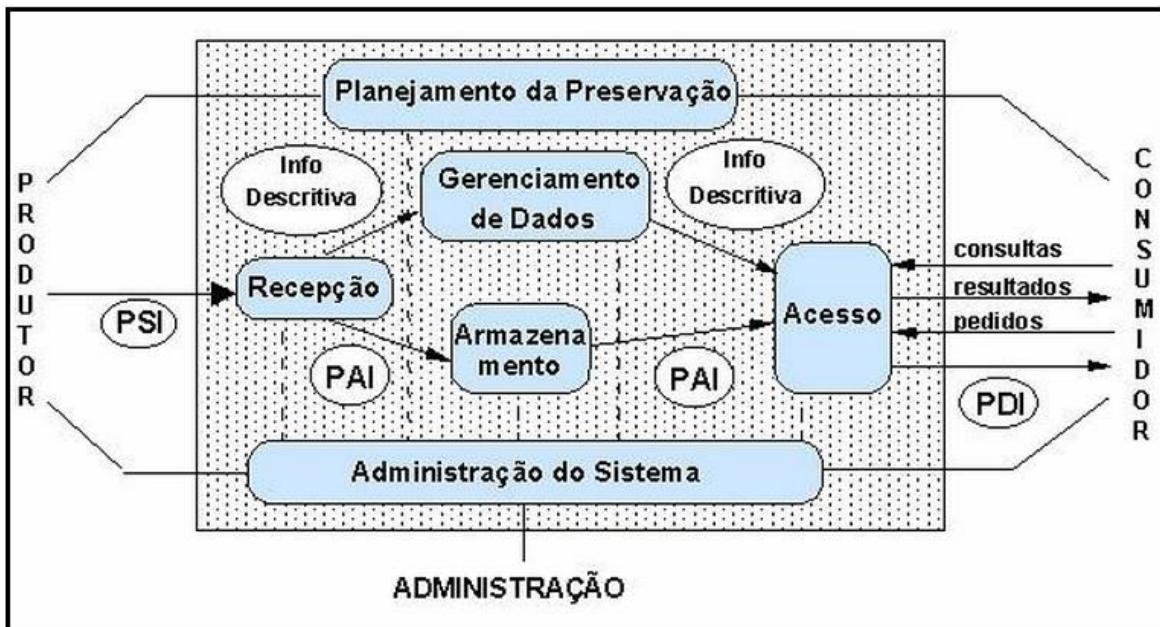
O modelo de informação SAAI, especialmente, o PAI e os seus Objetos de Informação, nos proporciona um viés de quais são as informações essenciais requeridas para o arquivamento digital de longo prazo. Em razão da sua pertinência, o modelo é usado na pesquisa para instaurar um delineamento inicial das informações abrangidas pelos metadados para preservação e acesso

utilizável a longo período de materiais digitais arquivados em repositórios pautados no SAAI, auxiliar na conceituação dos requisitos necessários para a preservação digital explicitados em Bullock (1999) e Thomaz (2004), e também servir de referência na constatação da garantia da preservação digital pelos padrões ou esquemas de metadados identificados na literatura.

3.4.4 Modelo funcional SAAI

O SAAI contém um modelo funcional para definição do conjunto de operações a serem executadas por um repositório dirigido para preservação, segundo exposto na Figura 3.

FIGURA 3 – Modelo funcional SAAI.



Fonte: Thomaz e Soares (2004).

Através de Arellano (2008), Thomaz (2004) e Thomaz e Soares (2004) constatam-se, num primeiro grau de detalhamento, seis entidades funcionais SAAI que administram o fluxo de informação entre as entidades que constituem o ambiente SAAI, e que também reconhecem os componentes funcionais e processos básicos existentes geralmente em arquivos direcionados para a preservação e o acesso de informação digital, os quais são descritas a seguir:

1. Recepção – aceita os PSI dos Produtores ou, ainda, de elementos internos sob o domínio da Administração do Sistema, e manipula os conteúdos para armazenamento e administração no interior do arquivo. Particularmente, obtém e confere a qualidade dos PSI, produz PAI baseando-se na formatação dos dados, padrões e políticas de documentação do arquivo, e cria ID dos PAI (os ícones/símbolos para navegação e metadados para busca e recuperação,

por exemplo). Deslocam, por fim, os PAI elaborados recentemente e as ID relacionadas tanto para a entidade Armazenamento quanto para a entidade Gerenciamento de Dados.

2. Armazenamento – associa-se com o armazenamento, a manutenção e a recuperação de PAI. Tais deveres compreendem adquirir recentes PAI da entidade Recepção e situá-los no âmbito de armazenamento permanente segundo critérios (os requisitos de suporte e as taxas de adoção presumida, por exemplo), gerenciar a ordem da área de armazenamento, reestruturar as mídias, efetuar hábitos de constatação de discrepâncias, prover capacidade de recuperação de imperfeições e oferecer cópias de PAI requeridos para a entidade Acesso.
3. Gerenciamento de Dados – preserva e acessa os dados administrativos adotados para gerir o arquivo assim como a ID que identifica e registra os seus acervos. Particularmente, gerencia o banco de dados do arquivo (protege os esquemas e as definições de visões e integridade referencial), proporciona suas atualizações (porta nova ID ou dados administrativos do arquivo) e examina os dados da entidade para a criação de relatórios.
4. Administração do Sistema – administra a rotina operacional de todo o arquivo. Pode requerer e mediar acordos de submissão com Produtores, averiguar submissões a fim de assegurar que obedeçam aos padrões do arquivo e gerir a configuração do *hardware* e *software* do sistema, analisar e aperfeiçoar a execução geral das operações do arquivo e atualização dos seus conteúdos, além de efetuar tal como preservar padrões e políticas do arquivo, prover assistência ao cliente/usuário e também responder as requisições incompletas.
5. Planejamento de Preservação – controla o ambiente SAAI e oferece diretrizes para assegurar o acesso contínuo por longo período da informação armazenada para a Comunidade Usuária Alvo, ainda que ocorra a obsolescência do âmbito computacional original. Dentre as suas atividades estão examinar os conteúdos do arquivo e indicar periodicamente a aplicação de migrações da informação arquivada, criar orientações para normas e políticas do arquivo e acompanhar as alterações no âmbito tecnológico, nas requisições de serviço e na Base de Conhecimento da Comunidade Alvo. Ademais, retrata modelos de pacotes de informação e efetua adequações de PAI e de PSI para certas submissões, bem como cria planos minuciosos de migração, arquétipos de *software*, e planos de avaliação que viabilizem o estabelecimento das respectivas metas de migração da entidade Administração do Sistema.
6. Acesso – auxilia os Consumidores no estabelecimento da existência, descrição, localização e disposição da informação depositada no SAAI e, além disso, propicia que os Consumidores requisitem e adquiram produtos de informação. Entre as suas obrigações está a comunicação com os Consumidores para obtenção de demandas, o uso de mecanismos para restringir o acesso (em especial para informações protegidas), gerir a realização de requerimentos para

assegurar a sua efetividade, fornecer respostas (PDI, resultados, diagnósticos) e concedê-las aos Consumidores.

O modelo SAAI fornece os fundamentos para o plano de implementação de um sistema de arquivamento de informação digital. A partir de uma abordagem genérica e simplificada das entidades que constituem o seu modelo de informação e modelo funcional, o SAAI descreve como os dados e informações digitais, tratados ou não em instituições arquivísticas, podem ser submetidos, organizados e armazenados num tipo de arquivo bem como mantidos, recuperados e dispostos à longo prazo, refletindo assim as atividades de um repositório digital.

A criação de padrões e orientações associadas ao SAAI²², concederá a cooperação e a interoperabilidade entre distintas comunidades com níveis variados de experiências na gestão, preservação e manutenção do acesso a longo período de materiais digitais, além de viabilizar o uso racional de recursos e uma maior padronização e coordenação na execução ou obtenção de processos, produtos e serviços comuns concernentes às demandas de preservação digital.

3.5 Subsídios para construção de políticas de preservação digital

A definição dos parâmetros e diretrizes para implantação de um plano de preservação e de manutenção do acesso às informações digitais a longo prazo, através de iniciativas nacionais e internacionais²³, são fatores cruciais para tomada de decisões racionais com uso sustentável de recursos e com garantias de perdas não permanentes de registros digitais. Devido ao caráter recente e pouco difuso das investigações e das práticas em prol da preservação digital no Brasil, torna-se de suma importância a identificação e a discussão de algumas metodologias que tem

²² O modelo de referência SAAI não oferece pressupostos no que se refere aos tipos de informações manipuladas e as especificações tecnológicas aplicadas pelo sistema de arquivamento digital. Em vista disto, cada comunidade pode utilizar o SAAI adequando-o ao seu respectivo ambiente organizacional, ademais, o modelo possibilita uma estrutura útil para o desenvolvimento de conjuntos de elementos de metadados de ou para a preservação de longo prazo. Com relação ao uso do SAAI em comunidades detentoras de diferentes tipos de recursos, há vários projetos internacionais já criados com grande notabilidade e valor na preservação digital, como: PANDORA, pela *National Library of Australia; Networked European Deposit Library (NEDLIB); CURL Exemplars in Digital Archives (CEDARS)*, efetuado no Reino Unido e patrocinado pelo *Joint Information Systems Committee (JISC)*; e PREMIS, como grupo de trabalho patrocinado pela OCLC/RLG (CAMPOS; SARAGOGO, 2007; SARAGOGO, 2004).

²³ Podem-se citar, como exemplo, *Digital Preservation Policy: Preserving Archival Digital Records Transferred from Commonwealth Agencies (National Archives of Australia)*, 2011; *Digital preservation and digitisation policy (National Museum of Australia)*, 2012; *Digital Preservation Policy (National Library of Australia)*, 2013; *Digital Preservation Policies: Guidance for archives (The National Archives, United Kingdom)*, 2011; *Building a preservation policy (British Library/Preservation Advisory Centre, England)*, 2013; *Digital Archiving Policy (Swiss Federal Archives, Switzerland)*, 2009; Recomendações para a produção de planos de preservação digital (DGLAB, Portugal), 2011; Programa Permanente de Preservação e Acesso a Documentos Arquivísticos Digitais – AN Digital: Política de Preservação Digital (Arquivo Nacional, Brasil), 2012; dentre outros.

sido recomendadas para a criação, a implementação e a manutenção de políticas e de programas de preservação digital, os quais refletem a realidade social, política e econômica nacional.

Grácio, Fadel e Valentim (2013) propõem uma notável política de preservação digital em IES respaldada em quinze elementos fundamentais, inter-relacionados e embasados nas TIC e na cultura informacional (e organizacional) da instituição, que podem ser distribuídos em três aspectos ou categorias: organizacional, legal e técnico. Somando-se as considerações de Grácio (2012) estes aspectos, capazes de contribuir na criação e determinação de modelos de política e administração da preservação da informação digital em IES, são interpretados conjuntamente e descritos da seguinte forma:

a) Aspectos organizacionais

Abrange elementos de gestão para fixação e estabilização institucional quanto à política e o desenvolvimento contínuo, a médio e longo prazo, das ações de preservação digital:

- ✚ Definição da missão, da visão e dos objetivos institucionais com a inserção da necessidade da implantação de uma política de preservação digital, refletindo a visibilidade acadêmica, científica, tecnológica, artística e cultural desejada, no presente e no futuro, pela instituição defronte à comunidade acadêmico-científica e à sociedade no todo. Os tipos de informação digital a serem preservadas, a quantidade de repositórios institucionais para implementação, o modo de inclusão dos ambientes nas infraestruturas tecnológicas e o conceito de autenticidade de um objeto digital também devem ser definidos neste domínio.
- ✚ Formação de uma equipe multidisciplinar especializada que constituirá o comitê gestor, com papéis e incumbências delimitadas, para a gerência adequada dos fluxos, dos processos e dos compromissos de preservação digital tal como dos objetos digitais produzidos na instituição, garantindo a sua autenticidade (salvaguarda do conteúdo original de sua criação/produção) e determinados níveis ou condições de acesso e recuperação a longo prazo. Para que a equipe tenha uma gestão coordenada e planejada, deve existir uma atuação integrada e em conjunto dos participantes, podendo estes serem de variadas esferas de atuação no âmbito institucional (administrativa, financeira, tecnológica, informacional, jurídica, assim por diante).
- ✚ Estabelecimento dos papéis e das responsabilidades dos profissionais que integram a equipe multidisciplinar, salientando-se distintos tratamentos e gerências para cada tipo de conteúdo. Envolve ainda a conscientização e determinação das obrigações e deveres dos membros que compõem a comunidade acadêmica (docentes, discentes e pessoal técnico-administrativo) quanto a preservação dos seus respectivos objetos digitais produzidos (as informações acadêmico-científicas e de extensão, além das informações administrativas).

- Disposição de recursos financeiros para aplicações contínuas em infraestrutura, tecnologia e capacitação de pessoal (e equipe técnica multidisciplinar). Definição de investimentos anuais com custos envolvidos avaliados, independentes de alterações administrativas, na instância superior ou a nível estratégico da instituição, que permitam a manutenção da política de preservação digital e a construção e preservação da memória da instituição bem como de sua produção intelectual (de natureza acadêmica, científica, técnica, artística e cultural).
- Definição de normas e de procedimentos institucionais para assegurar os direitos dos autores ou produtores dos objetos digitais, a autenticidade e o acesso ao conteúdo de cada tipo de objeto em todo o seu ciclo de vida (desde a sua criação/produção até o seu descarte). Estes atos administrativos, a serem registrados, deferidos e disseminados à comunidade (a equipe multidisciplinar, os produtores e os usuários dos objetos), devem ser baseados na legislação nacional e internacional vigente, permitindo a orientação e a compreensão dos seus membros quanto aos processos, aos métodos, as responsabilidades, os direitos e deveres, os princípios e valores e a conduta a ser seguida²⁴, construindo e consolidando-se assim uma cultura organizacional/informacional favorável direcionada para a preservação digital.

b) Aspectos legais

Compreende questões legais concernentes ao âmbito institucional e a legislação vigente a nível nacional e internacional, como:

- Adesão às leis nacionais sobre os aspectos que envolvam a preservação digital, como a cópia ou reprodução de objetos digitais diante das leis de direitos autorais²⁵ (um dos componentes da propriedade intelectual), como maneira de assegurar a legalidade e a padronização dos fluxos e dos processos de preservação digital, além de garantir a propriedade intelectual aos autores ou produtores dos objetos e sua autenticidade. Ainda que não exista uma legislação

²⁴ Como exemplo, é importante que os criadores/produtores: incluam metadados requeridos logo na fase de criação dos objetos digitais; adotem mídias de armazenamento, por meio de tomadas de decisões informadas, apropriadas para a sustentabilidade dos registros eletrônicos ao longo do tempo; utilizem formatos de arquivos digitais mais adequados para a preservação e o acesso à longo prazo; submetam corretamente objetos nos repositórios digitais, de acordo com os critérios, os padrões e as políticas institucionais; dentre outras (ARELLANO, 2008; BROWN, 2008; LIBRARY AND ARCHIVES CANADA, 2010; SARAGAMAGO, 2004).

²⁵ Hoje, a Lei nº 12.853, de 14 de agosto de 2013, dispõe acerca da gestão coletiva de direitos autorais, modifica, anula e adiciona dispositivos à Lei nº 9.610, de 19 de fevereiro de 1998, que regulamenta os direitos autorais no país (BRASIL, 1998, 2013). A partir de Grácio (2012) e de Grácio, Fadel e Valentim (2013), percebe-se que a Lei nº 9.610 não ampara os fluxos, os processos e as ações de preservação digital, pois preservar envolve a reprodução ou cópia sem fins lucrativos, com o propósito de tornar os objetos digitais acessíveis por longo período. Ademais, os autores ressaltam que a Lei nº 9.610 caracteriza-se por estar afastada do contexto tecnológico contemporâneo e por não fornecer subsídios em prol do direito de acesso às informações pelos cidadãos, essencialmente aquelas geradas pelas IES públicas que registram geralmente o conhecimento produzido para utilização dos indivíduos.

nacional corrente para preservação digital, as instituições requerem um suporte profissional jurídico para o conhecimento da legislação internacional vigente e de suas normas.

c) Aspectos técnicos

Envolve elementos técnicos acerca dos fluxos, dos processos e das ações de preservação digital, tais como:

- Utilização de modelos de referência, métodos, padrões e iniciativas internacionalmente testadas e aderidas, como meios de cooperação e de trocas de experiências e conhecimentos entre as instituições e seus profissionais (com níveis distintos de experiência em preservação digital), apoiando a padronização de componentes inerentes à gerência efetiva e eficaz da política de preservação digital, auxiliando na interoperabilidade dos objetos digitais.
- Existência, readequação e aprimoramento permanente de uma infraestrutura tecnológica (de preservação, com o propósito de proteger os objetos digitais e seus metadados; e de acesso, para viabilizar o acesso, a busca e a recuperação ideal dos objetos), frente a obsolescência do *hardware* e *software*, dos formatos e suportes (mídias) de armazenamento. As alterações e os avanços das TIC associado as propriedades tecnológicas do ambiente digital, favorecem modificações nos objetos originais, ameaçando assim a sua autenticidade e integridade.
- Criação e readequação constante de um programa permanente de capacitação e atualização do pessoal técnico (por meio de cursos, palestras etc.) relativo a manutenção, atualização e aperfeiçoamento da infraestrutura, compreendendo os sucessivos avanços das TIC tal como o nível de capacidade e a variedade de operações realizadas pelos profissionais. A equipe técnica compromete-se com as estratégias de preservação, o plano, o monitoramento e a avaliação tanto das transformações do meio digital como da tendência de obsolescência de tecnologias usadas, além de se instruir e qualificar em tecnologias emergentes e potenciais para adoção e, também, nos próprios aspectos que envolvem a preservação digital.

- Implantação, planejamento, atualização e investimentos ininterruptos em infraestrutura para repositórios digitais ou institucionais²⁶, permitindo através de tecnologias o gerenciamento, o armazenamento e a manutenção de dados bem como o acesso, a busca e a recuperação da informação digital no decorrer do tempo. Ademais, os repositórios permitem uma expansão da visibilidade institucional e viabilizam a divulgação, a disseminação e o livre acesso (*Open Access*), pelos membros da comunidade científica e da sociedade, quanto aos produtos de informação e de conhecimento gerados no interior das instituições, sobretudo públicas.
- Criação e revisão constante de critérios e de procedimentos para a seleção de objetos digitais importantes para serem geridos, armazenados e preservados pelas instituições. Estes critérios e métodos devem ser baseados nos objetivos institucionais e nas demandas da comunidade (os criadores/produtores dos objetos e os usuários destas informações) bem como em termos de definição de prioridades e de custo/benefício de realização da preservação. Além da seleção, deve-se elaborar e atualizar regularmente diretrizes relativas ao descarte de objetos, pautadas na legislação vigente e nos atos normativos internos, e também usar ferramentas específicas criadas e adotadas para este fim (a Tabela de Temporalidade, por exemplo).
- Identificação, adequação, aplicação e atualização continuada de diferentes estratégias de preservação para certos tipos de objetos digitais sob o domínio da instituição, considerando-se o contexto tecnológico original de criação e de tratamento assim como o monitoramento periódico das tecnologias existentes.
- Utilização de conjuntos de elementos ou metadados essenciais para descrição, identificação e representação do conteúdo informacional dos objetos digitais, com o intuito de assegurar efetivamente a sua autenticidade e o seu acesso, busca e recuperação. Dentre os padrões de metadados usados a informações disponibilizadas em IES, estão: o DC, um esquema mais simples, genérico e extensamente adotado; e o METS, tido mais complexo e aprimorado.

²⁶ Há alguns pacotes de *software* livre para implantação de repositórios institucionais, como: o *Flexible Extensible Digital Object Repository Architecture* (FEDORA), criado pelas Universidades de Virginia e de Cornell, é voltado para a construção de bibliotecas digitais e pode abranger serviços de preservação, pois baseia-se no modelo SAAI; e o OCLC *Digital Archive*, disponibilizado pela OCLC e interligado ao sistema de catalogação OCLC's *Conexion*, permite o acesso à longo prazo, o armazenamento e a preservação de objetos digitais, é respaldado no SAAI e adota o padrão de metadados METS para viabilizar a interoperabilidade com outros arquivos, isto é, a comunicação de recursos entre distintos sistemas tecnológicos ou configuração coletiva dos arquivos de modo a padronizá-los completamente. Outro *software* é o *DSpace*, criado pelo *Massachusetts Institute of Technology* (MIT) e produto da efetivação de padrões internacionalmente aceitos (o *Open Archive Initiative*, o SAAI e o *Dublin Core*, por exemplo), é dirigido para ser empregado em repositório que detém a produção intelectual das instituições, abrange diferentes tipos de materiais e dispõe de maiores alternativas de armazenamento e preservação de objetos digitais, é implantado por várias instituições e vêm sendo utilizado tanto para o compartilhamento da produção científica nas universidades como para prover uma união entre repositórios federativos e instituições (ARELLANO, 2008).

► Uso de suportes para infraestrutura de acesso com ampla capacidade de armazenamento, acesso, busca e recuperação efetiva dos objetos digitais. Os suportes digitais (ou seja, meios físicos ou dispositivos de armazenamento de informações digitais) devem ser averiguados periodicamente pela instituição e, em caso de constatação de problemas em sua integridade, necessitam ser substituídos por outros. Entre alguns dos fatores para a transferência de um suporte, estão: a degradação do suporte (decorrente ou não da inadequação de itens externos associados ao âmbito de armazenamento e preservação, como umidade do ar, temperatura, poluentes ambientais e ação de indivíduos responsáveis na sua preservação); a obsolescência do suporte e de seus componentes de acesso, leitura e interpretação dos *bits* armazenados (*hardware e software*); e a própria pretensão institucional de buscar suportes mais atuais.

Uma política de preservação digital deve ser dinâmica e readequada regularmente, em vista da rápida obsolescência tecnológica e dos constantes avanços e alterações das TIC. Inclui aspectos administrativos, econômicos, legais, culturais e técnicos, onde a consolidação de uma cultura organizacional/informacional consciente do valor da preservação digital e uma relação harmoniosa, comprometida e dedicada entre as instituições envolvidas (e seus profissionais), são primordiais para consecução dos processos e dos objetivos institucionais.

Embora possam existir alguns aspectos comuns, cada instituição adequará e implantará uma política específica ao seu contexto, permitindo o acesso por longo prazo às informações e conhecimentos em C&T produzidos e que são necessários para o progresso das organizações e das sociedades. A definição de uma política de preservação digital irá moldar basicamente como cada instituição efetuará os procedimentos que atenderão, de forma efetiva, desde as demandas de preservação que são presentes como também as prováveis necessidades futuras.

4 METADADOS: definições, classificações, metadados de preservação e esquemas

A concepção de metadados está fundamentada na tradicional atividade de catalogação em bibliotecas, cuja principal função é descrever um recurso informacional de forma única, multidimensionando suas formas de acesso, garantindo sua recuperação pelo usuário final. No domínio bibliográfico²⁷ temos como exemplo, o padrão ou formato de metadados MARC 21, que está baseado em regras de codificação definidas pela segunda edição do *Anglo-American Cataloguing Rules (AACR2)*²⁸. Já no domínio Web temos o *Dublin Core (DC)*, cuja finalidade é a de localizar recursos na Web.

O termo domínio Web deve ser entendido como o domínio relacionado aos diferentes tipos de ambientes informacionais digitais na Web, bem como os variados tipos de recursos disponibilizados nela por diferentes domínios do conhecimento. “As representações, entretanto, geradas nesses domínios, serão diferentes das representações geradas no domínio bibliográfico e correspondentes às necessidades informacionais próprias de cada um deles.” (ALVES; SANTOS, 2013, p. 16).

Novos desafios e demandas tem sido propiciados pela ágil e intensa produção e aquisição de conteúdos no universo digital, através do avanço das TIC junto a Internet. Diante desta visão torna-se válido o artigo de Santa Anna (2015), onde o autor aborda alguns aspectos alusivos às mudanças geradas nos típicos serviços biblioteconômicos ante a (r)evolução digital ou tecnológica da pós-modernidade. O autor alega que a utilização das TIC na organização, no tratamento e na representação informacional em meios digitais, implica na aparição de novos conceitos (por exemplo, metadados e ontologias), terminologias, procedimentos e instrumentos mais aprimorados, além de modernos profissionais catalogadores. Entre estes instrumentos está a *Resource Description and Access (RDA)* que, de acordo com o autor, objetiva aperfeiçoar as capacidades do AACR2 abrangendo as descrições para os objetos/documentos digitais, além dos ambientes clássicos das bibliotecas, mas potencialmente aplicáveis a quaisquer ambientes informacionais.

Outro ponto importante remete à complexidade existente na gestão de conteúdos digitais para a sua preservação ao longo do tempo, com salvaguarda de autenticidade e integridade. O

²⁷ O termo domínio bibliográfico é baseado na definição da *International Federation of Library Associations and Institutions (IFLA)* para Universo Bibliográfico que é o domínio relacionado às coleções de bibliotecas, arquivos, museus e outras comunidades de informação (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 1998). Assim, a acepção aqui adotada para domínio bibliográfico “[...] designa o campo de estudo voltado para o tratamento descritivo da informação (processo de representação do recurso) em bibliotecas.” (ALVES; SANTOS, 2013, p. 16).

²⁸ Disponível em: <<http://www.aacr2.org/>>. Acesso em: 15 set. 2015.

ambiente digital pode ser retratado, segundo Campos (2002) e Saramago (2002), como aquele que se caracteriza pela fragilidade, volatilidade e permanente dinâmica. Por sua vez, os objetos digitais a serem mantidos, sejam eles nascidos digitais ou digitalizados, são também dinâmicos e vulneráveis quanto a sua alta dependência de tecnologias que se tornam obsoletas rapidamente ou que são danificadas fisicamente (FORMENTON; GRACIOSO; CASTRO, 2015). Isto posto, métodos de preservação digital, tal como o uso de padrões de metadados, vêm sendo estudados para amenizar a perda definitiva de informações relevantes para as gerações atuais e futuras.

Assim, os aspectos supracitados permitem mostrar algumas das novas necessidades que refletiriam em uma extensão da concepção clássica de metadados. Além do seu papel descritivo, inerente na organização informacional em prol da eficácia dos sistemas de busca e recuperação de informação, os metadados podem ter a função de apoiar a preservação digital. Os metadados de preservação registram informações do conteúdo de objetos digitais e dados administrativos, estruturais e técnicos, propiciando gerir o arquivamento a longo prazo e o acesso contínuo aos objetos, com garantias de autenticidade e integridade. Dados sobre cadeia de custódia, medidas de preservação adotadas, direitos de propriedade e ambiente tecnológico exigido para o uso dos objetos, são algumas das informações abrangidas pelos metadados para preservação.

Neste contexto, múltiplos padrões ou esquemas de metadados tem sido desenvolvidos e propostos de modo a contemplar as necessidades da preservação digital. Por exemplo, temos o conjunto de metadados de preservação do PREMIS e, ainda, o conjunto de metadados técnicos do ANSI/NISO Z39.87 que, mesmo não sendo dirigidos exatamente para a preservação, buscam ser harmônicos com o PREMIS. Com base na revisão da literatura especializada, são abordadas algumas definições e classes para metadados; o conceito de metadados de preservação e as informações requeridas para a preservação de longo prazo; além dos principais padrões de metadados utilizados no suporte ao gerenciamento da preservação digital.

4.1 Definição e classificação para metadados

O termo metadados pode apresentar distintas definições e interpretações dependendo do contexto de seu emprego ou serventia. Embora a conceituação de metadados como “dados sobre dados” tenha-se enraizado em certas áreas do conhecimento, esta definição passou a ser limitada e simplória defronte à multiplicidade das finalidades e das atividades designadas aos metadados nos domínios vigentes de gerência da informação, como os ambientes digitais. As imposições advindas das problemáticas do meio digital e eletrônico possibilitaram uma reconfiguração da

concepção básica meramente descritiva de metadados, ampliando as suas conceitualizações de modo a contemplar novas ocupações e utilidades (SAYÃO, 2010; THOMAZ, 2004).

Para uma compreensão mais ampla do conceito de metadados e de suas aplicabilidades, algumas definições são relevantes para as discussões acerca do assunto. Há na literatura científica da área de Ciência da Informação várias definições e conceitos sobre metadados e padrões de metadados. Neste trabalho nos deteremos ao conceito de metadados determinado por Grácio (2012), pois este contempla as principais características e metodologias para garantir a preservação digital. Os metadados referem-se à “[...] um conjunto de dados, chamados ‘elementos’, cujo número varia de acordo com o padrão adotado, e que descreve o recurso, possibilitando a um usuário ou a um mecanismo de busca acessar e recuperar esse recurso.” (GRÁCIO, 2012, p. 121). Para Barbedo, Corujo e Sant’Ana (2011) e *National Museum of Australia* (2012), no âmbito da preservação digital, os metadados são dados, estruturados e padronizados, que descrevem e documentam as propriedades de um recurso provendo-lhe sentido, contextura e sistematização, compondo-se da catalogação ou indexação de informações para produção, localização, identificação, gestão, utilização e preservação de recursos no decorrer do tempo em diversos domínios.

Segundo Sayão (2010) os metadados podem ser agrupados em três classes conceituais:

- ⊕ Metadados descritivos – tratam-se daqueles que descrevem um recurso informacional para fins de descoberta e identificação. Podem conter elementos como, por exemplo, título, autor, resumo, palavras-chave e identificador persistente.
- ⊕ Metadados estruturais – refere-se às informações que registram de que maneira devem ser reconstituídos e organizados os recursos complexos formados de diversos componentes. Por exemplo, a forma como as páginas de um relatório, submetidas ao processo de digitalização à parte, são relacionadas entre si e arranjadas para compor um capítulo.
- ⊕ Metadados administrativos – provém informações que auxiliam as atividades de gerência do ciclo de vida dos recursos. Como exemplo, abrangem informações acerca de que modo e em que momento houve a produção do recurso e as motivações ou pretextos para isto. Podem incluir metadados para suporte à coordenação dos direitos relativos ao recurso e metadados técnicos que determinam as particularidades e dependências técnicas do recurso.

Através de Barbedo, Corujo e Sant’Ana (2011) são observadas múltiplas categorias funcionais de tipos de metadados, que são assim entendidas:

- ⊕ Metadados descritivos ou de identificação – visam à pesquisa, recuperação e identificação. Podem conter elementos como, por exemplo, título, autor, assunto e palavras-chave.

- Metadados estruturais – vinculam de forma hierárquica distintos objetos digitais (textos, imagens, áudios etc.) integrantes de um mesmo documento ou recurso informacional.
- Metadados administrativos – dispõem informações que suportam a gerência dos recursos arquivísticos eletrônicos. Incluem de que forma e em que ocasião os recursos foram gerados, espécie de ficheiro ou de arquivo de computador e demais informações técnicas, além dos titulares com direitos ou permissões de acesso.
- Metadados técnicos – especificam os aspectos técnicos dos arquivos e dos seus formatos.
- Metadados de preservação – incluem informações requeridas ao arquivamento e salvaguarda dos objetos digitais ao longo do tempo.

Barbedo, Corujo e Sant’Ana (2011) apontam também três subcategorias de tipos de metadados, isto é, os de gerência de controle e direitos, concernentes aos direitos de propriedade intelectual; os de controle e gestão; e os de âmbito de produção. Assim, o conjunto de categorias supracitadas expõem a variedade de informações, de papéis e de propósitos que os metadados podem vir a incorporar ou abranger. Para os propósitos da presente investigação destacaremos a categoria de metadados relativos à preservação de longo prazo, uma vez que são fundamentais para a consecução de uma efetiva gerência e preservação dos recursos/objetos digitais.

4.2 Metadados para a preservação digital

Os metadados de preservação são essenciais para a preservação digital. De acordo com Sayão (2010, p. 10) estes tipos de metadados são conceituados “[...] de uma forma simples e direta, como a informação que apoia e documenta a preservação de longo prazo de materiais digitais”. Em Saramago (2004, p. 1) comprehende-se por metadados de preservação “[...] como informação de apoio aos processos associados com a preservação digital de longo prazo”, sendo que para Arellano (2008) tratam-se daqueles associados ao conteúdo do recurso, seu contexto e estrutura de criação/produção, além das alterações ocorridas em todo o seu ciclo de vida.

Campos e Saramago (2007) e Saramago (2004) concebem os metadados de preservação em descritivos, administrativos e estruturais. Na divisão, os descritivos são tidos pelas autoras como aplicáveis às etapas de acesso e destinados a descrição dos recursos digitais, o registro das ações de gerência durante o tempo cabem aos administrativos, enquanto que os estruturais incluem o “quadro” tecnológico requerido para ideal representação e recuperação dos recursos, integrando assim as informações administrativas. Nestas duas últimas classes, os metadados administrativos e os estruturais, a atenção é pertinente posto que teremos as informações que especificam o conjunto de métodos e de estratégias adotadas para preservação.

Reunindo as ponderações de Barbedo, Corujo e Sant'Ana (2011), Campos e Saramago (2007), *National Museum of Australia* (2012), Saramago (2004), Sayão (2010) e Grácio (2012) pode-se identificar algumas das distintas informações e funções inter-relacionadas circunscritas pela obtenção, geração, guarda e manutenção de metadados para preservação, tais como:

- ✚ Registro do histórico, do *hardware*, *software* e outras dependências técnicas, das estruturas e das mudanças (de custódia legal e física ou de natureza tecnológica, social etc.) suportadas pelos recursos/objetos digitais, no decurso de todo o seu ciclo de vida (isto é, criação, seleção e descarte, identificação persistente, descrição e acesso, armazenamento e preservação).
- ✚ Documentação das tomadas de decisão e atos de preservação, dos métodos de preservação e seus resultados, das ações de gestão de coleções e direitos, além dos próprios metadados, de modo a apoiar os processos ou as iniciativas correntes e vindouras de preservação digital.
- ✚ Suporte à descoberta, localização, recuperação, referência única e persistente, preservação, acesso, utilização, comprovação de autenticidade e integridade, bem como gestão de direitos de propriedade intelectual de objetos arquivísticos e museológicos digitais ou eletrônicos.
- ✚ Assistência à uma ideal gerência dos objetos digitais, dos sistemas de informação (registros de correspondência e de documentos, registros de funcionamento dos sistemas e seus acessos por intermédio de práticas de auditorias mantidas pelos serviços de TIC, por exemplo), da comunidade usuária e das funcionalidades a serem oferecidas.

Em suma, os metadados de preservação registram informações do conteúdo de recursos digitais e dados administrativos, estruturais e técnicos para preservação digital. Estes metadados propiciam a preservação por longo prazo e o acesso contínuo aos objetos digitais, com garantias de autenticidade, integridade e confiabilidade. Neste cenário, a determinação e, possivelmente, adaptação de padrões de metadados torna-se uma ação indispensável e complexa, devendo-se incorporar os vários tipos específicos de processos e de recursos digitais tratados e preservados, o rol abundante de informações a serem registradas e fornecidas e, ainda, as decisões tomadas diante de um futuro de imprevisões intrínsecas à preservação em ambientes digitais.

Na construção de esquemas de metadados de preservação o desafio básico está, segundo Lavoie e Gartner (2005), na previsão e acerto das informações que de fato serão exigidas, pois: a) Alusivo a amplitude e profundidade dos metadados, é preciso ponderar alguns fatores como, por exemplo, o “grau” de preservação (guarda do conteúdo, apresentação etc.) e a duração do arquivamento; e b) A adequação dos elementos se definirá apenas num grande período posterior a sua implantação, ensejando assim a avaliação da eficácia do esquema e da deficiência ou não de informações para a preservação digital. Para os autores um esquema deste tipo deve ser:

- 1) Abrangente – o esquema deve possuir uma extensão que, referente a escopo e profundidade, conte cole as necessidades atuais e futuras de preservação do repositório concebido. Assim, preza-se por esquemas de maior amplitude, em vez de estruturas limitas que, possivelmente, serão estendidas com o tempo em virtude da necessidade de informações adicionais.
- 2) Orientado para a implantação – o esquema deve ser adaptável aos sistemas automatizados para coleta e gerenciamento de metadados. Além disto, deve ser projetado considerando os aspectos práticos de implementação. Como exemplo, fornecendo códigos ou vocabulários controlados para preenchimento de elementos, ao invés de basear em “texto livre”.
- 3) Interoperável – o esquema deve ser projetado para promover e simplificar as transações que envolvam os objetos digitais arquivados e os seus metadados associados ao longo de todo o seu ciclo de vida. Por exemplo, a submissão inicial a um repositório, a disseminação para um usuário ou a transferência para outro sistema de repositório.

Aludindo-nos à importância da definição de metadados na etapa de criação dos recursos digitais, algumas propostas são significantes. A título de exemplo, o Arquivo Nacional do Brasil instituiu o Programa AN Digital²⁹ que, através de Arquivo Nacional (2012), expõe uma política de preservação digital e processos de recolhimento de documentos arquivísticos digitais. Assim, a inserção dos documentos no repositório do AN Digital, exige que estes venham seguidos, por exemplo, de um conjunto de metadados e de informações contidas numa listagem descritiva:

- a. Metadados descritivos – inclui elementos básicos para suporte à presunção de autenticidade dos documentos, como autor, destinatário, produtor, âmbito e conteúdo, data de criação e de recebimento, código de referência original, título e dimensão; além de elementos apetecíveis, como assunto, data da transmissão e da captura ou arquivamento, código de classificação, indicação de anexo e de anotação, nome do setor incumbido pela execução da ação inclusa no documento, registro das migrações e data das suas ocorrências, restrição de acesso.
- b. Informações da listagem descritiva – abrange informações acerca das propriedades do acervo digital e histórico de criação e manutenção dos documentos no produtor/custodiador como, por exemplo, órgão ou entidade responsável pela produção e acumulação dos documentos; indicação dos processos para prevenção, descoberta e correção de perdas ou adulterações

²⁹ O Programa Permanente de Preservação e Acesso a Documentos Arquivísticos Digitais – AN Digital, instaurado em março de 2010, propõem-se a construir uma infraestrutura no Arquivo Nacional que possibilite o recolhimento, a descrição, o armazenamento, a preservação e o acesso aos documentos arquivísticos digitais criados/custodiados por órgãos ou entidades da Administração Pública Federal. Nesta acepção, o Programa AN Digital possui algumas atividades: a definição de procedimentos, de padrões e da abordagem de gestão e preservação digital do Arquivo Nacional; a elaboração e a implementação de um repositório; a aquisição de equipamentos e a organização de uma equipe especializada (ARQUIVO NACIONAL, 2010, 2012). Outras informações sobre o programa e o andamento de suas atividades podem ser obtidas através do Sistema de Gestão de Documentos de Arquivo – SIGA. Disponível em: <<http://www.siga.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm?sid=222>>. Acesso em: 26 abr. 2015.

dos documentos; reconhecimento dos formatos de arquivo digital; indicação das diretrizes para autenticação dos documentos; metadados para interpretar e representar os documentos, bem como a estrutura da base de dados, o esquema HTML e de metadados; dentre outras.

Nesta perspectiva, os produtores/custodiadores precisam ser orientados pelo repositório para uma adequada inserção e manutenção dos metadados descritivos. Em concordância com as boas práticas de preservação digital, a documentação gerada permitirá o depósito de recursos aceitos que incluem condições favoráveis à aplicação de processos futuros de preservação e de interpretação do contexto original de produção e salvaguarda. Deste modo, integrando-se às informações atribuídas pela instituição de preservação de longo prazo, os registros apoiarão a autenticidade e a integridade dos recursos digitais, desde a criação até a sua preservação.

4.3 Esquemas de metadados

A organização dos metadados dá-se em estruturas formais chamadas de esquemas de metadados, que consistem de conjuntos de elementos concebidos para um fim específico, como a descrição de um certo tipo de recurso de informação. Na literatura³⁰, as expressões “esquema”, “formato”, “sistema” e “conjuntos de elementos” têm sido usadas indistintamente para se referir a padrões de metadados (CHAN; ZENG, 2006; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004). De acordo com Zeng e Qin (2008, p. 323, tradução nossa) o esquema de metadados (*Metadata schema*) pode ser considerado como:

Uma especificação processavelmente por máquinas que define a estrutura, a codificação de sintaxe, regras, e formatos para um conjunto de elementos de metadados em uma linguagem formal num esquema. Na literatura o termo “*metadata schema*” usualmente refere-se ao conjunto de elementos na sua totalidade, assim como a codificação dos elementos e a estrutura com uma linguagem de marcação.

De fato, o “esquema” é uma entidade total contendo os componentes semânticos e de conteúdo (tidos como um “conjunto de elementos”), bem como a codificação dos elementos

³⁰ Entre os exemplos incluem as expressões: “sistema de metadados” em Campos e Saramago (2007); “formato de metadados” em Sayão (2010); “esquema de metainformação” em Barbedo et al. (2007); “modelo de metadados” em Arellano (2008); “*metadata scheme*”, “*metadata schema*” e “*element set*” em Chan e Zeng (2006).

com uma sintaxe ou linguagem de marcação, como SGML e XML³¹. Assim, um conjunto de elementos de metadados dispõe de dois componentes básicos: 1) Semântica – as definições ou os significados dos elementos e seus refinamentos; e 2) Conteúdo – as declarações ou as instruções de quais e como os valores devem ser atribuídos para os elementos. Para cada elemento definido, um padrão de metadados geralmente especifica regras de conteúdo para como o conteúdo deve ser criado ou incluído (por exemplo, como identificar o título principal), regras de representação para conteúdo (por exemplo, padrões de representação do tempo) e valores de conteúdo admissíveis (isto é, se os termos devem ser tomados a partir de um vocabulário controlado específico ou podem ser providos pelo autor, derivados do texto, ou aditados pelo trabalho de criadores de metadados sem uma lista de termos controlados). Pode haver ainda regras de sintaxe para codificação dos elementos e seu conteúdo (CHAN; ZENG, 2006; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004).

Da revisão de literatura pode-se identificar variados padrões ou esquemas de metadados usados para a descrição de recursos em distintos domínios, conforme exposto no Quadro 2.

³¹ XML é uma linguagem de marcação de dados produzida pelo *World Wide Web Consortium* (W3C), que através de um conjunto de etiquetas (*tags*), viabiliza a descrição textual e estruturada de informações. É um padrão aberto, independente das plataformas operacionais e dos fabricantes de *software*, comprensível por diferentes aplicações de *software* ou sistemas de informação e autoexplicativo. Esta linguagem mantém a semântica da informação (isto é, a alusão a um mesmo termo, mas com sentidos opostos em âmbitos distintos) e dispõe de uma estrutura flexível e simples. Proporciona o fácil intercâmbio de dados entre diversos sistemas e a formatação das etiquetas (que são extensíveis e definidas pelo usuário), a documentação de suas definições e a determinação do seu esquema. Por outro lado, a SGML é uma linguagem semelhante ao XML, que possibilita uma rica descrição formal da estrutura e dos conteúdos dos documentos para fins de gestão e manipulação de dados ou informações. A SGML é definida na norma ISO 8879:1986 (ANDRADE, 2007; LIBRARY AND ARCHIVES CANADA, 2010).

Quadro 2 – Padrões ou esquemas de metadados e seus escopos.

| Nome | Especificação |
|--|---|
| ANSI/NISO Z39.87, <i>Data Dictionary – Technical Metadata for Digital Still Images</i> ³² | Padrão de metadados técnicos para desenvolver, intercambiar e interpretar arquivos de imagem fixa digital, facilitar a interoperabilidade entre sistemas, serviços e software, assim como apoiar a gestão a longo prazo e o contínuo acesso a coleções de imagens digitais (BARBEDO et al. 2007; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004, c2011). |
| <i>Dublin Core (DC)</i> ³³ | Padrão para catalogação e suporte a descoberta de recursos eletrônicos na Web (CAMPOS; SARAMAGO, 2007; GRÁCIO, 2012; SAYÃO, 2010). |
| <i>Encoded Archival Description (EAD)</i> ³⁴ | Padrão para codificação de instrumentos arquivísticos de acesso, tais como inventários, índices, dentre outros (ANDRADE, 2007; BARBEDO et al. 2007; BARBEDO; CORUJO; SANT'ANA, 2011; CHAN; ZENG, 2006). |
| <i>Learning Object Metadata (LOM)</i> ³⁵ | Padrão que define um conjunto mínimo de atributos para gerir, localizar e avaliar recursos ou objetos de aprendizagem (CHAN; ZENG, 2006; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004). |
| <i>Machine Readable Cataloging (MARC)</i> ³⁶ | Padrão para a representação e a comunicação de registros ou informações bibliográficas entre bibliotecas, através da utilização de computadores (CHAN; ZENG, 2006; GRÁCIO, 2012; THOMAZ, 2004). |
| <i>Metadata Encoding and Transmission Standard (METS)</i> ³⁷ | Padrão para codificar metadados descritivos, administrativos e estruturais sobre objetos digitais num repositório (ALMEIDA; CENDÓN; SOUZA, 2012; BARBEDO et al. 2007; CAMPOS; SARAMAGO, 2007). |
| <i>Metadata Object Description Schema (MODS)</i> ³⁸ | Padrão de metadados descritivos derivado do esquema bibliográfico MARC 21, cujo o enfoque são os recursos eletrônicos e os serviços de biblioteca (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004). |
| <i>ISO/IEC Moving Picture Experts Group (MPEG) Multimedia Metadata</i> ³⁹ | Padrões para a descrição e a representação de recursos ou objetos audiovisuais e multimídia digitais, tais como MPEG-7 e MPEG-21 (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2004; SAYÃO, 2010). |
| <i>PREServation Metadata: Implementation Strategies (PREMIS) Data Dictionary</i> ⁴⁰ | Padrão que define um conjunto de elementos básicos para codificar, armazenar, gerir e intercambiar metadados de preservação no contexto de um sistema de repositório para preservação digital (ARELLANO, 2008; CAMPOS; SARAMAGO, 2007; SAYÃO, 2010). |
| <i>Record Export for Art and Cultural Heritage (REACH)</i> ⁴¹ | Padrão para a descrição de objetos no âmbito de coleções de museus (GRÁCIO, 2012). |

Fonte: Elaborado pelo autor.

Múltiplos padrões de metadados vêm sendo elaborados para uma extensa diversidade de usos e âmbitos específicos, todavia os esquemas são delimitados por seus próprios conjuntos

³² Disponível em: <http://www.niso.org/apps/group_public/project/details.php?project_id=69>. Acesso em: 30 maio 2015.

³³ Disponível em: <<http://dublincore.org/>>. Acesso em: 2 jun. 2015.

³⁴ Disponível em: <<http://www.loc.gov/ead/>>. Acesso em: 30 maio 2015.

³⁵ Disponível em: <<https://standards.ieee.org/findstds/standard/1484.12.1-2002.html>>. Acesso em: 1 jun. 2015.

³⁶ Disponível em: <<http://www.loc.gov/marc/>>. Acesso em: 30 maio 2015.

³⁷ Disponível em: <<http://www.loc.gov/standards/mets/>>. Acesso em: 30 maio 2015.

³⁸ Disponível em: <<http://www.loc.gov/standards/mods/>>. Acesso em: 2 jun. 2015.

³⁹ Disponível em: <<http://mpeg.chiariglione.org/standards>>. Acesso em: 30 maio 2015.

⁴⁰ Disponível em: <<http://www.loc.gov/standards/premis/>>. Acesso em: 1 jun. 2015.

⁴¹ Disponível em: <<http://www.oclc.org/research/activities/museumresources/reach.html>>. Acesso em: 2 jun. 2015.

de elementos de metadados, particularidades e domínios de utilização. A seguir são discutidos alguns dos principais e mais correntes padrões tratados na literatura especializada.

4.3.1 Padrão *Dublin Core*

O padrão DC tem sua origem em Chicago, na 2^a Conferência Internacional *World Wide Web*, em 1994, numa discussão sobre semântica e a *Web* enfocada na dificuldade da descoberta de recursos de informação. Posteriormente, a OCLC e o *National Center for Supercomputing Applications* (NCSA) organizaram um evento em Dublin, Ohio, em 1995. No evento, intitulado de "OCLC/NCSA *Metadata Workshop*", mais de 50 profissionais debateram como um conjunto semântico seria útil para a busca e a recuperação de recursos baseados na *Web*. Como resultado definiu-se um conjunto básico de elementos de metadados chamado "*Dublin Core*", em razão do local do evento⁴² (DUBLIN CORE METADATA INITIATIVE, c2015; HARPER, 2010).

Existem dois níveis no padrão DC: Simples e Qualificado. Sendo assim, o DC Simples inclui quinze elementos; e o DC Qualificado compreende três elementos adicionais (Audiência, Proveniência e Detentor de Direitos), assim como um grupo de refinamentos ou qualificadores de elementos, que aperfeiçoam a semântica dos elementos de maneira que possam ser úteis na descoberta de recursos. O DC apresenta algumas características, que refletem também os seus objetivos, dentre os quais são as seguintes (GRÁCIO, 2012; HILLMANN, 2005a):

- Simplicidade – o conjunto de elementos DC tem sido mantido pequeno e simples, de modo que sejam compreensíveis semanticamente. Um pesquisador ou responsável da coleção, não especialista, pode gerar facilmente registros descritivos para os recursos de informação.
- Interoperabilidade semântica – a existência de distintos esquemas de descrição intervém na capacidade das pesquisas e das descobertas entre áreas. Um conjunto de elementos comum, universalmente inteligível e apoiado, propicia maior interoperabilidade entre estas áreas.
- Consenso internacional – a participação de especialistas, órgãos, ou representantes de vários países na DCMI para fins de internacionalização na *Web* e de uma estrutura apropriada (que reflita a natureza multilíngue e pluricultural do universo da informação eletrônica), colabora para um consenso internacional e uma contínua melhora do DC (junto ao avanço da *Web*).

⁴² O conjunto de quinze elementos de metadados DC versão 1.1 são determinados na IETF *Request for Comments* (RFC) 5013, de agosto de 2007, na ISO 15836, de fevereiro de 2009, e na ANSI/NISO Z39.85-2012, de fevereiro de 2013. Um outro resultado significativo foi o surgimento, na década de 1990, da *Dublin Core Metadata Initiative* (DCMI), que respaldando-se em Grácio (2012) e Weibel e Koch (2000), pode ser vista como uma organização ou comunidade internacional voltada a proporcionar a disseminação do emprego de padrões de interoperabilidade de metadados e a elaboração de vocabulários específicos de metadados de apoio a descrição e a descoberta de recursos da *Web*, que oportunizem mais sistemas efetivos para a busca e a recuperação de informações.

- Extensibilidade – o padrão DC é um esquema simplificado de descrição, que na criação de esquemas, permite conjuntos de metadados adicionais para acudir as demandas de descrição e recuperação precisa de um recurso digital particular numa determinada comunidade. Estes elementos de metadados adicionais junto com os elementos DC, possibilitam que múltiplas comunidades em várias áreas usem o padrão DC, permutem e acessem as informações.
- Flexibilidade – os elementos são opcionais e repetíveis, não há ordem definida e podem ser alterados adotando um limitado conjunto de qualificadores, que são atributos utilizados para refinar (não estender) o significado do elemento.

Baseando-se em Campos e Saramago (2007), *Dublin Core Metadata Initiative* (2012) e Hillmann (2005c), os quinze elementos de metadados do padrão DC Simples versão 1.1 são:

1. Título – o nome ou o título atribuído ao recurso.
2. Criador – uma entidade principal responsável pela produção do conteúdo do recurso. Uma pessoa, uma organização, ou um serviço, são exemplos de Criador. Geralmente, o nome do Criador deve ser usado para indicar a entidade. Tais exemplos e modo de indicação da entidade aplicam-se, também, para os elementos Editor e Colaborador.
3. Assunto – o tópico ou a temática do conteúdo do recurso. Normalmente, o assunto será representado usando palavras-chave, frases-chave, ou códigos de classificação. Indica-se adotar um valor de um sistema de classificação formal ou vocabulário controlado⁴³.
4. Descrição – um relato do conteúdo do recurso. Pode conter, por exemplo, um índice, um resumo, uma representação gráfica, ou um relato de texto livre referente ao conteúdo.
5. Editor – uma entidade responsável por tornar o recurso disponível.
6. Colaborador – uma entidade responsável por fazer contribuições ao conteúdo do recurso.
7. Data – um período de tempo associado com um evento no ciclo de vida do recurso, como a data de criação ou de procedimento de preservação etc. Aconselha-se usar um esquema de codificação, como o perfil W3C-DTF⁴⁴ da norma ISO 8601⁴⁵.
8. Tipo – a natureza ou o gênero do conteúdo do recurso. Recomenda-se selecionar um valor a partir de um vocabulário controlado, tal como o DCMI *Type Vocabulary*⁴⁶.

⁴³ Entre os exemplos de esquemas de codificação para assunto, estão: a *United States (US) Library of Congress Classification* (LCC), a *Dewey Decimal Classification* (DDC), a *Universal Decimal Classification* (UDC), e a *US National Library of Medicine (NLM) Classification*, como um conjunto de recursos conceituais especificados; tal como o *US NLM Medical Subject Headings* (MeSH) e o *US Library of Congress Subject Headings* (LCSH), como um conjunto de conceitos categorizados e especificados (DUBLIN CORE METADATA INITIATIVE, 2012).

⁴⁴ Disponível em: <<http://www.w3.org/TR/NOTE-datetime>>. Acesso em: 21 jun. 2015.

⁴⁵ Disponível em: <<http://www.iso.org/iso/home/standards/iso8601.htm>>. Acesso em: 21 jun. 2015.

⁴⁶ Disponível em: <<http://dublincore.org/documents/dcmi-type-vocabulary/>>. Acesso em: 21 jun. 2015.

9. Formato – a manifestação física ou digital do recurso. Inclui o formato de arquivo, a forma física, ou as dimensões (tamanho e duração, por exemplo) do recurso. Indica-se o uso de um vocabulário controlado, como a lista de *Internet Media Types* ou *MIME Types*⁴⁷.
10. Identificador – uma referência não ambígua para o recurso num dado contexto. Orienta-se identificar o recurso por meio de uma sequência ou número de acordo com um sistema de identificação formal, como o *Uniform Resource Locator* (URL), o ISBN e o DOI. Esta forma de identificação ou referência a um recurso pode também ser aplicada para os elementos Fonte e Relação.
11. Fonte – uma referência a um outro recurso a partir do qual o recurso descrito é derivado no todo ou em parte.
12. Língua – o idioma do conteúdo intelectual do recurso. Sugere-se utilizar um vocabulário controlado, como o RFC 4646⁴⁸ em conjunto com a norma ISO 639⁴⁹.
13. Relação – uma referência para outro recurso do qual o recurso descrito é derivado ou está relacionado, como a parte, a versão, a tradução, ou a dependência de um recurso etc.
14. Cobertura – o âmbito espaço-temporal do conteúdo do recurso. Pode conter a designação de um lugar, de um período ou de uma entidade administrativa, além de datas, intervalos de datas e a especificação de um local via suas coordenadas geográficas. Orienta-se adotar um vocabulário controlado, como o *Getty Thesaurus of Geographic Names* (TGN)⁵⁰.
15. Direitos – informação sobre direitos do recurso. Inclui uma indicação dos vários direitos de propriedade alusivos ao recurso, como os direitos de propriedade intelectual.

De acordo com Hillmann (2005b) e Weibel e Koch (2000) a DCMI, defronte à demanda de uma representação mais detalhada dos recursos, definiu um conjunto de qualificadores para aperfeiçoar a semântica dos quinze elementos do DC, o qual caracterizam-se por: basearem em vocabulários controlados, notações e termos existentes mantidos por agências estabelecidas; serem flexíveis e abertos à propostas e melhorias futuras; e serem livres para uso ou não pelos implementadores, podendo haver a criação de qualificadores adicionais para aplicações locais ou domínios específicos, porém os mesmos não poderão ser entendidos por outras aplicações e, ainda, serão menos propensos de serem reutilizados por demais comunidades.

Para os autores a DCMI reconhece duas grandes classes de qualificadores:

⁴⁷ Disponível em: <<http://www.iana.org/assignments/media-types/media-types.xhtml>>. Acesso em: 21 jun. 2015.

⁴⁸ Disponível em: <<http://www.ietf.org/rfc/rfc4646.txt>>. Acesso em: 21 jun. 2015.

⁴⁹ Disponível em: <http://www.iso.org/iso/home/standards/language_codes.htm>. Acesso em: 21 jun. 2015.

⁵⁰ Disponível em: <<http://www.getty.edu/research/tools/vocabularies/tgn/>>. Acesso em: 21 jun. 2015.

- 1) Refinamento de elemento – os qualificadores tornam o significado do elemento mais estreito ou mais específico. Assim, um elemento refinado compartilha o significado do elemento não qualificado, contudo, com um escopo mais restrito ou detalhado.
- 2) Esquema de codificação – os qualificadores identificam esquemas de apoio na interpretação de um valor de elemento. Compreende vocabulários controlados (por exemplo, um termo de um sistema de classificação ou conjunto de cabeçalhos de assunto) e notações formais (por exemplo, uma sequência “2015-06-30” como a expressão padrão de uma data).

A Figura 4 apresenta a estrutura de um documento no formato de metadados *Dublin Core* expresso em linguagem XML, a seguir.

FIGURA 4 – Exemplo de um registro DC em XML.

```

<srw_dc:dc xmlns:srw_dc="info:srw/schema/1/dc-schema" xmlns="http://purl.org/dc/elements/1.1/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="info:srw/schema/1/dc-schema
  http://www.loc.gov/standards/sru/resources/dc-schema.xsd">
  <title>War! Or no war [graphic].</title>
  <creator>Odham, Thomas.</creator>
  <creator>Palmer, F. (Fanny), 1812-1876.</creator>
  <creator>F. & S. Palmer (Firm)</creator>
  <type>still image</type>
  <type>Lithographs 1840-1850. gmgpc</type>
  <type>Political cartoons 1840-1850. gmgpc</type>
  <publisher/>
  <date>1846.</date>
  <language>eng</language>
  <format>p</format>
  <format>p</format>
  <description>
    Another cartoon relating to the dispute between the United States and Great Britain over the northern boundary of American territory in Oregon. Opinion was sharply divided between support for a compromise claim of territory as far north as the forty-ninth parallel, and those who went for the more expansive 4.40 boundary. Palmer has two Irish immigrants disputing the issue. Wearing overalls, suspenders, boots, and stovepipe hats, the characters stand outside the Bowery Theater. (The Bowery section of New York City was known for its concentration of working-class Irish.) The man on the left says, "Ike! say the 49th & let's settle it amercably." The other responds, "No Sir-ree I goes for the hull of Oregon or none--I do & don't do nor-thin else." In the background, a woman sells apples and what look like nuts at a table. Behind her the theater's billboard advertises "Great Attraction. Adelgis. J.R. Scott. Cony & Blanchard."
  </description>
  <description>Weitenkampf,</description>
  <description>
    Another cartoon relating to the dispute between the United States and Great Britain over the northern boundary of American territory in Oregon. Opinion was sharply divided between support for a compromise claim of territory as far north as the forty-ninth parallel, and those who went for the more expansive 4.40 boundary. Palmer has two Irish immigrants disputing the issue. Wearing overalls, suspenders, boots, and stovepipe hats, the characters stand outside the Bowery Theater. (The Bowery section of New York City was known for its concentration of working-class Irish.) The man on the left says, "Ike! say the 49th & let's settle it amercably." The other responds, "No Sir-ree I goes for the hull of Oregon or none--I do & don't do nor-thin else." In the background, a woman sells apples and what look like nuts at a table. Behind her the theater's billboard advertises "Great Attraction. Adelgis. J.R. Scott. Cony & Blanchard."
  </description>
  <description>Weitenkampf,</description>
  <description>
    Another cartoon relating to the dispute between the United States and Great Britain over the northern boundary of American territory in Oregon. Opinion was sharply divided between support for a compromise claim of territory as far north as the forty-ninth parallel, and those who went for the more expansive 4.40 boundary. Palmer has two Irish immigrants disputing the issue. Wearing overalls, suspenders, boots, and stovepipe hats, the characters stand outside the Bowery Theater. (The Bowery section of New York City was known for its concentration of working-class Irish.) The man on the left says, "Ike! say the 49th & let's settle it amercably." The other responds, "No Sir-ree I goes for the hull of Oregon or none--I do & don't do nor-thin else." In the background, a woman sells apples and what look like nuts at a table. Behind her the theater's billboard advertises "Great Attraction. Adelgis. J.R. Scott. Cony & Blanchard."
  </description>
  <description>
    Forms part of: American cartoon print filing series (Library of Congress)
  </description>
  <description>
    Published in: American political prints, 1766-1876 / Bernard F. Reilly. Boston : G.K. Hall, 1991, entry 1846-4.
  </description>
  <subject>Bowery Boys (Gang)--1840-1850.</subject>
  <subject>Irish Americans--1840-1850.</subject>
  <coverage>Oregon--1840-1850.</coverage>
  <relation>
    American cartoon print filing series (Library of Congress)
  </relation>
  <identifier>http://hdl.loc.gov/loc.pnp/cph.3a41521</identifier>
  <identifier>http://hdl.loc.gov/loc.pnp/cph.3g05745</identifier>
  <rights>No known restrictions on publication.</rights>
</srw_dc:dc>
```

Fonte: Library of Congress (2015a).

O DC está entre os padrões mais usados por comunidades internacionais para descoberta de recursos no domínio *Web*. É representado por meio de diversas sintaxes como, por exemplo, codificado em HTML, ou em *Resource Description Framework* (RDF)⁵¹ usando XML. Embora o DC tenha uma admirável capacidade descritiva para prestar o acesso e seja um expoente de interoperabilidade na *Web*, este esquema é muito simples e menos eficaz quanto à preservação digital e à efetuação do papel dos metadados administrativos e estruturais de documentar, por exemplo, o histórico de estratégias de preservação adotadas no recurso digital e os componentes de *hardware* e *software* (CAMPOS; SARAMAGO, 2007; HILLMANN, 2005a).

4.3.2 Padrão MODS

Criado pela *Library of Congress* (LC) e disponível em 2002, o esquema de metadados bibliográficos MODS⁵² pode ser aplicado para objetos de bibliotecas digitais. Expresso em XML, o MODS contém um subconjunto de campos MARC 21 e utiliza elementos textuais, ao invés de códigos numéricos, possibilitando uma fácil compreensão. Dentre algumas das características do padrão, estão (GUENTHER, 2003; LIBRARY OF CONGRESS, 2014b):

- Os elementos MODS herdam a semântica dos elementos MARC, permitindo um alto grau de compatibilidade do padrão com os registros MARC em sistemas de bibliotecas.
- O MODS é mais simples que o formato MARC completo e propicia uma descrição mais rica dos recursos quando comparado ao padrão de metadados DC qualificado.
- A descrição de recurso original, a representação de um registro MARC simplificado, ou metadados empacotados com um recurso eletrônico, exemplificam o uso do MODS.
- Devido ser um esquema XML, o MODS pode vir a integrar, combinar ou ser usado com outros padrões baseados nesta sintaxe ou linguagem de codificação, como o METS.
- No MODS temos o reagrupamento de certos elementos MARC e, em alguns casos, o que está em vários elementos MARC são trazidos dentro de um único elemento MODS.

⁵¹ Disponível em: <<http://www.w3.org/RDF/>>. Acesso em: 27 jun. 2015.

⁵² O MODS é mantido pelo *Network Development and MARC Standards Office* da LC. No site oficial do padrão há descrições de diversos projetos que aplicam o MODS, tais como: o *Australian National Bibliographic Database Metadata Project* e o *MusicAustralia*, da Biblioteca Nacional da Austrália; o *Library of Congress Web Archives* (LCWA), o *American Memory and Global Gateway Legacy Conversion Project* e a *Performing Arts Encyclopedia*, da LC; além da *Oxford Digital Library* (ODL), da Universidade de Oxford no Reino Unido. Disponível em: <<http://www.loc.gov/standards/mods/registry.php>>. Acesso em: 19 out. 2015.

- A conversão de um registro MARC original para MODS e depois o retorno para MARC, pode resultar em perda de codificação ou de dados. Logo, o MARC XML⁵³, que concede a expressão de um registro MARC completo em XML, deve ser usado antes para garantir uma conversão sem perdas.

Atualmente na versão 3.6, o padrão MODS apresenta um conjunto de vinte elementos de metadados descritivos principais (LIBRARY OF CONGRESS, 2014a):

1. Informação de Título – as denominações atribuídas ao recurso. Inclui o título e o subtítulo do recurso, além da numeração e a nomeação de uma parte ou seção de um recurso. Preza-se pelo uso das regras contidas em um padrão de conteúdo, como o AACR2.
2. Nome – o nome de uma pessoa, família, organização, ou evento (conferências etc.) ligado ao recurso. Pode conter os nomes dos criadores/autores do conteúdo intelectual original ou derivado do recurso, dos tradutores e/ou editores sabidos etc.
3. Tipo de Recurso – as características e o tipo de conteúdo geral do recurso original, isto é, material catalográfico ou de natureza textual, imagem fixa ou em movimento, gravação de som musical, objeto tridimensional, *software* ou multimídia, material misto etc.
4. Gênero – uma classe mais específica que caracteriza um certo estilo, forma, ou conteúdo, tal como artístico, musical, composição literária etc. Indica-se o uso de valores por meio de listas ou vocabulários controlados de termos, como MARC *Genre Term*⁵⁴ e LCSH.
5. Informação de Origem – a procedência do recurso. Inclui, por exemplo, o editor, o local de sua origem/publicação, a edição e a frequência de publicação, a data de sua alteração, captura, criação, emissão, publicação, ou que está protegido por direitos autorais. Indica-se a adoção da norma ISO 3166⁵⁵ e das regras descritivas do AACR2.
6. Língua – o idioma do conteúdo do recurso. Indica-se utilizar valores de um vocabulário controlado, tal como o RFC 4646, o RFC 5646⁵⁶ e a norma ISO 639.
7. Descrição Física – as características físicas do recurso. Inclui, por exemplo, o meio ou a forma física do material, a origem do arquivo digital (nascido digital etc.), as dimensões (quantidade, tamanho, duração etc.), e o formato eletrônico de arquivo. Indica-se o uso de um vocabulário controlado, como a lista de *Internet Media Types* ou *MIME Types*.
8. Resumo – uma breve síntese do conteúdo do recurso. Pode conter, por exemplo, o escopo, a temática, um comentário, um relato ou uma análise do conteúdo do recurso.

⁵³ Disponível em: <<http://www.loc.gov/standards/marcxml/>>. Acesso em: 23 out. 2015.

⁵⁴ Disponível em: <<http://www.loc.gov/standards/valuelist/marcgt.html>>. Acesso em: 15 out. 2015.

⁵⁵ Disponível em: <http://www.iso.org/iso/country_codes/country_codes>. Acesso em: 20 out. 2015.

⁵⁶ Disponível em: <<http://www.ietf.org/rfc/rfc5646.txt>>. Acesso em: 15 out. 2015.

9. Índice – uma descrição dos conteúdos do recurso. Por exemplo, a indicação do conjunto de capítulos incluídos num livro ou em um documento textual.
10. Público Alvo – o nível e interesse intelectual do público para o qual o recurso se destina (adulto, adolescente, jardim de infância, historiador etc.). Indica-se usar valores de uma lista controlada de termos, como MARC *Target Audience Term*⁵⁷, ou texto livre.
11. Nota – informação textual geral sobre o recurso. Pode incluir, por exemplo, uma menção de responsabilidade (escrito por etc.) e um apontamento histórico.
12. Assunto – o tópico ou a temática do conteúdo do recurso. Inclui, por exemplo, o assunto representado via etiquetas (*tags*), o âmbito espaço-temporal, e um título ou nome utilizado como um assunto. Indica-se usar o perfil W3C-DTF da norma ISO 8601 e valores a partir de um vocabulário controlado, como o *Getty TGN* e os esquemas de classificação.
13. Classificação – o número de chamada ou de classificação do assunto do recurso mediante sistemas de classificação formal, como a LCC, a DDC, a UDC e a NLM *Classification*.
14. Item Relacionado – os outros recursos associados ao recurso que está sendo descrito. Pode incluir, por exemplo, informação relativa a predecessores ou sucessores, formas originais, partes constituintes, referências/citações, outras versões ou formatos etc.
15. Identificador – um código ou número padrão único que identifica precisamente o recurso, tal como o DOI, o ISBN, o URN, o SICI e outros identificadores persistentes.
16. Localização – a instituição ou repositório que mantém o recurso e a localização eletrônica na forma de um URL onde o recurso está disponível.
17. Condição de Acesso – as restrições ou a falta destas sobre o acesso, o uso e a reprodução do recurso, como os direitos de propriedade intelectual, as limitações de distribuição etc.
18. Parte – as partes físicas ou estruturais do recurso. Contém, por exemplo, a numeração de um tipo de parte do recurso (ou seja, volume, edição, capítulo, seção, parágrafo etc.) e as unidades de medição que integram a parte, como página, minuto etc.
19. Extensão – informação adicional não abarcada pelo MODS. Permite incluir elementos de outros esquemas para atender demandas locais de descrição e recuperação de recursos.
20. Informação de Registro – informação requerida para gerenciar e interpretar o registro de metadados. Pode conter a data em que o registro original MODS foi criado primeiramente e modificado pela última vez, a origem ou proveniência do registro (gerado por máquina ou não etc.), o idioma do registro, as regras usadas para o conteúdo da descrição (isto é, AACR2, RDA etc.) e a organização que criou ou alterou o registro original.

⁵⁷ Disponível em: <<http://www.loc.gov/standards/valuelist/marctarget.html>>. Acesso em: 15 out. 2015.

A Figura 5 apresenta a estrutura de um documento no formato de metadados MODS expresso em linguagem XML, a seguir.

FIGURA 5 – Exemplo de um documento MODS em XML.

```

<mods xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns="http://www.loc.gov/mods/v3" version="3.5" xsi:schemaLocation="http://www.loc.gov/mods/v3
      http://www.loc.gov/standards/mods/v3/mods-3-5.xsd">
    <titleInfo>
      <title>Models, Fantasies and Phantoms of Transition</title>
    </titleInfo>
    <name type="personal">
      <namePart type="given">Ash</namePart>
      <namePart type="family">Amin</namePart>
      <role>
        <roleTerm type="text">author</roleTerm>
      </role>
    </name>
    <typeOfResource>text</typeOfResource>
    <relatedItem type="host">
      <titleInfo>
        <title>Post-Fordism</title>
        <subTitle>A Reader</subTitle>
      </titleInfo>
      <name type="personal">
        <namePart type="given">Ash</namePart>
        <namePart type="family">Amin</namePart>
        <role>
          <roleTerm type="text">editor</roleTerm>
        </role>
      </name>
      <originInfo eventType="publication">
        <dateIssued>1994</dateIssued>
        <publisher>Blackwell Publishers</publisher>
        <place>
          <placeTerm type="text">Oxford</placeTerm>
        </place>
      </originInfo>
      <part>
        <extent unit="page">
          <start>23</start>
          <end>45</end>
        </extent>
      </part>
    </relatedItem>
    <identifier>Amin1994a</identifier>
  </mods>

```

Fonte: Library of Congress (2015d).

O esquema MODS fornece informações bibliográficas importantes que apoiam outros padrões expressos em XML, como o METS e o PREMIS. Sob o enfoque da preservação digital, três elementos MODS merecem atenção: Informação de Origem, Item Relacionado e Condição de Acesso. Estes elementos registram informações que são úteis para auxiliar os metadados de preservação na corroboração da proveniência, autenticidade e integridade dos objetos digitais, bem como no reconhecimento dos direitos associados ao recurso que podem intervir, sobretudo, na realização das ações de preservação, acesso e utilização dos seus respectivos conteúdos.

4.3.3 Padrão EAD

O padrão EAD tem seu início num projeto da biblioteca da Universidade da Califórnia, Berkeley, em 1993. Este projeto investigou a oportunidade e a viabilidade de desenvolver um padrão de codificação não-proprietário para instrumentos de pesquisa legíveis por computador, tais como índices, registros, inventários, ou guias elaborados por arquivos, bibliotecas, museus e repositórios de manuscritos, que proporcionam informações acerca de determinadas coleções de materiais arquivísticos. Assim sendo, o padrão deveria cumprir alguns critérios (LIBRARY OF CONGRESS, 2013):

- a) Expressar informações descritivas de instrumentos de pesquisa arquivísticos, de maneira inter-relacionada e extensiva.
- b) Conservar as relações hierárquicas existentes entre os níveis de descrição; e representar informações descritivas herdadas por um nível hierárquico a partir de outro.
- c) Possibilitar mover dentro de uma estrutura hierárquica informacional; além de apoiar a indexação e a recuperação de específicos elementos.

A princípio o EAD foi construído em linguagem SGML, tendo a participação da *Society of American Archivists* (SAA) e de uma equipe de vários especialistas. Posteriormente, o EAD foi restrukturado devido à demanda de uma maior compatibilidade com as normas de descrição arquivística, como a *General International Standard Archival Description* (ISAD(G)), e com a linguagem XML que, por exemplo, não necessita de certos softwares para o acesso ao EAD em SGML. Logo, a versões 1.0 e 2002 do EAD em XML foram lançadas, nesta ordem, em 1998 e 2002. Hoje, o EAD é mantido pela LC junto com a SAA, e recentemente conteúdos da nova versão EAD3 foram publicados em seu site oficial (LIBRARY OF CONGRESS, 2013, 2015b).

Conforme Andrade (2007), Barbedo et al. (2007), Gama e Ferneda (2010) e *Society of American Archivists* (2002), na versão de 2002 um documento EAD contém três partes:

- Cabeçalho EAD (<eadheader>) – informações relativas ao documento EAD em si. Pode conter, por exemplo, uma identificação única para o documento; informação bibliográfica sobre o documento, incluindo autor, título, subtítulo, e patrocinador, assim como a edição, editor, série, e notas relacionadas codificadas separadamente; informação sobre o idioma, principal e secundário, utilizado no documento, bem como referente a criação da versão codificada do documento, incluindo o agente, o local e a data de codificação; informação acerca de quaisquer alterações ou revisões feitas ao documento codificado.
- Conteúdo Prefacial (<frontmatter>) – informações que remetem à criação, publicação, ou uso do documento EAD. Pode conter, por exemplo, seções de texto, incluindo uma página

de rosto, prefácio, dedicatória, agradecimentos, e instruções para usar um instrumento de pesquisa; além de informações bibliográficas a partir do Cabeçalho EAD.

- Descrição Arquivística (<archdesc>) – informações pertinentes à descrição do acervo e de seus materiais constituintes. Pode conter, por exemplo, informações sobre os materiais arquivísticos descritos, incluindo as condições para o seu acesso; subdivisão e a forma de sua organização; origem e as circunstâncias de seu recebimento; cópias, e a existência, a localização e disponibilidade dos originais; condições para o seu uso ou reprodução, após a concessão do acesso; as características físicas e os requisitos técnicos; histórico de sua custódia/propriedade; os itens relacionados por proveniência, mas que foram fisicamente separados ou removidos; escopo e a abrangência temática do seu conteúdo.

A Figura 6 apresenta a estrutura de um documento no formato de metadados EAD expresso em linguagem XML, a seguir.

FIGURA 6 – Exemplo de um documento EAD em XML.

```

<ead xmlns="urn:isbn:1-931666-22-9" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="urn:isbn:1-931666-22-9 http://www.loc.gov/ead/ead.xsd" id="mferd347e3">
  <eadheader id="mferd347e4" repositoryencoding="iso15511" relatedencoding="MARC21" countryencoding="iso3166-1" scriptencoding="iso15924" dateencoding="iso8601" langencoding="iso639-2b">...</eadheader>
  <archdesc level="collection" relatedencoding="MARC21" type="register">
    <did id="mferd847e47">
      <head>Collection Summary</head>
      <unittitle label="Title" encodinganalog="245$a">
        Thomas Bailey Aldrich Papers
        <unitdate label="Span Dates" type="inclusive" normal="1873/1901" encodinganalog="245$f" era="ce" calendar="gregorian">1873-1901</unitdate>
      </unittitle>
      <unitid label="ID No." encodinganalog="590" countrycode="US" repositorycode="US-DLC">MSS84264</unitid>
      <origination label="Creator">
        <persname encodinganalog="100" source="lcnav">Aldrich, Thomas Bailey, 1836-1907</persname>
      </origination>
      <physdesc label="Extent">...</physdesc>
      <langmaterial label="Language">
        Collection material in
        <language encodinganalog="041" langcode="eng">English</language>
      </langmaterial>
      <repository label="Location" encodinganalog="852">
        <corpname>...</corpname>
        <address>
          <addressline>Washington, D.C.</addressline>
        </address>
      </repository>
      <abstract label="Summary" encodinganalog="520$a">
        Author and editor. Outgoing correspondence of Aldrich and his wife addressed primarily to Edward Howard House relating to personal matters. Also includes a volume containing a holograph and typescript copy of "Monody on the Death of Wendell Phillips."
      </abstract>
    </did>
    <!-- ead:controlaccess template -->
    <!-- ead:controlaccess template: otherwise -->
    <controlaccess id="mferd847e88" altrender=":::F005=^20090615113748.0^">...</controlaccess>
    <descgrp id="mferd847e110" type="admininfo">...</descgrp>
    <scopecontent id="mferd847e138" encodinganalog="520">
      <head>Scope and Content Note</head>
      <p>
        The papers of Thomas Bailey Aldrich (1836-1907) span the years 1873-1901 and consist of correspondence and a bound volume containing both a holograph and typescript of "Monody on the Death of Wendell Phillips.". The correspondence contains letters sent by Aldrich and his wife concerning personal and social matters. Most of the correspondence is to Aldrich's friend, Edward Howard House.
      </p>
    </scopecontent>
    <arrangement id="mferd847e143" encodinganalog="351$a">...</arrangement>
    <otherfindaid encodinganalog="555$a" id="lccnURI">...</otherfindaid>
    <!-- $otherfindaidTestCheck: yes -->
    <dsc id="mferd847e148" type="combined">
      <head>Container List</head>
      <thead valign="bottom">...</thead>
      <c01 id="mferd847e157" level="file">
        <did>
          <container type="box">1</container>
          <unittitle encodinganalog="245$a">
            Correspondence,
            <unitdate encodinganalog="245$f" era="ce" calendar="gregorian">1873-1901</unitdate>
          </unittitle>
          <physdesc>
            <extent encodinganalog="300$a">(2 folders)</extent>
          </physdesc>
        </did>
      </c01>
      <c01 id="mferd847e168" level="file">
        <did>
          <container type="box">1</container>
          <unittitle encodinganalog="245$a">
            "Monody on the Death of Wendell Phillips," holograph and typescript,
            <unitdate encodinganalog="245$f" era="ce" calendar="gregorian">1884</unitdate>
          </unittitle>
        </did>
      </c01>
    </dsc>
  </archdesc>
</ead>

```

Fonte: Library of Congress (2014d).

Na preservação digital o uso do EAD é cabível pois, para Barbedo, Corujo e Sant'Ana (2011), trata-se de um esquema completo e que atende as normas de descrição arquivística. Os padrões de metadados descritivos, tal como o DC, o MODS e o próprio EAD, apesar de serem

mais propícios à descoberta e recuperação de recursos ao invés da preservação de longo prazo, são esquemas importantes para o registro de algumas informações de apoio aos metadados de preservação, no tocante a proveniência, autenticidade, integridade, contexto técnico, direitos, e a custódia/propriedade dos recursos armazenados num repositório de preservação digital.

4.3.4 Padrão ANSI/NISO Z39.87

O padrão ANSI/NISO Z39.87 tem sua origem no “*Image Metadata Workshop*”, evento patrocinado pela NISO, CLIR e RLG, em abril de 1999, que se propôs a promover um trabalho cooperativo para determinar um conjunto de elementos de metadados para documentar atributos técnicos de imagens fixas digitais. Na norma ANSI/NISO Z39.87-2006 (R2011) os elementos visam ser harmônicos com os metadados de preservação do PREMIS, sendo que um esquema em XML, aludido como NISO *Metadata for Images in XML (MIX)*⁵⁸, está em construção para intercambiar e/ou armazenar os elementos de metadados técnicos especificados na ANSI/NISO Z39.87-2006 (NATIONAL INFORMATION STANDARDS ORGANIZATION, c2011).

Através de *National Information Standards Organization* (c2011) a versão atual de 2011 agrupa os elementos do padrão ANSI/NISO Z39.87 em cinco principais seções, a saber:

1. Informação Básica do Objeto Digital – esta seção inclui elementos que se aplicam a todos os objetos digitais, não somente aos arquivos de imagens digitais. Sendo assim, este tipo de informação pode ser considerada como metadados de preservação e possibilita algum cruzamento entre o Z39.87 e os outros padrões, tal como o PREMIS.
 - ObjectIdentifier, FormatDesignation e fileSize – a identificação única, o formato (por exemplo, TIFF) e o tamanho em *bytes* do objeto digital.
 - FormatRegistry – uma referência à outras informações detalhadas do formato.
 - byteOrder – a ordem de *byte* onde os números de vários *bytes* são armazenados.
 - Compression – o nível e o esquema de compressão de dados utilizado no objeto digital ou arquivo de imagem que está sendo descrito.
 - Fixity – contém informações para verificar se um arquivo mudou ou foi alterado de uma forma não documentada ou não autorizada.

⁵⁸ O MIX, atualmente na versão 1.0, é mantido pelo *Network Development and MARC Standards Office* da LC, e tem sido elaborado em parceria com o *NISO Technical Metadata for Digital Still Images Standards Committee* e outros especialistas interessados. Expresso em linguagem XML o esquema MIX é voltado para a codificação de um conjunto de elementos de metadados técnicos que são requeridos para o gerenciamento de coleções de imagens digitais. Disponível em: <<http://www.loc.gov/standards/mix/>>. Acesso em: 5 jul. 2015.

2. Informação Básica da Imagem – esta seção reúne elementos vitais para a reconstrução do objeto digital como uma imagem visível em monitores. Não há presunção da exatidão da imagem exibida, apenas que uma imagem de visual razoável pode ser reconstruída.
 - BasicImageCharacteristics – as especificações alusivas ao tamanho da imagem digital, as informações para a correta interpretação dos valores de *pixel* etc.
 - SpecialFormatCharacteristics – as informações sobre determinados formatos de arquivo de imagem como, por exemplo, arquivos *Joint Photographic Experts Group* (JPEG) 2000, *MrSID* e *DjVu*.
3. Metadados de Captura da Imagem – esta seção abrange elementos referentes a metadados técnicos ou metadados administrativos. Algumas das informações podem ser obtidas com base no próprio arquivo de imagem, enquanto que outras informações terão de ser dadas pela instituição gestora do processo de captura ou criação das imagens digitais.
 - SourceInformation – o objeto imagético. Por exemplo, inclui o tipo de material analógico digitalizado para criar uma imagem digital, como uma fotografia, um microfilme, uma escultura, um livro ou documento textual etc.
 - GeneralCaptureInformation – as informações sobre a data e o horário de criação da imagem, o produtor da imagem (a nível organizacional) e a classificação do dispositivo (por exemplo, câmera digital e *scanner* de transmissão) usado para a criação dos dados da imagem digital.
 - ScannerCapture – as configurações do *scanner* usado na criação ou digitalização da imagem, como o modelo de *scanner*, o fabricante, o *software* de captura etc.
 - DigitalCameraCapture – as propriedades da câmera digital usada para capturar a imagem, como o modelo de câmera, o fabricante, o sensor de imagem etc.
 - orientation – a orientação de exibição correta da imagem em um dispositivo de visualização, no que diz respeito a sua dimensão horizontal e vertical.
 - methodology – a metodologia e fundamentação para a digitalização de um objeto digital ou de uma coleção.
4. Metadados de Avaliação da Imagem – esta seção objetiva manter os atributos da imagem relativos à sua qualidade. Os elementos servem como indicadores para avaliar a exatidão da imagem seja no contexto atual de uso como também futuramente após a aplicação de técnicas ou métodos de preservação, sobretudo, aquelas pautadas na migração.
 - SpatialMetrics e ImageColorEncoding – as medidas quantitativas de alto nível de desempenho de imagem, como resolução e codificação de cor numa imagem.

- *TargetData* – os parâmetros de referência que são fundamentais para correções ou reconstruções fiéis do documento de origem.
5. Histórico de Mudança – esta seção atende o papel de documentar os processos aplicados a dados da imagem (por exemplo, qualquer edição e transformação do arquivo) ao longo do ciclo de uma imagem. Os elementos não se destinam a serem utilizados para reverter operações de edição e para autenticar uma imagem digital.
- *ImageProcessing* – o sumário das operações de transformação (migração) ou de edição realizadas na imagem, que pode ser usado tanto para uma avaliação futura de qualidade dos dados da imagem como para entender a proveniência de uma determinada geração de um arquivo de imagem.
 - *Previous Image Metadata* – contém metadados técnicos de gerações ou versões anteriores da imagem, se porventura das operações ou dos processos executados no arquivo de imagem tem como resultado uma nova versão.

A Figura 7 apresenta a estrutura de um documento no formato de metadados MIX expresso em linguagem XML, a seguir.

FIGURA 7 – Exemplo de um documento MIX em XML.

```

<mix xmlns="http://www.loc.gov/mix/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/mix/ http://www.loc.gov/mix/mix.xsd">
  <BasicImageParameters>
    <Format>
      <MIMEType>image/tiff</MIMEType>
      <ByteOrder>big-endian</ByteOrder>
      <Compression>...</Compression>
      <PhotometricInterpretation>...</PhotometricInterpretation>
      <Segments>...</Segments>
      <PlanarConfiguration>1</PlanarConfiguration>
    </Format>
    <File>
      <ImageIdentifier imageIdentifierLocation="system of identifier">unique persistent identifier</ImageIdentifier>
      <FileSize>1001000</FileSize>
      <Checksum>
        <ChecksumMethod>checksum</ChecksumMethod>
        <ChecksumValue>2224446888</ChecksumValue>
      </Checksum>
      <Orientation>3</Orientation>
      <DisplayOrientation>0</DisplayOrientation>
      <TargetedDisplayAR>...</TargetedDisplayAR>
    </File>
    <PreferredPresentation>recommended target viewing device or application</PreferredPresentation>
  </BasicImageParameters>
  <ImageCreation>
    <SourceType>daggerotype</SourceType>
    <SourceID>LCCN</SourceID>
    <ImageProducer>JJT, Inc.</ImageProducer>
    <Host>
      <HostComputer>Computer in use at time of image creation</HostComputer>
      <OperatingSystem>Windows</OperatingSystem>
      <OSVersion>2000</OSVersion>
    </Host>
    <DeviceSource>digital still scanner</DeviceSource>
    <ScanningSystemCapture>
      <ScanningSystemHardware>
        <ScannerManufacturer>Scitex</ScannerManufacturer>
        <ScannerModel>
          <ScannermodelName>Leaf Volare</ScannermodelName>
          <ScannerModelNumber>100SX</ScannerModelNumber>
          <ScannerModelSerialNo>1122334455</ScannerModelSerialNo>
        </ScannerModel>
      </ScanningSystemHardware>
      <ScanningSystemSoftware>
        <ScanningSoftware>Leaf</ScanningSoftware>
        <ScanningSoftwareVersionNo>4.0</ScanningSoftwareVersionNo>
      </ScanningSystemSoftware>
      <ScannerCaptureSettings>...</ScannerCaptureSettings>
    </ScanningSystemCapture>
    <DigitalCameraCapture>
      <DigitalCameraManufacturer>PhaseOne</DigitalCameraManufacturer>
      <DigitalCameraModel>H_20</DigitalCameraModel>
    </DigitalCameraCapture>
    <CameraCaptureSettings>...</CameraCaptureSettings>
    <Sensor>MonochromeArea</Sensor>
    <DateTimeCreated>2002-04-17T00:00:00.001</DateTimeCreated>
    <Methodology>...</Methodology>
  </ImageCreation>
  <ImagingPerformanceAssessment>...</ImagingPerformanceAssessment>
  <ChangeHistory>
    <ImageProcessing>
      <DateTimeProcessed>2001-10-06T00:00:00.001</DateTimeProcessed>
      <SourceData>http://path/to/file</SourceData>
      <ProcessingAgency>JJT, Inc.</ProcessingAgency>
      <ProcessingSoftware>
        <ProcessingSoftwareName>Adobe Photoshop</ProcessingSoftwareName>
        <ProcessingSoftwareVersion>version 5.5</ProcessingSoftwareVersion>
      </ProcessingSoftware>
      <ProcessingActions>ICC profile added</ProcessingActions>
    </ImageProcessing>
    <PreviousImageMetadata/>
  </ChangeHistory>
</mix>
```

Fonte: Cundiff (2015).

O ANSI/NISO Z39.87 não tem como escopo metadados para o registro da proveniência, autenticidade, integridade, ou dos direitos e propriedade intelectual de imagens. Na preservação digital o uso do padrão pode ser contextualizado através de dois projetos notáveis: em Portugal,

o RODA da DGLAB e da UMinho; e nos Estados Unidos, a UHDL. Para a guarda de metadados de preservação, estruturais e descritivos o RODA utiliza, nessa ordem, o PREMIS, o METS e o EAD, tendo o Z39.87 de apoio ao PREMIS na guarda de metadados técnicos de imagens. Por outro lado, a UHDL preserva registros METS gerados por softwares (*JHOVE* e *7train*) através de metadados descritivos DC e de metadados técnicos MIX (BARBEDO et al., 2007; CHEN; REILLY, 2011; NATIONAL INFORMATION STANDARDS ORGANIZATION, c2011).

4.3.5 Padrão METS

O METS teve como antecedente o projeto *Making of America II* (MOA2), que forneceu um esquema de codificação para metadados para obras textuais e baseadas em imagens. Usando a linguagem XML, o METS⁵⁹ concede um padrão para codificação de metadados – descritivos, administrativos e estruturais – tanto para a administração de objetos de bibliotecas digitais num repositório como também para o intercâmbio destes objetos digitais entre repositórios (ou entre repositórios e seus usuários). Dependendo do seu uso, um documento METS pode ser utilizado para a estruturação dos pacotes de informação PSI, PAI ou PDI, no contexto de gerência das informações digitais no modelo de referência SAAI (LIBRARY OF CONGRESS, 2011).

Segundo Barbedo et al. (2007), Campos e Saramago (2007), Grácio (2012), *Library of Congress* (2011) e Sayão (2010) um documento METS compõe sete principais seções:

1. Cabeçalho METS – contém metadados descrevendo o documento METS em si. Pode incluir, por exemplo, função do agente ou parte em relação ao registro (criador do registro, arquivista responsável pelo material original etc.), data e horário de criação/atualização do registro, identificadores alternativos (por exemplo, número de registro OCLC, *Library of Congress Control Number* etc.) e o estado do processamento do registro METS.
2. Metadados Descritivos – pode ter apontadores para metadados descritivos que são externos ao documento METS (por exemplo, um registro MARC num catálogo *online* ou um registro EAD mantido num servidor *Web*), ou metadados descritivos incorporados internamente, ou ambos. Várias instâncias de metadados descritivos, internas e externas, podem coexistir num

⁵⁹ O padrão METS, atualmente na versão 1.10, é mantido pelo *Network Development and MARC Standards Office* da LC, e tem sido desenvolvido como uma iniciativa da DLF. No site oficial do METS temos descrições de vários projetos que utilizam o esquema, como: a Biblioteca Nacional Digital (BND) da Biblioteca Nacional de Portugal; a *Biblioteca Virtual del Patrimonio Bibliográfico* do Ministério da Cultura e das Comunidades Autônomas, na Espanha; o *OCLC Digital Archive*; e o *kopal – Co-operative Development of a Long-Term Digital Information Archive* da Biblioteca Nacional da Alemanha junto com a Biblioteca Estadual e Universitária de Gotinga e outras organizações. Disponível em: <<http://www.loc.gov/standards/mets/mets-registry.html>>. Acesso em: 29 jun. 2015.

mesmo documento METS. Inclui, por exemplo, o tipo de identificador de localização (DOI, URN etc.) e qual a forma de metadados está a ser referenciada (DC, MODS, MIX etc.).

3. Metadados Administrativos – inclui informações técnicas sobre a criação e o armazenamento dos arquivos que compõem o objeto digital, informações de direitos e propriedade intelectual (direitos autorais e dados de licenciamento), metadados acerca do material fonte original a partir do qual o objeto foi derivado e informações relativas a proveniência dos arquivos que integram o objeto (ou seja, relações de origem/destino ou de original/derivado entre arquivos e informações de migrações/transformações). Os metadados administrativos, assim como os descritivos, podem ser externos ao documento METS ou codificados internamente.
4. Seção de Arquivos – lista todos os arquivos que constituem as versões eletrônicas do objeto da biblioteca digital. Os elementos <file> podem ser agrupados em elementos <fileGrp>, de maneira a possibilitar a subdivisão e a identificação de arquivos por versão do objeto digital (por exemplo, podem existir elementos <fileGrp> separados para as versões PDF, as versões de texto codificadas em XML, as imagens originais e digitalizadas de arquivo etc.).
5. Mapa Estrutural – é o núcleo do documento METS. Esboça uma estrutura hierárquica para os objetos digitais e liga os elementos desta estrutura aos arquivos com conteúdos, bem como aos metadados concernentes a cada elemento. Indica-se hierarquicamente como os diversos componentes do objeto da biblioteca digital se correlacionam reciprocamente propiciando, desta maneira, que seus elementos constituídos sejam navegados pelos usuários.
6. Ligações Estruturais – possibilita aos criadores METS registrar a existência de hiperligações entre itens na hierarquia esboçada no Mapa Estrutural. Deste modo, esta seção tem um valor especial caso se planeje utilizar o METS para arquivar *sites* e manter um registro da estrutura do hipertexto dos *sites* de forma separada dos arquivos HTML do *site* em si.
7. Comportamento – pode ser utilizada para associar comportamentos executáveis ao conteúdo no objeto METS. Cada comportamento apresenta um elemento de definição de interface que representa uma definição abstrata do conjunto de comportamentos representado por uma seção de comportamento específica. Ademais, cada comportamento contém um elemento de mecanismo que identifica um módulo de código executável capaz de implantar e executar os comportamentos definidos abstratamente pela definição de interface.

A Figura 8 apresenta a estrutura de um documento no formato de metadados METS contendo metadados no formato MODS expresso em linguagem XML, a seguir.

FIGURA 8 – Exemplo de um documento METS contendo metadados MODS em XML.

```

<mets:mets xmlns:mets="http://www.loc.gov/METS/" xmlns:lcc="http://www.loc.gov/mets/profiles"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:vp="http://www.loc.gov/mets/profiles/videoProgram" xmlns:mods="http://www.loc.gov/mods/v3"
  xmlns:rights="http://www.loc.gov/rights/" xmlns:xlink="http://www.w3.org/1999/xlink"
  OBJID="loc.natlib.ihas.200198140" xsi:schemaLocation="http://www.loc.gov/METS/
  http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/mods/v3
  http://www.loc.gov/standards/mods/mods.xsd" PROFILE="lc:simpleVideoStream">
  <mets:metsHdr LASTMODDATE="2014-12-22T13:39:17.407-05:00"/>
  <mets:mdSec ID="MODS1">
    <mets:mdWrap MDTYPE="MODS">
      <mets:xmlData>
        <mods:mods version="3.4">
          <mods:titleInfo>
            <mods:title>Cupid Takes a Holiday</mods:title>
          </mods:titleInfo>
          <mods:physicalDescription>
            <mods:form>videorecording</mods:form>
            <mods:reformattingQuality>access</mods:reformattingQuality>
            <mods:extent>35mm black and white film</mods:extent>
            <mods:digitalOrigin>reformatted digital</mods:digitalOrigin>
          </mods:physicalDescription>
          <mods:targetAudience authority="marctarget">general</mods:targetAudience>
          <mods:note displayLabel="Aspect Ratio">4:3</mods:note>
          <mods:identifier type="local" displayLabel="MODSID">77430</mods:identifier>
          <mods:identifier type="local" displayLabel="IHASDigitalID">200198140</mods:identifier>
          <mods:identifier type="membership">ihas</mods:identifier>
          <mods:identifier type="local" displayLabel="IHASMODSID">77430</mods:identifier>
          <mods:identifier type="membership">kayefine</mods:identifier>
          <mods:identifier type="membership">The Danny Kaye and Sylvia Fine Collection</mods:identifier>
          <mods:relatedItem type="host">
            <mods:titleInfo>
              <mods:title>The Danny Kaye and Sylvia Fine Collection</mods:title>
            </mods:titleInfo>
            <mods:location>
              <mods:url>
                http://lcweb2.loc.gov/diglib/ihas/html/kayefine/kayefine-home.html
              </mods:url>
            </mods:location>
          </mods:relatedItem>
          <mods:relatedItem type="host">
            <mods:titleInfo>
              <mods:title>Performing Arts Encyclopedia</mods:title>
            </mods:titleInfo>
            <mods:location>
              <mods:url>http://www.loc.gov/performingarts</mods:url>
            </mods:location>
          </mods:relatedItem>
          <mods:location>
            <mods:physicalLocation valueURI="http://id.loc.gov/authorities/names/n78023545">Motion Picture
            Section</mods:physicalLocation>
            <mods:physicalLocation authority="marcorg">DLC</mods:physicalLocation>
          </mods:location>
          <mods:recordInfo>
            <mods:recordContentSource>IHAS</mods:recordContentSource>
            <mods:recordChangeDate encoding="marc">141210</mods:recordChangeDate>
            <mods:recordIdentifier source="IHAS">loc.natlib.ihas.200198140</mods:recordIdentifier>
          </mods:recordInfo>
        </mods:mods>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:mdSec>
  <mets:mdSec/>
  <mets:fileSec>
    <mets:fileGrp USE="MASTER">
      <mets:file ID="Fd32691e38" GROUPID="G1" MIMETYPE="video/mp4" SIZE="124719462">
        <mets:fLocat LOCTYPE="URL"
          xlink:href="http://lcweb2.loc.gov/natlib/ihas/warehouse/kayefine.4/200198140/seg01/0001.mp4"
          LABEL="D71BEB587EE2021CE0438C93F028021C"/>
      </mets:file>
    </mets:fileGrp>
  </mets:fileSec>
  <mets:structMap>
    <mets:div DMDID="MODS1" TYPE="vp:videoProgram">
      <mets:div TYPE="vp:programSegment" DMDID="MODS1">
        <mets:div TYPE="vp:video">
          <mets:fptr FILEID="Fd32691e38"/>
        </mets:div>
      </mets:div>
    </mets:div>
  </mets:structMap>
</mets:mets>

```

Fonte: Library of Congress (2014c).

O padrão METS dispõe um mecanismo flexível para organizar e vincular várias formas de metadados aos objetos digitais num repositório. Permite a expressão das relações complexas entre os metadados; a associação dos objetos a códigos executáveis (*softwares* ou serviços) para interação com seu conteúdo; e a descrição concomitante com padrões de metadados descritivos ou administrativos já em uso por uma instituição. O METS é relevante na preservação digital, pois está escrito em XML, entretanto, a flexibilidade na implementação do METS pode inibir o intercâmbio de registros causando assim problemas de interoperabilidade entre sistemas; e na utilização do METS com o PREMIS pode ocorrer duplicações ou, ainda, uma parcial harmonia entre seus elementos (LAVOIE; GARTNER, 2013; RODRIGUES, 2008; SAYÃO, 2010).

4.3.6 Padrão PREMIS

O PREMIS denota o nome de um grupo de trabalho internacional patrocinado pela RLG e OCLC, no período de 2003 a 2005. Tal grupo de trabalho produziu um relatório final chamado “PREMIS Data Dictionary for Preservation Metadata”, que define um conjunto de metadados básicos, implementáveis e de ampla aplicação, para apoiar a preservação digital em sistemas de repositórios. O Dicionário de Dados PREMIS⁶⁰, atualmente na versão 3.0, propicia orientações para o preenchimento e o gerenciamento dos metadados de preservação, caracterizando-se por (CAPLAN, 2009; SAYÃO, 2010):

- Definir unidades semânticas e não elementos de metadados. As unidades semânticas são porções de informação e os elementos de metadados são formas definidas de representar estas informações num registro de metadados, esquema ou banco de dados. O PREMIS define apenas o que a maioria dos repositórios precisam saber para efetuar suas funções de preservação e o que eles devem ser capazes de exportar para demais sistemas.
- Não ter como escopo certas categorias de metadados, tais como: metadados descritivos; metadados técnicos detalhados de formatos de arquivo (áudio, vídeo etc.); informações detalhadas sobre mídia ou *hardware*; informações atinentes a direitos e permissões que

⁶⁰ O Dicionário de Dados e esquema PREMIS é mantido pelo *Network Development and MARC Standards Office* da LC. No site oficial do padrão temos descrições de múltiplos projetos que utilizam o PREMIS, como: a *Virtual Library of Historical Newspapers*, do Ministério da Cultura da Espanha; o *National Digital Newspaper Program (NDNP)*, da LC; o *Digitaal Magazijn*, da Biblioteca Nacional da Holanda; o *Scalable Preservation and Archiving Repository (SPAR)*, da Biblioteca Nacional da França; o *Digital Data Archive (DDA)*, dos Arquivos Nacionais da Escócia; e o *National Library of Sweden Digital Archive*, da Biblioteca Nacional da Suécia. Disponível em: <<http://www.loc.gov/standards/premis/registry/>>. Acesso em: 17 jul. 2015.

não afetam de modo direto as atividades de preservação; metadados acerca das políticas ou práticas de um determinado repositório; dentre outras.

- Enfocar o repositório e sua gestão, onde os principais usos do PREMIS está para o plano e a avaliação de repositórios de preservação (isto é, a criação e/ou a escolha de *softwares* para implementar o repositório), e a troca de pacotes de informação armazenados entre os repositórios, sobretudo, com o uso do PREMIS em esquema XML.
- Permitir a inclusão e o registro de metadados não abrangidos no escopo PREMIS como, por exemplo, metadados técnicos de imagens digitais definidos na norma Z39.87. Além disto, no âmbito do PREMIS recomenda-se a utilização de valores retirados a partir de vocabulários controlados.

Com base em Barbedo et al. (2007), Campos e Saramago (2007), Caplan (2009) e Sayão (2010), o Dicionário de Dados PREMIS organiza-se de acordo com um modelo de dados que correlaciona cinco tipos de entidades compreendidas nos processos de preservação digital:

1. Entidade Intelectual – alude-se a um conjunto coerente de conteúdos que é tido como um item ou unidade de informação para fins de gestão e descrição, por exemplo, um livro, um mapa, uma imagem, ou um banco de dados. O PREMIS não define metadados para a descrição das entidades intelectuais, visto que há muitos padrões de metadados descritivos para escolher. No RODA, a aplicação desta entidade é assegurada pelo EAD, tendo a norma Z39.87 para guarda de metadados técnicos de imagem digital. Um objeto no sistema de preservação deve ser associado com a entidade intelectual incluindo um identificador da mesma (por exemplo, DOI, ISBN etc.) nos metadados do objeto.
2. Objeto – entende-se como uma unidade de informação discreta em formato digital e que compõe o que é de fato armazenado e gerido no repositório. Pode ter o registro de certas informações, por exemplo, um identificador único para o objeto (tipo e valor); o nome original, o tamanho e o formato do objeto; informação sobre a criação, as propriedades úteis (a serem definidas e mantidas), os inibidores (encriptação, proteção por senha etc.) para o acesso, o uso, ou a migração, e o âmbito do objeto (isto é, o *hardware* e o *software* para processá-lo e as dependências em outros objetos); informação de fixidez; relações com outros objetos e tipos de entidades. O PREMIS define três tipos de objetos:
 - a. Arquivo – uma sequência de *bytes* reconhecida por um sistema operacional, ou seja, um arquivo de computador, como um arquivo PDF ou JPEG. Pode conter permissões de acesso e dados sobre o tamanho e a data da última alteração.

- b. Fluxo de Bits – um conjunto de dados (*bits*) num determinado arquivo, que tem propriedades comuns para propósitos de preservação e que devem estar contidos em um cabeçalho de arquivo ou numa outra estrutura.
 - c. Representação – um conjunto de arquivos, incluindo metadados estruturais, que são requeridos para processar ou apresentar uma entidade intelectual. A título de exemplo, na preservação de uma página *Web* o objeto de representação permite ao repositório identificar o conjunto de arquivos relacionados (arquivos HTML, imagens JPEG, áudio etc.) e descrever qualquer aspecto da totalidade (a página *Web* no seu todo etc.), que podem ser distintas daquelas de suas partes.
3. Eventos – trata-se da agregação de informações acerca de ações que afetam os objetos no repositório. Pode incluir, por exemplo, um identificador único para o evento (tipo e valor); o tipo de evento (criação, ingestão, migração etc.); a data e a hora de ocorrência do evento; uma descrição detalhada do evento; um resultado codificado do evento e sua descrição detalhada, os objetos e os agentes envolvidos no evento e seus papéis. Assim, um registro preciso e confiável dos eventos é crucial para a manutenção da proveniência e a subsequente demonstração da autenticidade de um objeto.
4. Agentes – remete-se a pessoas, organizações, ou aplicações de *software*, que executam papéis em eventos e declarações de direitos. Pode incluir, por exemplo, um identificador único para o agente (tipo e valor); o nome do agente; e a designação da espécie de agente (pessoa, organização ou *software*). O repositório pode optar por utilizar o identificador de agente para apontar para informações registradas externamente, ou usar um padrão separado para o registro de informações adicionais sobre agentes. A título de exemplo, um agente poderia desempenhar a função de autor e detentor dos direitos de uma obra, de autor (porém não detentor dos direitos), ou de depositante de outros trabalhos.
5. Direitos – refere-se à agregação de informações sobre direitos e permissões úteis para a preservação dos objetos no repositório. Manifesta-se na declaração de direitos as ações que o repositório tem o direito de executar e a base para reivindicar este direito (ou seja, direitos autorais, licença ou estatuto). Pode incluir, por exemplo, um identificador único para a declaração (tipo e valor); a ação que a declaração permite e quaisquer restrições sobre a mesma (por exemplo, um repositório pode manter a versão digitalizada de uma obra de domínio público); o período de tempo no qual a declaração se aplica; o objeto a que a declaração se aplica bem como os agentes envolvidos e suas funções.

A Figura 9 apresenta a estrutura de um documento no formato de metadados METS contendo metadados nos formatos MODS, MIX e PREMIS expressos em linguagem XML, a seguir.

FIGURA 9 – Exemplo de um documento METS contendo metadados MODS, MIX e PREMIS em XML.

```

<mets:mets xmlns:lc="http://www.loc.gov/mets/profiles/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:mets="http://www.loc.gov/METS/"
  xmlns:photo="http://www.loc.gov/mets/profiles/photoObject" xmlns:premis="info:lc/xmlns/premis-v2" PROFILE="lc:photoObject"
  OBJID="loc.natlib.gottlieb.09601" xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd">
  <mets:mdSec ID="MODS">
    <mets:mdWrap MDTYPE="MODS">
      <mets:xmlData>
        <mods:mods xmlns:mods="http://www.loc.gov/mods/v3" xsi:schemaLocation="http://www.loc.gov/mods/v3
          http://www.loc.gov/standards/mods/v3/mods-3-3.xsd" ID="ver1">
          <mods:titleInfo>...</mods:titleInfo>
          <mods:name type="personal">...</mods:name>
          <mods:typeOfResource>still image</mods:typeOfResource>
          <mods:genre authority="marc">photograph</mods:genre>
          <mods:genre authority="gmgpc">Portrait photographs-1930-1950.</mods:genre>
          <mods:genre authority="gmgpc">Film negatives-1930-1950.</mods:genre>
          <mods:originInfo>...</mods:originInfo>
          <mods:physicalDescription>...</mods:physicalDescription>
          <mods:note>Gottlieb Collection Assignment No. 048</mods:note>
          <mods:note>Original negative and contact print not served.</mods:note>
          <mods:note>Purchase William P. Gottlieb</mods:note>
          <mods:note type="version">original negative</mods:note>
          <mods:subject authority="lcsh">...</mods:subject>
          <mods:subject authority="lcsh">...</mods:subject>
          <mods:subject authority="lcsh">...</mods:subject>
          <mods:classification authority="lcc">LC-GLB13- 0960</mods:classification>
          <mods:location>...</mods:location>
          <mods:location>...</mods:location>
          <mods:identifier type="stock number">LC-GLB13-0960 DLC</mods:identifier>
          <mods:accessCondition type="restrictionOnAccess">Original negative and contact print not served.
          </mods:accessCondition>
          <mods:recordInfo>...</mods:recordInfo>
          <mods:relatedItem type="otherVersion" ID="ver2">...</mods:relatedItem>
        </mods:mods>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:mdSec>
  <mets:amdSec>
    <mets:techMD ID="object1">
      <mets:mdWrap MDTYPE="PREMIS:OBJECT">
        <mets:xmlData>
          <premis:object xsi:type="premis:file" xsi:schemaLocation="info:lc/xmlns/premis-v2
            http://www.loc.gov/standards/premis/v2/premis-v2-0.xsd">
            <premis:objectIdentifier>...</premis:objectIdentifier>
            <premis:preservationLevel>...</premis:preservationLevel>
            <premis:significantProperties>...</premis:significantProperties>
            <premis:objectCharacteristics>
              <premis:compositionLevel>0</premis:compositionLevel>
              <premis:fixity>...</premis:fixity>
              <premis:size>28808896</premis:size>
              <premis:format>...</premis:format>
              <premis:creatingApplication>...</premis:creatingApplication>
              <premis:creatingApplication>...</premis:creatingApplication>
              <premis:objectCharacteristicsExtension>
                <mix:mix xmlns:mix="http://www.loc.gov/mix/v20" xsi:schemaLocation="http://www.loc.gov/mix/v20
                  http://www.loc.gov/standards/mix/mix20/mix20.xsd">
                  <mix:BasicDigitalObjectInformation>...</mix:BasicDigitalObjectInformation>
                  <mix:BasicImageInformation>...</mix:BasicImageInformation>
                  <mix:ImageCaptureMetadata>...</mix:ImageCaptureMetadata>
                  <mix:ImageAssessmentMetadata>...</mix:ImageAssessmentMetadata>
                </mix:mix>
              </premis:objectCharacteristicsExtension>
            </premis:objectCharacteristics>
            <premis:originalName>0001h.tif</premis:originalName>
            <premis:storage>...</premis:storage>
            <premis:environment>...</premis:environment>
            <premis:relationship>...</premis:relationship>
            <premis:relationship>...</premis:relationship>
            <premis:relationship>...</premis:relationship>
            <premis:linkingEventIdentifier>...</premis:linkingEventIdentifier>
            <premis:linkingEventIdentifier>...</premis:linkingEventIdentifier>
            <premis:linkingIntellectualEntityIdentifier>...</premis:linkingIntellectualEntityIdentifier>
            <premis:linkingIntellectualEntityIdentifier>...</premis:linkingIntellectualEntityIdentifier>
          </premis:object>
        </mets:xmlData>
      </mets:mdWrap>
    </mets:techMD>
    <mets:techMD ID="object2">...</mets:techMD>
    <mets:digiprovMD ID="event1">...</mets:digiprovMD>
    <mets:digiprovMD ID="event2">...</mets:digiprovMD>
    <mets:digiprovMD ID="event3">...</mets:digiprovMD>
    <mets:digiprovMD ID="event4">...</mets:digiprovMD>
    <mets:digiprovMD ID="event5">...</mets:digiprovMD>
    <mets:digiprovMD ID="event6">...</mets:digiprovMD>
    <mets:digiprovMD ID="agent1">...</mets:digiprovMD>
  </mets:amdSec>
  <mets:fileSec>...</mets:fileSec>
  <mets:structMap>...</mets:structMap>
</mets:mets>

```

Fonte: Library of Congress (2010).

Promovido pela LC, o PREMIS *Maintenance Activity* oferece um esquema XML que corresponde diretamente ao Dicionário de Dados, permitindo a descrição de objetos, eventos, agentes e direitos, assim como a utilização do PREMIS representado em XML para a gerência e o intercâmbio de metadados entre sistemas de preservação. Diversas iniciativas tem buscado definir boas práticas para usar o PREMIS e o METS juntos, pois: não há uma perfeita harmonia entre as seções de informações nos dois esquemas (em especial, para informação sobre agente); e há alguma sobreposição que, por exemplo, pode causar variações na representação dos dados, dificultando a interoperabilidade (CPLAN, 2009; LIBRARY OF CONGRESS, 2015c).

4.4 Análise dos padrões de metadados à luz da preservação digital

Os padrões DC, MODS, EAD, ANSI/NISO Z39.87, METS e PREMIS tem um conjunto de características em comum e algumas particularidades. Este capítulo visa refletir os esquemas supracitados sob o objetivo de sinalizar quais dos seus elementos de metadados principais são úteis no suporte a preservação digital em repositórios institucionais. Não fazem parte do escopo desta análise questões relacionadas à interoperabilidade, à equivalência, à correspondência, ou a harmonização entre os elementos dos padrões indicados. As ponderações feitas respaldaram-se nos requisitos para a preservação digital, no modelo de informação SAAI e nas informações registradas e expressas pelos metadados de preservação (ver os subcapítulos 3.2, 3.4.3 e 4.2).

O Quadro 3 sintetiza os aspectos básicos dos padrões e os elementos de metadados julgados nesta investigação como importantes para a preservação digital.

Quadro 3 – Padrões e elementos de metadados de apoio a preservação digital.

| Padrão | Características | Elementos de metadados úteis para a preservação digital |
|------------------|--|--|
| DC Qualificado | <ul style="list-style-type: none"> - Esquema flexível e extensível de metadados descritivos para descoberta de recursos <i>Web</i>. - Mantido pela DCMI vêm sendo muito usado por comunidades internacionais. - Pode ser representado em XML e propicia interoperabilidade de dados na <i>Web</i>. | <ul style="list-style-type: none"> • Descrição • Data • Formato • Identificador • Fonte • Relação • Cobertura • Direitos • Detentor de Direitos • Proveniência |
| MODS | <ul style="list-style-type: none"> - Esquema XML aplicável aos objetos de bibliotecas digitais e mantido pela LC. - Elementos descritivos herdam a semântica dos elementos MARC 21. - É mais simples que MARC completo e mais rico que o DC Qualificado. | <ul style="list-style-type: none"> • Informação de Origem • Descrição Física • Índice • Item • Relacionado • Identificador • Localização • Parte • Condição de Acesso • Informação de Registro |
| EAD | <ul style="list-style-type: none"> - Esquema XML e uma DTD para codificar instrumentos de pesquisa arquivísticos. - Mantido pela LC junto com a SAA. - Compatível com as normas de descrição arquivística, tal como a ISAD(G). | <ul style="list-style-type: none"> • Cabeçalho EAD • Descrição Arquivística |
| ANSI/NISO Z39.87 | <ul style="list-style-type: none"> - Inclui metadados técnicos para gestão de coleções de imagens fixas digitais. - Implementado e adotado via esquema XML NISO MIX mantido pela LC. - Compreende metadados para preservação e metadados administrativos. | <ul style="list-style-type: none"> • Informação Básica do Objeto Digital • Informação Básica da Imagem • Metadados de Captura da Imagem • Metadados de Avaliação da Imagem • Histórico de Mudança |
| METS | <ul style="list-style-type: none"> - Esquema XML mantido pela LC para gestão de objetos de biblioteca digital. - Pode implantar e padronizar os pacotes PSI, PAI ou PDI no modelo SAAI. - Organiza e expressa relações hierárquicas e hiperligações entre os objetos e suas partes. - Inclui DC, MODS, EAD e MIX na seção de Metadados Descritivos, e pode ter o PREMIS na seção de Metadados Administrativos. | <ul style="list-style-type: none"> • Cabeçalho METS • Mapa Estrutural • Ligações Estruturais • Comportamento |
| PREMIS | <ul style="list-style-type: none"> - Esquema XML de metadados para suporte a preservação digital em repositórios. - É mantido pela LC, focaliza o sistema e sua gerência e define unidades semânticas. - Pode ter apoio do DC, MODS, EAD e MIX na verificação da autenticidade, integridade, procedência ou direitos relativos aos objetos mantidos num repositório digital. | <ul style="list-style-type: none"> • Entidade Intelectual • Objeto • Eventos • Agentes • Direitos |

Fonte: Elaborado pelo autor.

A princípio, é notável frisar que todos estes padrões apontados podem ser representados na linguagem computacional XML. Visto que é um padrão aberto, não proprietário e entendível por aplicações variadas, a linguagem XML apoia as demandas de preservação de longo prazo e propicia o intercâmbio de dados entre sistemas, além de favorecer o uso combinado de

diferentes esquemas, como o METS e o PREMIS com as informações descritivas do DC, MODS, EAD ou MIX. Ademais, temos a DCMI, no caso do DC, e o *Network Development and MARC Standards Office* da LC, para os outros padrões, onde padronizam a descrição e a representação das informações através de vocabulários controlados ou esquemas de codificação internacionalmente reconhecidos.

Utilizar conjuntamente distintos padrões de metadados torna-se plausível em razão dos tipos de recursos para serem descritos e dos vários processos de preservação digital. Não apenas os metadados para preservação são importantes, mas também os metadados descritivos, estruturais, administrativos e técnicos auxiliam o alcance da preservação por longo prazo. Desta maneira, não há como afirmar qual é o único esquema de metadados existente que assegure totalmente a preservação digital e, sim, que diferentes padrões podem trabalhar juntos, de modo a se auto complementarem para o registro eficaz das informações requeridas na gestão da preservação e do acesso utilizável ao longo do tempo de objetos/recursos digitais.

Dentre os elementos de metadados que são relevantes e recorrentes em todos os padrões indicados anteriormente, estão os identificadores. Uma vez que os objetos digitais são frágeis, dinâmicos e volúveis, os identificadores provêm o acesso contínuo e confiável para os objetos como a localização ou referência única e indubitável para estes em todo o seu ciclo de vida. O registro de informações de identificadores locais ou persistentes manifestam os metadados para preservação delineados na Informação de Descrição de Preservação (em especial, referência) do modelo SAAI, além do requisito de localizar e rastrear o objeto ao longo do tempo.

Os padrões DC, MODS e EAD são mais aplicáveis a descrição de recursos para fins de sua descoberta, busca, recuperação ou identificação. Embora a eficácia dos esquemas apontados esteja intrinsecamente atrelada à etapa de acesso do que propriamente para a preservação, alguns dos seus elementos são relevantes para auxiliar os metadados de preservação PREMIS. Assim, as informações descritivas destes padrões apoiam tanto os requisitos de manter a autenticidade, a proveniência e o contexto do objeto digital como os metadados de preservação na Informação de Descrição de Preservação (principalmente, contexto e proveniência) do modelo SAAI.

Sob o enfoque da preservação por longo prazo, determinados elementos de metadados dos esquemas DC Qualificado, MODS e EAD requerem uma maior atenção, tais como:

- Descrição, Data, Formato, Identificador, Fonte, Relação, Cobertura, Direitos, Detentor de Direitos e Proveniência, no esquema DC Qualificado.
- Informação de Origem, Descrição Física, Índice, Item Relacionado, Identificador, Parte, Localização, Condição de Acesso e Informação de Registro, no esquema MODS.
- Cabeçalho EAD e Descrição Arquivística, no esquema EAD.

Embora esteja fora do seu escopo, os metadados técnicos ANSI/NISO Z39.87 ou MIX podem ser usados com o PREMIS na comprovação da procedência, autenticidade e integridade de imagens digitais. Além de metadados sobre os aspectos técnicos de imagens, o ANSI/NISO Z39.87 possui também metadados de preservação e administrativos. Isto posto, todas as seções de metadados Z39.87 devem ser levadas em consideração, visto que amparam os metadados de preservação na Informação de Descrição de Preservação (contexto, proveniência e fixidade) e os requisitos de salvaguardar a autenticidade, a proveniência e o contexto do objeto digital.

Por sua vez, o METS pode atuar na padronização dos pacotes de informação PSI, PAI ou PDI no modelo SAAI, armazenando metadados – descritivos, estruturais, administrativos e de preservação – associados a um objeto digital num sistema. É útil para simplificar a ordem e a gerência das partes constituintes de objetos complexos (Seção de Arquivos) e seus metadados, ou expressar as relações hierárquicas e hiperligações (Mapa Estrutural e Ligações Estruturais) entre os vários arquivos que compõem os objetos ou entre diferentes objetos digitais. Algumas seções num documento METS são notáveis para a preservação, dando suporte aos requisitos de manter o contexto e as funcionalidades do objeto digital e para os metadados de preservação na Informação de Descrição de Preservação (especialmente, contexto), tais como:

- Cabeçalho METS, Mapa Estrutural, Ligações Estruturais e Comportamento.

Em razão do PREMIS refletir diretamente o modelo de informação SAAI e os requisitos de preservação digital, todas as suas entidades ou seções de metadados são significantes. Tendo em vista as discussões realizadas neste trabalho consideramos que, num sistema de repositório baseado no SAAI, o METS pode ser utilizado como um esquema central para a administração dos objetos digitais, englobando os padrões DC, MODS, EAD e MIX na seção de Metadados Descritivos, além do padrão PREMIS na seção de Metadados Administrativos. Todavia, temos que ponderar também a ocorrência de problemas de duplicação, sobreposição e equivalência entre os conjuntos de metadados no uso concomitante destes diferentes esquemas.

5 CONSIDERAÇÕES FINAIS

Os desafios da preservação digital estão associados à natureza dos objetos digitais e seu ambiente, à obsolescência tecnológica e aos direitos de propriedade intelectual. É desejável não apenas salvaguardar o acesso contínuo aos conteúdos, mas também assegurar que os objetos sejam localizados ao longo do tempo somado à preservação do leiaute, das funcionalidades, da origem e do contexto, o qual suportam a comprovação de sua autenticidade e integridade. Estes aspectos refletem basicamente os requisitos para consecução da preservação de longo prazo.

Diferentes estratégias para preservação digital foram observadas nesta investigação. Em razão de sua ampla aplicação em instituições e da não exigência de grandes recursos, a migração demonstra ser útil para a maioria das políticas institucionais de preservação digital. Outro método importante remete à adoção de metadados para preservação, que nos possibilita justificar tanto a realização do presente trabalho como expor uma atual função dos metadados, isto é, apoiar a administração dos processos relativos ao arquivamento e à manutenção do acesso à informação digital a longo prazo, com garantias de autenticidade, de integridade e de confiabilidade.

A estrutura conceitual do Sistema Aberto de Arquivamento de Informação (SAAI) define as atividades e os metadados necessários em um sistema de repositório para a preservação digital. Um ponto interessante no modelo é a estipulação dos usuários de informação (incluindo suas demandas e interesses) e a mediação de submissões de recursos (e seus metadados) de criadores de informação. O recurso depositado precisa ser relevante para os usuários e atender adequadamente as boas práticas para a preservação estabelecidas mediante a criação, implantação e manutenção de políticas.

De um modo geral, as políticas ou programas de preservação digital englobam aspectos administrativos, econômicos, legais, culturais e técnicos. No contexto das IES, um dos grandes desafios de políticas desta natureza está na construção e sustentação de uma cultura institucional ciente dos problemas e, sobretudo, dos benefícios da preservação de longo prazo. Uma vez que esta cultura esteja consolidada, oportunizará a incorporação e a continuidade do programa nos objetivos institucionais, a formação de equipes de trabalho empenhadas e comprometidas e de uma comunidade adepta aos procedimentos, normas e demais exigências necessárias.

Além da definição de um *software* para implementação do repositório institucional, faz-se necessária uma política para preservação digital que contemple a seleção de padrões ou esquemas de metadados descritivos, estruturais, administrativos, técnicos e de preservação. Os padrões de metadados podem propiciar a interoperabilidade dos objetos digitais, oferecer ou

orientar o uso de vocabulários controlados e esquemas de codificação, bem como possibilitar a descrição conjunta ou a incorporação de metadados provindos de outros esquemas XML.

Na pesquisa foram identificados, sistematizados e analisados padrões de metadados para a preservação digital, articulados à Ciência da Informação e a áreas afins. Os metadados descritivos e técnicos do DC, MODS, EAD e ANSI/NISO Z39.87 ou MIX, possuem um papel mais direcionado a amparar o METS e o PREMIS tanto na descoberta e no registro dos aspectos técnicos do objeto digital como na comprovação de sua procedência, autenticidade, contexto e integridade. Englobando metadados descritivos, estruturais, administrativos e de preservação do PREMIS, o METS pode gerir os objetos digitais atuando como um Pacote de Submissão de Informação (PSI), Pacote de Arquivamento de Informação (PAI) ou Pacote de Disseminação de Informação (PDI) num SAAI.

Determinados elementos de metadados dos padrões identificados puderam ser sinalizados nesta investigação como sendo úteis para a preservação digital em repositórios. Nos padrões DC e MODS, por exemplo, os elementos indicados no Quadro 3 incluem algumas das informações previstas nos metadados de preservação PREMIS, como os direitos de propriedade intelectual e seus detentores, a identificação única para o objeto digital, os outros objetos relacionados ao objeto que está sendo descrito e as dependências técnicas do objeto.

Existe uma certa carência de estudos nacionais na Ciência da Informação e áreas afins acerca da indicação de padrões de metadados aplicáveis à preservação digital. Deste modo, os resultados neste trabalho dispõem um respaldo teórico, técnico e sistematizado de esquemas que podem ser utilizados em repositórios institucionais construídos para atender a preservação e o acesso a longo prazo de informações digitais. Os elementos de metadados julgados na pesquisa como úteis na preservação digital, contribuirão para a escolha dos padrões segundo as demandas da instituição, ou para o reconhecimento do conjunto de informações necessárias para garantir a preservação e a administração de conteúdos digitais em C&T num sistema.

Destaca-se que a garantia de preservação digital a longo prazo só será possível com a adoção efetiva de padrões ou esquemas de metadados, pois são eles que determinarão a descrição, a representação, a consistência e a persistência do recurso/objeto digital no ambiente informacional, além de definir a interoperabilidade entre sistemas.

Entre as propostas para futuros trabalhos está em analisar de que maneira os metadados dos esquemas identificados nesta dissertação podem ser melhor harmonizados. Assim, a partir das experiências de organizações relatadas em pesquisas nacionais e internacionais da Ciência da Informação e de áreas afins, o trabalho construiria e recomendaria um planejamento para o uso dos distintos padrões defronte aos problemas de duplicações e sobreposições entre os seus

elementos, o qual dificultam a interoperabilidade. Este plano apoiaria as instituições na tomada de decisões racionais com a utilização consciente de recursos e com perdas não permanentes de registros na aplicação dos esquemas em repositórios voltados à preservação digital.

REFERÊNCIAS

- ALMEIDA, M. B.; CENDÓN, B. V.; SOUZA, R. R. Metodologia para implantação de programas de preservação de documentos digitais a longo prazo. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 17, n. 34, p. 103-130, maio./ago., 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p103/22622>>. Acesso em: 27 mar. 2015.
- ALVES, R. C. V.; SANTOS, P. L. V. A. C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013. 196 p.
- ANDRADE, R. S. Aspectos introdutórios da representação de informação arquivística: a Norma Brasileira de Descrição Arquivística (Nobrade), a Descrição Arquivística Codificada (EAD-DTD) e o projeto Archives Hub. **Ponto de Acesso**, Salvador, v. 1, n. 2, p. 70-100, jul./dez. 2007. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/1589/1813>>. Acesso em: 23 maio 2015.
- ARELLANO, M. Á. M. **Critérios para a preservação digital da informação científica**. 2008. 354 f. Tese (Doutorado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2008. Disponível em: <http://bdtd.bce.unb.br/tedesimplificado/tde_busca/arquivo.php?codArquivo=4547>. Acesso em: 11 jan. 2015.
- ARQUIVO NACIONAL. **AN digital**: política de preservação digital. Versão 1. [Rio de Janeiro], abril 2012. 27 p. Disponível em: <http://www.siga.arquivonacional.gov.br/media/andigital/and_politica_preservao_digital_v1.pdf>. Acesso em: 30 mar. 2015.
- _____. **Programa permanente de preservação e acesso a documentos arquivísticos digitais**: ANDigital. Rio de Janeiro, julho 2010. 22 p. Disponível em: <http://www.siga.arquivonacional.gov.br/media/andigital/an_digital_completo_2010.pdf>. Acesso em: 26 abr. 2015.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR 15472**: sistemas espaciais de dados e informações: modelo de referência para um sistema aberto de arquivamento de informação (SAAI). Rio de Janeiro, 2007. 95 p.
- AULER, D.; BAZZO, W. A. Reflexões para a implementação do movimento CTS no contexto educacional brasileiro. **Ciência & Educação**, Bauru, v. 7, n. 1, p. 1-13, 2001. Disponível em: <<http://www.scielo.br/pdf/ciedu/v7n1/01.pdf>>. Acesso em: 30 jan. 2015.
- BAGGIO, C. C.; FLORES, D. Documentos digitais: preservação e estratégias. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 27, n. 1, p. 11-24, jan./jun. 2013. Disponível em: <<http://www.seer.furg.br/biblos/article/view/2654/2395>>. Acesso em: 9 jan. 2015.
- BARBEDO, F. et al. RODA: repositório de objectos digitais autênticos. **Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas**, Lisboa, n. 9, 2007. Disponível em:

<<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/535/320>>. Acesso em: 13 maio 2014.

BARBEDO, F.; CORUJO, L.; SANT'ANA, M. **Recomendações para a produção de planos de preservação digital**. Lisboa: Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB), nov. 2011. 111 p. Disponível em: <http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2014/02/Recomend_producao_PPD_V2.1.pdf>. Acesso em: 20 fev. 2015.

BARNES, B. **Sobre ciencia**. Barcelona: Labor, 1987.

BODÊ, E. C. **Preservação de documentos digitais**: o papel dos formatos de arquivo. 2008. 153 f. Dissertação (Mestrado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2008. Disponível em: <http://bdtd.bce.unb.br/tedesimplificado/tde_busca/arquivo.php?codArquivo=4529>. Acesso em: 5 set. 2015.

BRASIL. **Lei n. 12.853**, de 14 de agosto de 2013. Altera os arts. 5º, 68, 97, 98, 99 e 100, acrescenta arts. 98-A, 98-B, 98-C, 99-A, 99-B, 100-A, 100-B e 109-A e revoga o art. 94 da Lei nº 9.610, de 19 de fevereiro de 1998, para dispor sobre a gestão coletiva de direitos autorais, e dá outras providências. Brasília, DF, 2013. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2013/Lei/L12853.htm>. Acesso em: 21 mar. 2015.

_____. **Lei n. 9.610**, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Brasília, DF, 1998. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L9610.htm>. Acesso em: 21 mar. 2015.

BROWN, A. **Digital preservation guidance note 2: selecting storage media for long-term preservation**. Issue 2. [London]: The National Archives, aug. 2008. 7 p. Disponível em: <<https://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>>. Acesso em: 13 jan. 2015.

BULLOCK, A. **Preservation of digital information: issues and current status**. [Ottawa]: The National Library of Canada, apr. 22, 1999. Disponível em: <<http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm>>. Acesso em: 15 jan. 2015.

CAMPOS, F. M. G.; SARAMAGO, M. L. Preservação digital de longo prazo em instituições patrimoniais: reutilização e adaptação de metadados. **Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas**, Lisboa, n. 9, p. 1-7, 2007. Disponível em: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/540/330>>. Acesso em: 11 mar. 2015.

CAMPOS, F. M. Informação digital: um novo património a preservar. **Cadernos BAD – Cadernos de Biblioteconomia, Arquivística e Documentação**, Lisboa, n. 2, p. 8-14, 2002. Disponível em: <<http://www.bad.pt/publicacoes/index.php/cadernos/article/view/861/860>>. Acesso em: 8 set. 2015.

CANADIAN DIGITAL PRESERVATION INITIATIVE (CDPI). Disponível em: <<http://www.canadiana.ca/en/canadian-digital-preservation-initiative>>. Acesso em: 20 fev. 2015.

CANADIANA.ORG. Disponível em: <<http://www.canadiana.ca/>>. Acesso em: 20 fev. 2015.

CAPLAN, P. **Understanding PREMIS**. [Washington, DC]: Library of Congress Network Development and MARC Standards Office, feb. 2009. 26 p. Disponível em: <<http://www.loc.gov/standards/premis/understanding-premis.pdf>>. Acesso em: 9 jul. 2015.

CASTELFRANCHI, Y. et al. As opiniões dos brasileiros sobre ciência e tecnologia: o ‘paradoxo’ da relação entre informação e atitudes. **História, Ciências, Saúde – Manguinhos**, Rio de Janeiro, v. 20, supl., p. 1163-1183, nov. 2013. Disponível em: <<http://www.scielo.br/pdf/hcsm/v20s1/0104-5970-hcsm-20-s-1163.pdf>>. Acesso em: 30 jan. 2015.

CEREZO, J. A. L. Ciência, tecnologia e sociedade: o estado da arte na Europa e nos Estados Unidos. In: SANTOS, L. W. et al. (Org.). **Ciência, tecnologia e sociedade**: o desafio da interação. Londrina: IAPAR, 2002. p. 3-38.

CHAN, L. M.; ZENG, M. L. Metadata interoperability and standardization: a study of methodology part i: achieving interoperability at the schema level. **D-Lib Magazine**, v. 12, n. 6, june 2006. Disponível em: <<http://www.dlib.org/dlib/june06/chan/06chan.html>>. Acesso em: 23 maio 2015.

CHEN, M.; REILLY, M. Implementing METS, MIX, and DC for Sustaining Digital Preservation at the University of Houston Libraries. **Journal of Library Metadata**, v. 11, n. 2, p. 83-99, may 2011. Disponível em: <<http://www.tandfonline.com/doi/full/10.1080/19386389.2011.570662#abstract>>. Acesso em: 7 jul. 2015.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Carta para a Preservação do Patrimônio Arquivístico Digital**. Rio de Janeiro: United Nations Educational, Scientific and Cultural Organization (UNESCO), 2005. 22 p. Disponível em: <<http://www.conarq.arquivonacional.gov.br/media/carta.pdf>>. Acesso em: 20 fev. 2015.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). Disponível em: <<http://public.ccsds.org/default.aspx>>. Acesso em: 28 fev. 2015.

CORRÊA, A. M. G. **Preservação digital**: autenticidade e integridade de documentos em bibliotecas digitais de teses e dissertações. 2010. 96 f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2010. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27151/tde-05112010-105831/pt-br.php>>. Acesso em: 5 set. 2015.

COUNCIL ON LIBRARY AND INFORMATION RESOURCES (CLIR). Disponível em: <<http://www.clir.org/>>. Acesso em: 20 fev. 2015.

CUNDIFF, M. **MIX Example Documents**. Example MIX XML Documents. Test MIX document instance (version 0.2). Nov. 2015. Disponível em: <http://www.loc.gov/standards/mix/instances/mix_test.xml>. Acesso em: 10 nov. 2015.

DIGITAL LIBRARY FEDERATION (DLF). Disponível em: <<http://www.diglib.org/>>. Acesso em: 20 fev. 2015.

DIGITAL PRESERVATION COALITION (DPC). Disponível em: <<http://www.dpconline.org/>>. Acesso em: 20 fev. 2015.

DIGITAL PRESERVATION OUTREACH AND EDUCATION (DPOE). Disponível em: <<http://www.digitalpreservation.gov/education/>>. Acesso em: 20 fev. 2015.

DUBLIN CORE METADATA INITIATIVE (DCMI). DCMI Usage Board. **DCMI Metadata Terms**. June 2012. Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 15 jun. 2015.

_____. **History of the Dublin Core Metadata Initiative**. c2015. Disponível em: <<http://dublincore.org/about/history/>>. Acesso em: 17 jun. 2015.

FOOT, M. M.; PEACH, C. **Building a preservation policy**. London: British Library/Preservation Advisory Centre, may 2013. 17 p. Disponível em: <http://www.bl.uk/aboutus/stratpolprog/collectioncare/publications/booklets/building_a_preservation_policy.pdf>. Acesso em: 30 mar. 2015.

FORMENTON, D.; GRACIOSO, L. S.; CASTRO, F. F. Revisitando a preservação digital na perspectiva da ciência da informação: aproximações conceituais. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 13, n. 1, p. 170-191, jan./abr. 2015. Disponível em: <<http://periodicos.sbu.unicamp.br/ojs/index.php/rdbc/article/view/1587/1571>>. Acesso em: 21 set. 2015.

GAMA, F. A.; FERNEDA, E. A Mediação da Informação nos Arquivos Permanentes: serviços de referência arquivística no ambiente digital. **Informação & Informação**, Londrina, v. 15, n. 2, p. 148-169, jul./dez. 2010. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/7352/7005>>. Acesso em: 8 jul. 2015.

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

GONZÁLEZ GARCÍA, M.; LÓPEZ CEREZO, J. A.; LUJÁN, J. L. **Ciencia, tecnología y sociedad**: una introducción al estudio social de la ciencia y la tecnología. Madrid: Tecnos, 1996.

GRÁCIO, J. C. A. **Preservação digital na gestão da informação**: um modelo processual para as instituições de ensino superior. São Paulo, SP: Cultura Acadêmica, 2012. 214 p. Disponível em: <http://www.culturaacademica.com.br/_img/arquivos/Preservacao_digital_na_gestao_da_info_rmacao-WEB_v2.pdf>. Acesso em: 13 maio 2015.

GRÁCIO, J. C. A.; FADEL, B.; VALENTIM, M. L. P. Preservação digital nas instituições de ensino superior: aspectos organizacionais, legais e técnicos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 18, n. 3, p. 111-129, jul./set. 2013. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1612/1196>>. Acesso em: 15 mar. 2015.

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

HARPER, C. A. Dublin Core Metadata Initiative: beyond the element set. **Information Standards Quarterly (ISQ)**, v. 22, n. 1, p. 19-28, winter 2010. Disponível em: <http://www.niso.org/apps/group_public/download.php/3670/FE_DCMI_Harper_isqv22no1.pdf>. Acesso em: 13 jun. 2015.

HESLOP, H.; DAVIS, S.; WILSON, A. **An approach to the preservation of digital records**. Canberra: National Archives of Australia, dec. 2002. 24 p. Disponível em: <http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm16-47161.pdf>. Acesso em: 20 fev. 2015.

HILLMANN, D. **Using Dublin Core**. Nov. 2005a. Disponível em: <<http://dublincore.org/documents/usageguide/>>. Acesso em: 10 jun. 2015.

_____. **Using Dublin Core: dublin core qualifiers**. Nov. 2005b. Disponível em: <<http://dublincore.org/documents/usageguide/qualifiers.shtml>>. Acesso em: 28 jun. 2015.

_____. **Using Dublin Core: the elements**. Nov. 2005c. Disponível em: <<http://dublincore.org/documents/usageguide/elements.shtml>>. Acesso em: 15 jun. 2015.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA (IBICT). Rede Brasileira de Serviços de Preservação Digital – Rede Cariniana. **Dicionário de preservação digital**. [2015?]. Disponível em: <<http://cariniana.ibict.br/index.php/dicionario-de-preservacao-digital>>. Acesso em: 5 set. 2015.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS (IFLA) Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records: final report. **UBCIM Publications - New Series**, vol. 19. München: K. G. Saur, 1998. 136 p. Disponível em: <<http://www.ifla.org/files/assets/cataloguing/frbr/frbr.pdf>>. Acesso em: 17 nov. 2015.

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM (IIPC). Disponível em: <<http://netpreserve.org/>>. Acesso em: 20 fev. 2015.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). Disponível em: <<http://www.iso.org/iso/home.html>>. Acesso em: 28 fev. 2015.

_____. **ISO 15489-1**: information and documentation: records management: part 1: general. Geneva, 2001. 19 p.

INTERNET ARCHIVE. Disponível em: <<https://archive.org/index.php>>. Acesso em: 20 fev. 2015.

LATOUR, B. **Ciencia en acción**. Barcelona: Labor, 1992.

LAVOIE, B.; GARTNER, R. Preservation metadata. 2nd edition. **DPC Technology Watch Report**, v. 13, n. 3, p. 1-36, may 2013. Disponível em: <<http://dx.doi.org/10.7207/twr13-03>>. Acesso em: 28 set. 2015.

_____. _____. **DPC Technology Watch Report**, v. 5, n. 1, p. 1-21, sept. 2005. Disponível em: <<http://www.dpconline.org/docs/reports/dpctw05-01.pdf>>. Acesso em: 29 abr. 2015.

LIBRARY AND ARCHIVES CANADA (LAC). **File format guidelines for preservation and long-term access**. [Ottawa], [2010?]. 93 p. Disponível em: <<http://www.collectionscanada.gc.ca/obj/012018/f2/012018-2200-e.pdf>>. Acesso em: 20 fev. 2015.

LIBRARY OF CONGRESS. **Collection Cartoon Prints, American. War! Or no war**. [2015a?]. Disponível em: <<https://lccn.loc.gov/2008661463/dc>>. Acesso em: 10 nov. 2015.

_____. **Development of the Encoded Archival Description DTD**. Dec. 2013. Disponível em: <<http://www.loc.gov/ead/eaddev.html>>. Acesso em: 27 jul. 2015.

_____. **Encoded Archival Description Official Site**. 2015b. Disponível em: <<http://www.loc.gov/ead/index.html>>. Acesso em: 31 jul. 2015.

_____. **METS: an overview & tutorial**. Sept. 2011. Disponível em: <<http://www.loc.gov/standards/mets/METSOversetw.v2.html>>. Acesso em: 13 maio 2015.

_____. **MODS User Guidelines**. MODS elements and attributes. Version 3. Nov. 2014a. Disponível em: <<http://www.loc.gov/standards/mods/userguide/>>. Acesso em: 6 out. 2015.

_____. **MODS: uses and features**. Dec. 2014b. Disponível em: <<http://www.loc.gov/standards/mods/mods-overview.html>>. Acesso em: 21 out. 2015.

_____. **PREMIS: Preservation Metadata Maintenance Activity**. 2015c. Disponível em: <<http://www.loc.gov/standards/premis/>>. Acesso em: 17 jul. 2015.

_____. **Sample MODS Version 3.5 XML Documents**. Book chapter (XML). Sept. 2015d. Disponível em: <<http://www.loc.gov/standards/mods/v3/modsbook-chapter.xml>>. Acesso em: 10 nov. 2015.

_____. **The Danny Kaye and Sylvia Fine Collection**. Cupid Takes a Holiday [video recording]. Oct. 2014c. Disponível em: <<http://memory.loc.gov/diglib/ihas/loc.natlib.ihas.200198140/mets.xml>>. Acesso em: 10 nov. 2015.

_____. **Thomas Bailey Aldrich papers, 1873-1901**. Sept. 2014d. Disponível em: <<http://rs5.loc.gov/findingaids/master/mss/eadxmlmss/2007/ms007107.xml>>. Acesso em: 10 nov. 2015.

- _____. **Using PREMIS with METS.** Introduction. Oct. 2010. Disponível em: <<http://www.loc.gov/standards/premis/louis-2-0.xml>>. Acesso em: 10 nov. 2015.
- LONG, A. S. **Long-term preservation of web archives:** experimenting with emulation and migration methodologies. [S.l.]: International Internet Preservation Consortium, dec. 2009. 54 p. Disponível em: <<http://www.netpreserve.org/sites/default/files/resources/Methodologies.pdf>>. Acesso em: 6 fev. 2015.
- LUSENET, Y. Digital heritage for the future. **Cadernos BAD – Cadernos de Biblioteconomia, Arquivística e Documentação**, Lisboa, n. 2, p. 15-27, 2002. Disponível em: <<http://www.bad.pt/publicacoes/index.php/cadernos/article/view/863/862>>. Acesso em: 8 set. 2015.
- MARCONDES, C. H.; SAYÃO, L. F. À guisa de introdução: repositórios institucionais e livre acesso. In: SAYÃO, L. et al. (Org.). **Implantação e gestão de repositórios institucionais:** políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 9-21. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ufba/473/3/implantacao_repositorio_web.pdf>. Acesso em: 12 nov. 2015.
- NATIONAL ARCHIVES OF AUSTRALIA. **Digital preservation policy:** preserving archival digital records transferred from commonwealth agencies. Version 1, 2. [Canberra], july 2011. Disponível em: <<http://www.naa.gov.au/about-us/organisation/accountability/operations-and-preservation/digital-preservation-policy.aspx>>. Acesso em: 30 mar. 2015.
- NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PRESERVATION PROGRAM (NDIIPP). Disponível em: <<http://www.digitalpreservation.gov/index.php>>. Acesso em: 20 fev. 2015.
- NATIONAL DIGITAL STEWARDSHIP ALLIANCE (NDSA). Disponível em: <<http://www.digitalpreservation.gov/ndsa/>>. Acesso em: 20 fev. 2015.
- NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). **Data Dictionary:** technical metadata for digital still images. Baltimore: NISO, c2011. 104 p. Disponível em: <http://www.niso.org/apps/group_public/download.php/14698/z39_87_2006_r2011.pdf>. Acesso em: 1 jul. 2015.
- _____. **Understanding metadata.** Bethesda: NISO Press, 2004. 16 p. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 30 abr. 2015.
- NATIONAL LIBRARY OF AUSTRALIA. **Digital preservation policy.** 4th ed. [Canberra], feb. 2013. Disponível em: <<http://www.nla.gov.au/policy-and-planning/digital-preservation-policy>>. Acesso em: 30 mar. 2015.
- NATIONAL MUSEUM OF AUSTRALIA. **Digital preservation and digitisation policy.** Version 2.2. Canberra, aug. 2012. 11 p. Disponível em:

<http://www.nma.gov.au/__data/assets/pdf_file/0013/1453/POL-C-028_Digital_preservation_and_digitisation-2.2_public.pdf>. Acesso em: 30 mar. 2015.

ONLINE COMPUTER LIBRARY CENTER (OCLC). Disponível em: <<http://www.oclc.org/home.en.html>>. Acesso em: 28 fev. 2015.

PEARSON, D.; DEL POZO, N. Explaining preservation actions: a working document. **National Library of Australia Staff Papers**, Canberra, p. 1-42, 2009. Disponível em: <<http://www-prod.nla.gov.au/openpublish/index.php/nlasp/article/view/1532/1843>>. Acesso em: 6 fev. 2015.

PINHEIRO, N. A. M.; SILVEIRA, R. M. C. F.; BAZZO, W. A. Ciência, tecnologia e sociedade: a relevância do enfoque CTS para o contexto do ensino médio. **Ciência & Educação**, Bauru, v. 13, n. 1, p. 71-84, 2007. Disponível em: <<http://www.scielo.br/pdf/ciedu/v13n1/v13n1a05.pdf>>. Acesso em: 30 jan. 2015.

PRESERVING AND ACCESSING NETWORKED DOCUMENTARY RESOURCES OF AUSTRALIA (PANDORA). Disponível em: <<http://pandora.nla.gov.au/>>. Acesso em: 20 fev. 2015.

RESEARCH LIBRARIES UNITED KINGDOM (RLUK). Disponível em: <<http://www.rluk.ac.uk/>>. Acesso em: 20 fev. 2015.

RODRIGUES, N. A. Introdução ao METS: preservação e intercâmbio de objetos digitais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 13, n. 26, p. 1-16, 2008. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p172/6644>>. Acesso em: 28 set. 2015.

RUSSELL, K.; SERGEANT, D. The Cedars project: implementing a model for distributed digital archives. **RLG DigiNews**, v. 3, n. 3, jun. 1999. Disponível em: <<http://worldcat.org/arcviewer/2/OCC/2009/07/29/H1248898656915/viewer/file2.html#pubinfo>>. Acesso em: 11 nov. 2015.

SANTA ANNA, J. A (r)evolução digital e os dilemas para a catalogação: os cibertecários em atuação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 13, n. 2, p. 312-328, maio/ago. 2015. Disponível em: <<http://periodicos.bc.unicamp.br/ojs/index.php/rdbci/article/view/8634632/3388>>. Acesso em: 14 set. 2015.

SANTOS, L. W.; ICHIKAWA, E. Y. CTS e a participação pública na ciência. In: SANTOS, L. W. et al. (Org.). **Ciência, tecnologia e sociedade**: o desafio da interação. Londrina: IAPAR, 2002. p. 239-271.

SARAMAGO, M. L. Metadados para preservação digital e aplicação do modelo OAIS. **Actas dos Congressos Nacionais de Bibliotecários, Arquivistas e Documentalistas**, Lisboa, n. 8, p. 1-6, 2004. Disponível em: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/640/637>>. Acesso em: 11 mar. 2015.

_____. Preservação digital a longo prazo: boas práticas e estratégias. **Cadernos BAD – Cadernos de Biblioteconomia, Arquivística e Documentação**, Lisboa, n. 2, p. 54-68, 2002. Disponível em: <<http://www.apbad.pt/CadernosBAD/Caderno22002/Saramago.pdf>>. Acesso em: 20 set. 2015.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibl: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 15, n. 30, p. 1-31, 2010. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1/19527>>. Acesso em: 10 mar. 2015.

SELLTIZ, C. et al. **Métodos de pesquisa nas relações sociais**. São Paulo: Herder, 1967.

SEVERINO, A. J. **Metodologia do trabalho científico**. 23. ed. rev. e atual. São Paulo: Cortez, 2007. 304 p.

SOCIETY OF AMERICAN ARCHIVISTS. Encoded Archival Description Working Group. **Encoded Archival Description Tag Library**: version 2002. Chicago, 2002. 308 p. (EAD Technical Document, n. 2). Disponível em: <http://www2.archivists.org/sites/all/files/EAD_TagLibrary_2002.pdf>. Acesso em: 23 jul. 2015.

SWISS FEDERAL ARCHIVES. **Digital archiving policy**. Bern, c2009. 39 p. Disponível em: <<http://www.bar.admin.ch/dokumentation/00445/00527/index.html?lang=en>>. Acesso em: 30 mar. 2015.

THE EUROPEAN LIBRARY (TEL). Disponível em: <<http://www.theeuropeanlibrary.org/tel4/>>. Acesso em: 20 fev. 2015.

THE NATIONAL ARCHIVES. **Digital preservation policies**: guidance for archives. Surrey, 2011. 16 p. Disponível em: <<http://www.nationalarchives.gov.uk/documents/information-management/digital-preservation-policies-guidance-draft-v4.2.pdf>>. Acesso em: 20 fev. 2015.

THOMAZ, K. P. **A preservação de documentos eletrônicos de caráter arquivístico**: novos desafios, velhos problemas. 2004. 388 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/VALA-68ZRKF/doutorado__katia_de_padua_thomaz.pdf?sequence=1>. Acesso em: 11 jan. 2015.

THOMAZ, K. P.; SOARES, A. J. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramZero**, Rio de Janeiro, v. 5, n. 1, fev. 2004. Disponível em: <http://www.dgz.org.br/fev04/F_I_art.htm>. Acesso em: 18 fev. 2015.

UNITED STATES GEOLOGICAL SURVEY (USGS) FUNDAMENTAL SCIENCE PRACTICES ADVISORY COMMITTEE. **USGS Guidelines for the Preservation of Digital Scientific Data**. [Virginia], apr. 2014. 6 p. Disponível em: <http://www.digitalpreservation.gov/ndsa/working_groups/documents/USGS_Guidelines_for_the_Preservation_of_Digital_Scientific_Data_Final.pdf>. Acesso em: 20 fev. 2015.

VALENTE, M. E.; CAZELLI, S.; ALVES, F. Museus, ciência e educação: novos desafios. **História, Ciências, Saúde – Manguinhos**, Rio de Janeiro, v. 12, supl., p. 183-203, 2005. Disponível em: <<http://www.scielo.br/pdf/hcsm/v12s0/09>>. Acesso em: 30 jan. 2015.

WEIBEL, S. L.; KOCH, T. The Dublin Core Metadata Initiative: mission, current activities, and future directions. **D-Lib Magazine**, v. 6, n. 12, dec. 2000. Disponível em: <<http://www.dlib.org/dlib/december00/weibel/12weibel.html>>. Acesso em: 24 jun. 2015.

ZENG, M. L.; QIN, J. **Metadata**. New York: Neal-Schuman Publishers, 2008. 365 p.