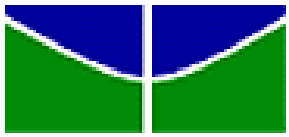


Universidade de Brasília
Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação
Departamento de Ciência da Informação e Documentação

Miguel Ángel Márdero Arellano

CRITÉRIOS PARA A PRESERVAÇÃO DIGITAL DA INFORMAÇÃO CIENTÍFICA

Brasília – DF
2008

Universidade de Brasília
Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação
Departamento de Ciência da Informação e Documentação

CRITÉRIOS PARA A PRESERVAÇÃO DIGITAL DA INFORMAÇÃO CIENTÍFICA

Miguel Ángel Márdero Arellano

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação, do Departamento de Ciência da Informação e Documentação da Universidade de Brasília como requisito parcial para obtenção do grau de Doutor em Ciência da Informação.

Orientador: Prof. Dr. Murilo Bastos da Cunha

Área de concentração: Transferência da Informação
Linha de pesquisa: Gestão da Informação e do Conhecimento

Brasília – DF
2008

Márdero Arellano, Miguel Ángel.

Critérios para a preservação digital da informação científica / Miguel Ángel Márdero Arellano. – Brasília: Universidade de Brasília, 2008.
xix, 354 f. : il.

Tese (Doutorado em Ciência da Informação) – Universidade de Brasília.
Possui gráficos, tabelas e referências.

1. Preservação Digital. 2. Critérios de Preservação Digital. 3. Repositório Digital. 4. Modelo de Referência OAIS. 5. Informação Científica. 6. MCT.

UNIVERSIDADE DE BRASÍLIA (UnB)
Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação (FACE)
Departamento de Ciência da Informação e Documentação (CID)
Programa de Pós-Graduação em Ciência da Informação (PPGCInf)

FOLHA DE APROVAÇÃO

Título: "Critérios para a preservação digital da informação científica".

Autor: Miguel Angel Márdero Arellano

Área de concentração: Transferência da Informação

Linha de pesquisa: Gestão da Informação e do Conhecimento.

Tese submetida à Comissão Examinadora designada pelo Colegiado do Programa de Pós-graduação em Ciência da Informação do Departamento de Ciência da Informação e Documentação da Universidade de Brasília como requisito parcial para obtenção do título de **Doutor em Ciência da Informação.**

Tese aprovada em: 04 de dezembro 2008.

Aprovado por:

Prof. Dr. Murilo Bastos da Cunha
Presidente - (UnB/PPGCInf)

Profa. Dra. Ivette Kafure – UnB/CID
Membro Interno – (UnB/CID)

Rubens Ribeiro Gonçalves da Silva
Membro Externo – (UFBA)

Prof. Dr. Hélio Kuramoto
Membro Externo – (IBICT)

Profa Dra Sely Maria de Souza Costa
Membro Interno – (UnB/PPGCInf)

Profa Dra Elmira Luzia Melo Soares Simeão
Suplente – (UnB/PPGCInf)

Agradecimentos

Ao meu orientador, professor Murilo Bastos da Cunha

Aos professores e funcionários do CID-UnB

A minha amiga, bibliotecária e *proofreader*, Tereza Walter

A minha amiga, colega e revisora, Margaret de Palermo

A minha amiga, colega e bibliotecária da BPCT, Elise do Nascimento

Aos colegas e chefes do Ibict

Aos colegas das unidades de pesquisa do MCT

Aos amigos de longe e de perto

Aos humanos e animais que gostam de mim...

"Confiança - o senhor sabe - não se tira das coisas feitas ou perfeitas; ela rodeia é o quente da pessoa"

João Guimarães Rosa
GRANDE SERTÃO: VEREDAS
(pg. 49)

RESUMO

O presente trabalho tentou responder ao desafio de identificar um conjunto de critérios de preservação digital essenciais para os sistemas de informação científica nas unidades de pesquisa do Ministério da Ciência e Tecnologia do Brasil. Para a elaboração da proposta de critérios, formularam-se quatro objetivos específicos: a) caracterizar o estado atual das práticas de preservação digital; b) identificar na literatura científica os critérios usados na construção dos sistemas de preservação digital em funcionamento; c) identificar as necessidades de preservação digital do grupo de instituições selecionadas para o estudo, e d) identificar os conceitos e aplicações usados que podem interferir ou não na implementação de procedimentos padronizados de preservação digital da informação científica. Constatou-se, por meio da revisão da literatura, a inexistência de avaliações dos fundamentos teóricos dos critérios e requisitos de preservação digital adotados em projetos internacionais, e que são poucos os estudos que calculam os benefícios observados em sua aplicação. Tanto a arquivologia quanto a ciência da informação têm na preservação digital um objeto de estudo que engloba todas as tarefas envolvidas no fluxo informacional, mas são áreas que estão começando a propor padrões, normas, políticas e procedimentos para o arquivamento digital. Um critério que se destaca no desenvolvimento das práticas de preservação digital citadas na literatura é o da confiabilidade, determinante na formação do conjunto de critérios aqui propostos. A metodologia adotada baseou-se no uso das técnicas de questionário, entrevista, observação e análise etnográfica de conteúdo, sob a perspectiva de David Altheide. Para poder conhecer as abordagens teóricas e metodológicas que têm sido publicadas, realizou-se um estudo bibliométrico descritivo, coletando dados de referências bibliográficas sobre o tema da preservação digital do período de 1996 até 2007. A análise dessa produção científica utilizou o *Statistical Package for the Social Sciences-SPSS*. A análise dos dados qualitativos coletados na pesquisa de campo mostrou que a maioria dos responsáveis pelos centros de informação e bibliotecas do Ministério da Ciência e Tecnologia percebe a preservação digital como os procedimentos relacionados com a capacidade de armazenamento, migração e disseminação de conteúdos que, em sua maioria, são administrados de forma centralizada. Apenas os representantes de uma instituição reconheceram a importância das políticas institucionais na definição das funções de inserção e gerenciamento dos conteúdos arquivados na sua biblioteca digital. Os critérios de preservação digital propostos estão relacionados aos aspectos institucionais (Confiabilidade, Responsabilidade Política e Sustentabilidade Econômica) e aos aspectos gerenciais (Inclusão em repositórios digitais, Transparência e Acessibilidade de longo prazo). Estes critérios são considerados elementos determinantes das propriedades relativas aos sistemas de gerenciamento de informação científica. Entre as sugestões para estudos futuros, recomenda-se o desenvolvimento de pesquisas sobre a aprendizagem e aplicação de normas e padrões pelos integrantes das equipes que gerenciam os sistemas de informação científica do governo, a inclusão de uma discussão detalhada sobre metadados de preservação para informação científica e sobre o gerenciamento de longo prazo dos direitos autorais no ambiente digital.

Palavras-chave: Preservação digital. Critérios de preservação digital. Repositório digital. Modelo de referência OAIS. Informação científica. Ministério da Ciência e Tecnologia.

ABSTRACT

This thesis aims to propose a Digital Preservation Criteria Set which is crucial for scientific information systems in the research units of the Science and Technology Ministry of Brazil. Four objectives were formulated in proposing the criteria: a) describe the digital preservation procedures currently being used; b) identify in scientific literature the key criteria used in building digital preservation systems; c) identify the digital preservation demands from the sampled research units; and d) identify both concepts and procedures that might interfere with standard digital preservation procedures of scientific information. A literature review has highlighted that a theoretical evaluation of both digital preservation criteria and requirements from international projects is nonexistent. It also highlighted that there are a very limited number of studies assessing the benefits produced by the application of this criteria and these requirements. Within the context of digital preservation both Archivistic and Information Science systems deal with tasks related to information flow. Only recently, are these areas proposing patterns, norms, policies and procedures directed to digital archiving. Trustworthy is a criteria for digital preservation procedures that stands out in the Scientific literature and is critical in identifying the digital preservation criteria set for scientific information. The proposed digital preservation criteria set is related to institutional aspects (Trustworthy, Political Responsibility and Economic Sustainability) and management aspects (Digital Repository Inclusion, Transparency, and Long-term Accessibility). These criteria are considered determinant elements in the management proprieties of administration systems for scientific information. The methodology in data collection included surveys, interviews, observation, and ethnographic content analysis according to David Altheide's approach. A descriptive bibliometric study was conducted in order to identify both the theoretical and methodological approaches in digital preservation literature from 1996 to 2007. Data analysis was performed using the Statistical Package for the Social Sciences-SPSS. Qualitative data from field research indicated that most of the information unit managers consider digital preservation as procedures related to storage capacity, migration and content dissemination. They also realize that these procedures are administered in a centralized manner. Only one research unit's participants recognized the importance of institutional policies in defining data ingest functions and management of contents archived in the institutional digital library. Suggestions for future studies include investigating the acquisition and application of Norms and Standards by government staff responsible for the administration of scientific information systems and extensive discussions both on Preservation Metadata relating to scientific information and Long-term management of propriety rights within the digital environment.

Key-words: Digital preservation. Digital preservation criteria. Digital repository. OAIS reference model. Scientific Information. Ministério da Ciência e Tecnologia.

Lista de Figuras

Figura 1 - Contexto das necessidades de preservação digital da informação científica 53
Figura 2 - Expectativa de vida de várias mídias para armazenamento de informação 55
Figura 3 - Modelos do processamento nas perspectivas de preservação digital 60
Figura 4 - Diferenças entre estratégias de preservação 62
Figura 5 - Um documento digital encapsulado 69
Figura 6 - Esquema de metadados da NZNL 79
Figura 7- Estrutura de metadados de preservação das UNT *Libraries* 82
Figura 8 - Registro de metadados da *Victorian Electronic Records Strategy* 84
Figura 9 - Ferramenta para extração de metadados de preservação da NLNZ .. 86
Figura 10 - Modelo de referência *Open Archival Information System* (OAIS) ou Sistema Aberto para Arquivamento de Informação (SAAI) 90
Figura 11 - Modelo funcional OAIS 91
Figura 12 - Exemplo de monitoramento de caixas LOCKSS 116
Figura 13 - Página do OJS/SEER opção Arquivamento da Revista no LOCKSS 119
Figura 14 - Crescimento dos arquivos/repositórios 126
Figura 15 - Primeira interface do DSpace da Cornell University 139
Figura 16 - Interface da Biblioteca Digital Memória do INPE 146
Figura 17 - Ciclo de vida dos objetos digitais 151
Figura 18 - Modelo de Repositório Digital Confiável (TDR) da RLG e OCLC 180
Figura 19 - Diagrama da integração dos dois pilares da preservação digital baseado no modelo TDR da RLG-OCLC 181
Figura 20 - Modelo funcional OAIS no projeto SHERPA 183
Figura 21 - Subtemas e locais das publicações 216
Figura 22 - Contribuição dos Subtemas no período 1996 a 2007 217
Figura 23 - Freqüência do tipo de autoria em função do ano de publicação 218
Figura 24 - Tipos de propostas apresentadas no período 1996 a 2007 219
Figura 25 - Tipologia de documentos 223
Figura 26 - Tendência da produção de documentos no período 1996 a 2007 223
Figura 27 - Tipo de produção em função do ano de produção 224

Lista de Quadros

Quadro 1 - Relevância dos requisitos de preservação 45
Quadro 2 - Expectativa de vida de algumas mídias para fins de 56
preservação digital
Quadro 3 - Métodos de preservação digital 61
Quadro 4 - Ferramentas para editoração eletrônica de revistas 118
Quadro 5 - Estratégias de preservação no fluxo dos objetos digitais na 123
biblioteca digital
Quadro 6 - Gerenciamento dos dados preservação digital 149
Quadro 7 - Estratégias de preservação adotadas pelos 21 sistemas 159
operacionais
Quadro 8 - Classificação das unidades de pesquisa do Ministério da 198
Ciência e Tecnologia
Quadro 9 - Local de preservação e acesso da informação científica em 200
formato digital produzida nas unidades de pesquisa do MCT
Quadro 10 - Listagem de respondentes das unidades de pesquisa do 207
MCT entrevistados na fase exploratória e de pesquisa de
campo
Quadro 11 - Plano de procedimento de observação 208
Quadro 12 - Questões debatidas em cada ano do período 221
Quadro 13 - Tipo de material segundo e setor responsável no centro de .. 231
informação e biblioteca
Quadro 14 - Bases de dados usadas nos centros de informação e 233
bibliotecas para o registro .. dos materiais bibliográficos
Quadro 15 - Ferramenta de indexação utilizada segundo o tipo de 235
material
Quadro 16 - Tipo de acesso disponibilizado pelas unidades de pesquisa 237
do MCT segundo o tipo de material
Quadro 17 - Formas de tratamento da produção científica da unidade de 238
pesquisa
Quadro 18 - Métodos de preservação aplicados aos materiais 240
bibliográficos nas unidades de pesquisa
Quadro 19 - Listagem de entrevistados das unidades de pesquisa do 249
MCT
Quadro 20 - Entrevistados por setor das unidades de pesquisa do MCT .. 252
Quadro 21 - Função dos entrevistados nas unidades de pesquisa 253
Quadro 22 - Conteúdo dos documentos produzidos por servidores do 272
MCT sobre preservação digital

Lista de Tabelas

Tabela 1 - Produção dos autores em função dos subtemas de interesse .. 215
Tabela 2 - Freqüência de publicações e autoria em função do subtema ... 216
Tabela 3 - Tipos de propostas para preservação digital 219
Tabela 4 - Questões debatidas no período 1996-2007 220
Tabela 5 - Propostas e questões debatidas ao longo do período 1996-2007 222
Tabela 6 - Distribuição de freqüência da área do conhecimento 225
Tabela 7 - Tipos de Publicação em função da área de conhecimento 225
Tabela 8 - Título e país de origem dos periódicos com artigos sobre ..... preservação digital 226
Tabela 9 - Servidores nas bibliotecas das unidades de pesquisa do MCT 245
Tabela 10 - Produção bibliográfica sobre preservação digital no MCT ..... (1998-2007) 270

Lista de Abreviaturas e Siglas

ABEC	Associação Brasileira de Editores Científicas
ACII	*American Standard Code for Information Exchange*
AEB	Agência Espacial Brasileira
AEC	Análise Etnográfica de Conteúdo
AHDS	Arts and Humanities Data Service
AIA	Aerospace Industries Association/Boing Co.
AINFO-NTIA	AINFO/National Telecommunications and Information Administration
AIP	American Institute of Physics
AQD	Análise Qualitativa de Documentos
Arc	Australian Research Council
ARK	*Archival Resource Key*
ASCII	*American Standard Code for Information Interchange*
ASP	*Active Server Pages*
BDJur	Biblioteca Digital Jurídica
BDTD	Biblioteca Digital de Teses e Dissertações
BIREME	Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
BL	British Library
BMC	BioMed Central
BOAI	*Budapest Open Archive Initiative*
BPCT	Biblioteca de Política em Ciência e Tecnologia
C&T	Ciência e Tecnologia
CAPES	Coordenação de Aperfeiçoamento do Ensino Superior
CBPF	Centro Brasileiro de Pesquisas Físicas
CCSDS	*Consultative Committee for Space Data Systems*
CDI	Coordenação de Documentação e Informação Científica
CDPP	Centre de Données de la Physique des Plasmas
CD-R	*Compact Disk - Recordable*
CD-ROM	*Compact Disk – Read Only Memory*
CDU	Classificação Universal Decimal
CEDAR	*Consortium of University Research Libraries (CURL) Exemplars in Digital Archives*

CEITEC	Centro de Excelência em Tecnologia Eletrônica Avançada
CENDI	*U.S. Federal Information Managers Group*
CENPRA	Centro de Pesquisas Renato Archer
CERN	European Organization for Nuclear Research
CETEM	Centro de Tecnologia Mineral
CGEE	Centro de Gestão e Estudos Estratégicos
CGGI	Coordenação-Geral de Gestão e Inovação
cgi	*Computer Graphic Interface*
CLIR	Council of Library Information Research
CNEN	Comissão Nacional de Energia Nuclear
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNRI	Corporation for National Research Iniciatives
Conarq	Conselho Nacional de Arquivos
CORDIS	Community Research & Development Information Service
CPA	Commission on Preservation and Access
CRiB	*Conversion and Recommendation of Digital Object Formats*
CSDGM	Content Standard for Digital Geospatial Metadata
DARPA	Defense Advanced Research Projects Agency
DBMS	*Database Management System*
DCC	Digital Curation Centre
DIAS	Digital Information Archiving Systems Dutch National Library
DICI	Diálogo Científico (IBICT)
DIOnAS	*Data Ingest and Online Access Sub-System*
DIP	*Dissemination Information Package*
DiVA	Electronic Publishing Centre, Uppsala University Library
DOI®	*Digital Object Identifier*
DPC	Digital Preservation Coalition
DROID	*Digital Record Object Identification*
DRS	*Digital Rosetta Stone*
DTD	*Document Type Definition*
DTIC	*Defense Technical Information Center*
DTS	*Data Time Stamping*
DVD	*Digital Video Disk*

EAD	*Encoding Archives Description*
EBSCO	Elton B. Stephens Company
ECPA	European Commission on Preservation and Access
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ERA	Electronic Record Archives
EROS	Earth Resources Observation Systems Data Center
ERPANET	*Electronic Resource Preservation and Access Network*
ESD	Elsevier Science Direct – also part of the Dutch National Library
FAPESP	Fundação de Apoio à Pesquisa do Estado de São Paulo
Fedora	*Flexible Extensible Digital Object Repository Architecture*
FINEP	Financiadora de Estudos e Projetos do Estado de São Paulo
GIF	*Graphics Interchange Format*
GILS	*Government Information Locator Service*
GPO	U.S. Government Printing Office
HD-ROM	*High-Density - Read Only Memory*
HP	Hewlett-Packard Co.
HTML	*Hyper Text Markup Language*
HTTP	*Chat & Hyper Text Transfer Protocol*
NGO	*Non Government Organization*
Ibict	Instituto Brasileiro de Informação em Ciência e Tecnologia
IBM	International Business Machines
ICSTI	International Council for Scientific and Technical Information
ICT	Informação em Ciência e Tecnologia
IDSM	Instituto de Desenvolvimento Sustentável Mamirauá
IES	Instituição de Ensino Superior
IFLA	International Federation of Libraries Association
IMAGE	*Imager for Magnetopause-to-Aurora Global Exploration*
IMLS	Institute of Museum and Library Services
IMPA	Associação Instituto Nacional de Matemática Pura e Aplicada
INB	Indústrias Nucleares do Brasil
INIST-CNRS	Institut de l´Information Scientifique et Technique
INPA	Instituto Nacional de Pesquisas da Amazônia
INPE	Instituto Nacional de Pesquisas Espaciais

INSA	Instituto Nacional do Semi-Árido
INT	Instituto Nacional de Tecnologia
InterPARES	*The International Research on Permanent Authentic Records in Electronic Systems*
IPR	*Intellectual Propriety Rights*
ISBN	*Internation Standard Book Number*
ISO	International Organization for Standadization
IUCR	International Union of Crystallography
JISC	Joint Information Systems Committee
JPG	*Joint Photographic Experts*
JSTOR	Electronic-Archiving Initiative
KB	National Library of the Netherlands
KB	Koninklijke Bibliotheek
LATEX	*LAmport Tex*
LC	Library of Congress
LNA	Laboratório Nacional de Astrofisica
LNCC	Laboratório Nacional de Computação Científica
LNLS	Laboratório Nacional de Luz Síncrotron / ABTLus
LOCKSS	*Lots of Copies Keep Stuff Safe*
LOTAR	*Long Term Archiving and Retrieval and Product Data Within the Aerospace Industry*
LSDA	*Life Science Data Archive*
MAST	Museu de Astronomia e Ciências Afins
MCT	Ministério de Ciência e Tecnologia
MD5	*Message-Digest algorithm 5 (para criptografar)*
METS	*Metadata Encoding & Transmission Standard*
MIME	*Multipurpose Internet Mail Extensions*
MIT	Massachussetts Institute of Technology
MITCET	Council on Educational Technology
MIX	*Metadata for Image in XML*
MKA	*Meta Knowledge Archive*
MoReq	*Model Requirements for the Management of Electronic Records*
MPEG	Museu Paraense Emílio Goeldi
MS DOS	*Microsoft Disk Operating System*
N2T	*Name-to-thing*

NARA	National Archives and Records Administration
NASA	Goddard Space Flight Center Library
NDIIP	*National Digital Information Infrastructure for Preservation Program*
NEDLIB	Network European Deposit Library
NEH	*The National Endowment for the Humanities*
NESTOR	*Network of expertise in Long-Term Storage of Digital Resources*
NHPRC	National Historical Publications and Records Commission
NISO	National Information Standards Organization
NLA	National Library of Australia
NLM	Profiles in Science, National Library of Medicine
NLNZ	National Library of New Zealand
NMM	National Motor Museum
NPO	National Preservation Office
NSF	National Science Foundation
Nuclep	Nuclebrás Equipamentos Pesados
NZLS	*New Zeland Locator Service*
OAI	Open Archives Initiative
OAI-PMH	*Open Archive Initiative Protocol Metadata Harvesting*
OAIS	*Open Archive Information System*
OARINZ	Open Access Repositories in New Zealand
OCLC	Online Computer Library Center
OCW	*Open Courseware*
OE	Objetivo Específico
OhioLINK	*Ohio Library and Information Network*
OJS	*Open Journal Systems*
ON	Observatório Nacional
OpenDOAR	*Directory of Open Access Reposiories*
P&D	Pesquisa e Desenvolvimento
PAI	Pacote de Arquivamento de Informação
PADI	*Preserving Access to Digital Information*
PANDORA	*Australia's Web Archive*
PDF	*Portable Document Format*
PDI	Pacote de Disseminação de Informação

PKP	*Public Knowledge Project*
PLANETS	*Preservation and Long-term Access through Networked Services*
PLEDGE	*PoLicy Enforcement in Data Grid Environments*
PPS	Photographic Preservation Society
PREMIS	*Preservation Metadata: Implementation Strategies*
PRO	Public Record Office
PSI	Pacote de Submissão de Informação
PubMed Central	National Library of Medicine
PURL	*Persistent Uniform Resource Locators*
RDF	*Resource Description Framework*
RI	*Repositório Institucional*
RIDI	Repositório Institucional Digital do IBICT
RLG	Research Library Group
RNP	Associação Rede Nacional de Ensino e Pesquisa
ROAR	*Registry of Open Access Repositories*
SAAI	Sistema de Arquivamento Aberto de Informação
SBPC	Sociedade Brasileira para o Progresso da Ciência
SciELO	*Scientific Electronic Library Online*
SDSC	*San Diego Supercomputer Center*
SEER	Sistema Eletrônico de Editoração de Revistas
SGML	*Standard Generalized Markup Language*
SHERPA	*Securing a Hybrid Environment for Research Preservation and Access*
SIP	Pacote de Submissão de Informação
SNBU	Seminário Nacional de Bibliotecas Universitárias
SOA	Arquitetura de Serviço Orientado
SPSS	*Statistical Package for Social Sciences*
TCC	Trabalho de Conclusão de Curso
TCI	Tesauro da Ciência da Informação
TCL	*Tool Command Language*
TDR	*Trust Digital Repository*
TOM	*Typed Object Model*
TRAC	*Trustworthy Respository Audit & Certification: Criteria and Checklist*

UNESCO	United Nations Educational, Scientific and Cultural Organization
UPS	*Universal Preprint Service*
URL	*Uniform Resources Location*
URLib	*Uniform Repository for a Library*
URN	*Uniform Resource Name*
USP	Universidade de São Paulo
VERS	*Victorian Electronic Records Strategy – Austrália*
VHS	*Video Home System*
VRA	*Visual Resources Association*
WEB	Sinônimo de World Wide Web
WIPO	*World Intellectual Property Organization*
XINGÓ	Instituto de Desenvolvimento Científico e Tecnológico Xingó
XML	*Extensible Markup Language*

SUMÁRIO

AGRADECIMENTOS iii
RESUMO iv
ABSTRACT v
LISTA DE FIGURAS vi
LISTA DE QUADROS vii
LISTA DE TABELAS viii
LISTA DE ABREVIATURAS E SIGLAS ix

1 INTRODUÇÃO 20
1.1 APRESENTAÇÃO DO TEMA 21
1.2 DEFINIÇÃO DO PROBLEMA 26
1.3 OBJETIVOS 29
1.3.1 Objetivo geral 29
1.3.2 Objetivos específicos 29

2 REVISÃO DE LITERATURA 30
2.1 CIÊNCIA E PRESERVAÇÃO DIGITAL 31
2.1.1 Tecnologia e evolução dos documentos 36
2.1.2 Arquivologia e preservação digital 38
2.1.3 Ciência da informação e preservação digital 41
2.2 PRESERVAÇÃO DIGITAL 43
2.2.1 Preservação da informação em objetos digitais 43
2.2.2 Necessidades de preservação digital 50
2.2.3 Estratégias para a preservação digital 54
2.2.3.1 Preservação da seqüência de *bits* 59
2.2.3.2 Preservação da tecnologia e do objeto digital 61
2.2.3.3 Migração 62
2.2.3.4 Emulação e encapsulamento 68
2.2.3.5 Conservação de *hardware* e *software* 71
2.2.4 Metadados de preservação 72
2.2.4.1 Iniciativas de implementação de metadados de preservação 82

2.2.4.2 A norma OAIS 87

2.3 PRESERVAÇÃO DIGITAL DA PRODUÇÃO CIENTÍFICA 99

2.3.1 Preservação digital e acesso livre 99

2.3.1.1 OAI e a preservação do acesso de longo prazo 100

2.3.1.2 Acesso livre e a comunidade científica 102

2.3.1.3 Autodepósito e preservação digital 105

2.3.1.4 “Preservação” e “preservação” 107

2.3.1.5 Tecnologias para o acesso livre 109

2.3.2 Preservação de periódicos eletrônicos 110

2.3.2.1 Arquivamento de periódicos eletrônicos 113

2.3.2.2 LOCKSS 114

2.3.2.3 Gerenciamento editorial e preservação digital 117

2.3.3 Preservação em repositórios digitais 120

2.3.3.1 Bibliotecas digitais 122

2.3.3.2 Repositórios digitais 124

2.3.3.3 Repositórios institucionais 127

2.3.3.4 Repositórios institucionais nas bibliotecas 131

2.3.3.5 *Softwares* para repositórios institucionais 133

2.3.3.6 Repositórios DSpace 138

2.3.3.7 Repositórios e identificadores persistentes 143

2.3.3.8 Repositório uniforme do Inpe 146

2.4 GERENCIAMENTO DA PRESERVAÇÃO DIGTAL 147

2.4.1 Infra-estrutura tecnológica 153

2.4.2 Infra-estrutura organizacional 155

2.4.3 Consolidação das iniciativas de preservação digital 157

2.5 SUSTENTABILIDADE DA PRESERVAÇÃO DIGITAL 162

2.5.1 Sustentabilidade comercial 166

2.6 POLÍTICAS DE PRESERVAÇÃO DIGITAL 169

2.6.1 Permissões para preservar 171

2.6.2 Gerenciamento dos direitos de propriedade 174

2.7 MAPEAMENTO DOS CRITÉRIOS DE PRESERVAÇÃO DIGITAL 177

2.7.1 Atributos de repositórios digitais 184

2.7.2 Certificação de repositórios digitais 185

2.8 CONCLUSÕES DA REVISÃO DE LITERATURA 189

3 METODOLOGIA 193

3.1 PROBLEMA INVESTIGADO 196

3.2 UNIVERSO DE PESQUISA E AMOSTRA 196

3.3 INSTRUMENTOS DE COLETA DE DADOS 201

3.3.1 Estudo bibliométrico 201

3.3.2 Questionários 202

3.3.3 Pesquisa de campo 203

3.3.4 Entrevistas 204

3.3.5 Observações de curta duração 208

4 **ANÁLISE DOS DADOS** 210

4.1 ANÁLISE BIBLIOMÉTRICA 210

4.1.1 Bibliometria descritiva da preservação digital 211

4.1.2 Avaliação da produtividade 214

4.1.3 Temáticas e questões discutidas 228

4.2 ANÁLISE DOS DADOS DOS QUESTIONÁRIOS 229

4.2.1 Características dos centros de informação 230

4.2.2 Perfil profissional das equipes 243

4.3 ANÁLISE DOS DADOS DAS ENTREVISTAS 246

4.3.1 Perfil profissional dos entrevistados 250

4.3.2 Análise dos conteúdos das entrevistas 253

4.3.2.1 Sobre as próprias funções e atividades de preservação 254

4.3.2.2 Sobre o tratamento dos objetos digitais na instituição 258

4.4 ANÁLISE ETNOGRÁFICA DE CONTEÚDO 262

4.4.1 Observações de campo 264

4.4.2 Os documentos 267

4.4.2.1 Materiais de divulgação 268

4.4.2.2 Documentos oficiais 268
4.4.2.3 Material bibliográfico 269
4.4.3 Questões discutidas 270
4.5 DETERMINAÇÃO DE CRITÉRIOS 273
4.5.1 Critérios institucionais 276
4.5.1.1 Confiabilidade 277
4.5.1.2 Responsabilidade política 278
4.5.1.3 Sustentabilidade Econômica 279
4.5.1.4 Inclusão em repositórios digitais 280
4.5.1.5 Transparência 281
4.5.1.6 Acessibilidade de longo prazo 282
5 CONCLUSÕES E SUGESTÕES 284
REFERÊNCIAS 290
APÊNDICE A – Questionário 319
APÊNDICE B – Roteiros de entrevistas 321
APÊNDICE C – Lista de projetos e iniciativas de preservação digital 323
APÊNDICE D – Centros de informação e bibliotecas das unidades de pesquisa do Ministério da Ciência e Tecnologia 330
APÊNDICE E – Produção bibliográfica dos servidores do Ministério da Ciência e Tecnologia sobre a preservação digital 347
APÊNDICE F – Glossário de termos técnicos 350

1 INTRODUÇÃO

O propósito desta investigação foi identificar os critérios de preservação digital estabelecidos nas políticas, programas e projetos internacionais e as recomendações que instituições no exterior e no Brasil estão oferecendo para a preservação da informação científica e tecnológica em formato digital.

Apresentar-se-á um trabalho teórico sobre o estado-da-arte e serão aglutinados os critérios de preservação digital, dentro da proposta de um conjunto mínimo de critérios que possam ser aplicados nas unidades de pesquisa do Ministério da Ciência e Tecnologia (MCT).

A proposição inicial é apresentar o tema da preservação digital, as definições conceituais mais utilizadas e sua aplicação na realidade em um tipo específico de instituições de pesquisa.

A preservação digital é uma preocupação dos profissionais da informação, devido principalmente ao uso extensivo que tem sido feito das tecnologias digitais e do registro de informações em soluções cada vez mais sofisticadas e complexas.

Os antecedentes considerados significativos são, especificamente, a ausência de trabalhos científicos que tratem sobre o contexto desses registros e a falta de padronização que possa ajudar no tratamento desses dados. A pesquisa a ser relatada pretende contribuir para as investigações das transformações que os sistemas de informação vêm enfrentando.

Para realizar o trabalho, foram selecionados alguns métodos a serem empregadas em diferentes momentos da investigação. Em primeiro lugar foram feitos levantamentos bibliográficos, principalmente em textos que abrangem os anos de 1996 ao final de 2007.

Identificados os principais autores e instituições dedicadas à pesquisa na área, foram realizadas análises bibliométricas para se chegar aos produtores das normas e padrões que poderiam ser usados na construção do conjunto de critérios adequados a um sistema de preservação digital no Brasil.

Nesse sentido, foi preciso conhecer a atuação de um grupo de instituições científicas brasileiras, representadas por órgãos públicos – as unidades de pesquisa do Ministério da Ciência e Tecnologia. Para chegar à compreensão de sua realidade, foram enviados questionários e realizadas visitas à maioria dessas

instituições, observando atividades que a equipe responsável identificou como relacionadas à preservação digital; e, por último, fez-se um registro sonoro das entrevistas com as pessoas envolvidas com o tratamento dos dados digitais nessas unidades.

Na análise dos dados foi mencionada a definição de documento dentro da técnica de Análise Etnográfica de Conteúdo (AEC), na qual ele é uma representação simbólica que pode ser registrada e recuperada. Essa técnica é utilizada na análise de padrões de surgimento e uso de conceitos na divulgação de notícias e eventos ligados a situações complexas. Nesta pesquisa, a AEC foi usada para conseguir identificar os padrões que levam à construção de conceitos relacionados com a preservação digital, que servem para configurar a noção que as instituições empregam para justificar as atividades que têm sido aplicadas no tratamento dos registros digitais.

Este trabalho não pretendeu fazer um levantamento de todos os sistemas de preservação digital existentes, mas reconhecer, dentre os mais citados na literatura, os critérios que possam interessar às organizações que lidam com informação científica em formato digital, os desafios que elas enfrentam na preservação do acesso e no gerenciamento dos repositórios digitais de registros da ciência. Espera-se que esses critérios possam contribuir para o desenvolvimento de sistemas de gerenciamento de preservação e na implementação de políticas, programas e projetos dedicados ao tema.

1.1 APRESENTAÇÃO DO TEMA

O tema desta pesquisa é a preservação da informação científica em documentos digitais, para poder formular o conjunto de critérios que caracterizam sua implementação e funcionamento em instituições de pesquisa do governo brasileiro. A proposição surge da necessidade de produzir uma estrutura teórica e uma abordagem metodológica adequada ao problema da preservação da informação científica em formato digital.

A investigação procurou ordenar uma linha de pensamento teórica e conceitual para produzir conhecimentos e instrumentos apropriados para tratar uma

parte fundamental do fluxo interno da gestão da informação digital que é o planejamento das atividades de preservação de longo prazo.

Com base em resoluções da United Nations Educational, Scientific and Cultural Organization (UNESCO), como a "Carta sobre a Preservação Digital" (2003) e "Recomendações sobre *Software* Livre para Repositório e Sistema de Preservação" (2007), pode-se afirmar que existe necessidade urgente de salvar o patrimônio cultural e científico digital, para garantir o acesso a esses recursos às funcionalidades dos registros eletrônicos autênticos.

Os documentos digitais são frágeis e não existe segurança suficiente nos sistemas eletrônicos para garantir sua preservação digital de longo prazo. Essa fragilidade pode estar em sua rápida degradação física, na obsolescência tecnológica, na complexidade e nos custos. Também deve ser mencionada a necessidade da formação de profissionais especializados em todas as áreas envolvidas no processo de preservação digital que inclui bibliotecários, arquivistas, museólogos e o pessoal de tecnologia da informação.

A dependência social na informação digital vai além do simples acesso rápido aos dados, mas se expressa no uso desses dados como provas das funções e atividades de pessoas, grupos e instituições.

O problema da preservação digital está no conteúdo dos objetos digitais, nas informações armazenadas e na maneira como foram armazenadas. Não são apenas seqüências de zeros e uns ou cadeias de *bits* (*bitstream*)[1], mas o conjunto de decisões que definiram a formação básica do objeto informacional como um objeto único. Isso envolve a descrição de documentos que podem ser representados de diferentes formas, mas o seu conteúdo será sempre interpretado do mesmo modo, no momento da sua recuperação, ou seja, será a mesma cadeia de *bits* com seu significado original.

A maioria dos bibliotecários e arquivistas aceita que a preservação digital depende de modo geral da migração de material digital que permite que ele seja compatível com novas tecnologias. Mas a migração é muito mais complexa do que apenas transferir o *bitstream* de uma mídia para outra. A estrutura interna e o conteúdo do material devem ser preservados e transferidos igualmente, para que, dessa forma, o "novo" objeto seja uma representação fiel do original.

---

[1] NA: o *bit* ou dígito binário é a menor unidade de informação utilizada nas áreas de computação e informática, o *bitstream* é a camada mais elementar para dotar de significado um objeto digital.

Os objetos digitais, freqüentemente complexos, são compostos de tipos heterogêneos fechados e sem possibilidade de modificação depois de criados. O que os cientistas apreciam nesses objetos ou formatos pode estar em conflito com os métodos de preservação. Isso produz tensão entre cientistas que desejam ter uma mídia digital “viva e respirando” e os preservacionistas que apenas querem que o objeto seja estável para seu uso infinito.

Os arquivos digitais com informações científicas podem ter seu significado expresso pelos *softwares* no qual foram criados, já que eles são dependentes desses programas, mas a sua preservação não é apenas salvar o programa que interpreta o arquivo. Os documentos digitais precisam também da máquina na qual são lidos, complicando ainda mais sua retenção e recuperação.

A dificuldade da preservação digital deve-se principalmente à necessidade de retenção do objeto digital e do seu significado. A carência maior está na definição de técnicas de preservação digital capazes de compreender e reproduzir a forma e a função original do objeto, para garantir sua autenticidade e acessibilidade, pois eles não são apenas objetos físicos.

Como respostas a esse desafio, surgem estratégias de preservação digital que procuram incorporar todos os aspectos relacionados ao problema tecnológico: custos, legislação, gestão, acesso, políticas e critérios. São formas de reunir soluções parciais ante um problema complexo no qual estão envolvidos, entre outros itens, a migração, emulação, arqueologia digital, criptografia, metadados, formatos-padrão e *software* livre.

O tema da preservação dos documentos digitais está cada vez mais presente entre os criadores e gerenciadores de coleções digitais. Para aqueles que começam a enfrentar a possibilidade de não poder recuperar informações produzidas há 5 ou 10 anos, é vital contar com estratégias que garantam a persistência desses dados.

No caso específico das informações científicas, a importância do registro e arquivamento dos dados produzidos durante a pesquisa é fundamental. Desde que a internet se estabeleceu como uma plataforma de comunicação entre cientistas e para a localização de informações relevantes para o desenvolvimento da ciência têm surgido novas formas de organização de informação digital, contando com o respaldo de instituições de ensino e pesquisa para sua manutenção e uso permanente.

As técnicas de preservação que eram conhecidas pelas bibliotecas e pelos centros de informação se modificaram e encontraram novas práticas num contexto de rede de computadores, em que a informação não fica apenas num só lugar. A preservação da informação em formato digital precisa de um conjunto de práticas técnicas e de gerenciamento que mudam constantemente.

Com o surgimento de pacotes de *software*, pode-se afirmar que a área de preservação digital está chegando à maturidade. Algumas dessas soluções saíram das áreas da computação e da ciência da informação, como ferramentas e serviços para bibliotecas e arquivos. Entre os sistemas mais citados estão o DSpace, o Fedora, o LOCKSS (*Lots of Copies Keep Stuff Safe*), o PANDAS, o OCLC *Digital Archive*, e o DIAS (*Digital Information Archive System*)[2] da IBM. Todos eles procedem de organizações de diferentes perfis: provedores de serviço para biblioteca, grupos de pesquisa de bibliotecas especializadas, equipes de repositórios universitários e bibliotecas nacionais trabalhando em parceria com empresas terceirizadas, mostrando a necessidade de integração entre todos os interessados no assunto.

Para a informação científica, a preservação dos documentos nascidos digitais ou que são versões de documentos analógicos tem caráter de dependência das tecnologias existentes, mas também, a necessidade de reformulação das práticas tradicionais de aquisição, armazenamento e acesso.

Pretende-se mostrar neste estudo os aspectos da preservação da informação científica, que têm trazido um conjunto inovador de práticas e de decisões gerenciais que não eram contempladas no passado. A ciência precisa da preservação dos resultados das pesquisas desenvolvidas por instrumentos confiáveis para que esses resultados possam ser utilizados futuramente por outros cientistas.

Os subsídios para esta pesquisa vieram da literatura sobre o item preservação, que destaca alguns temas tratados com freqüência durante o desenvolvimento das principais iniciativas internacionais.

Apesar da crescente relevância que os meios e técnicas digitais de gerenciamento da informação digital vêm assumindo no mundo, existe pouca

---

[2] Muitas Cópias Deixam as Coisas Seguras; Arquivo Digital; Sistema Digital de Arquivamento da Informação.

avaliação dos fundamentos teóricos e metodológicos das estratégias de preservação digital.

A literatura da área, em sua ampla maioria, tem sido produzida por pesquisadores preocupados com a herança cultural e artística da humanidade. Há necessidade de um corpo de conhecimentos ligado à preservação de materiais digitais, compreendendo o fluxo da informação, o processamento, o acesso e o uso otimizado. A carência de trabalhos científicos avaliando as tecnologias de preservação de coleções importantes de Ciência e Tecnologia (C&T) tem sido apontada por raros profissionais da informação (GLADNEY, 2006).

Estudos realizados enfocam alguns métodos de preservação de objetos digitais e mostram a sua limitação ante o enorme crescimento desses registros. Tanto nas áreas da arquivística, quanto da ciência da informação, existe ainda o desafio de estabelecer padrões, políticas e estratégias para as melhores práticas de arquivamento digital.

Atualmente, vários estudos vêm centrando-se no modelo de preservação digital das bibliotecas e repositórios digitais, enfocando a necessidade de futuros usuários disporem de materiais autênticos e certificados por instituições reconhecidas. A descrição em metadados de todos os detalhes que expressem a história de criação de um objeto digital está sendo considerada uma metodologia que pode garantir a autenticidade de um registro eletrônico.

No âmbito nacional, a definição das políticas, obrigações e metodologias mais apropriadas à preservação dos documentos eletrônicos deve levar em consideração a implementação de modelos, a fim de verificar se eles atendem às expectativas dos usuários de instituições nacionais de pesquisa. Recomenda-se, também, comprovar se eles estão em concordância com os padrões internacionais já testados, que promovem a longo prazo o arquivamento digital da produção científica. Tal discussão não pode ficar isolada ou postergada. Ela é parte de um tema de debate internacional que, certamente, deve incluir o Brasil.

Nesta pesquisa pretende-se identificar o conjunto de critérios de preservação digital a serem propostos às instituições ligadas ao MCT. Esse conjunto não foi testado nessas organizações, mas analisaram-se as percepções dos responsáveis pelos acervos digitais nessas instituições, bem como o estado das coleções digitais dessas unidades de pesquisa.

1.2 DEFINIÇÃO DO PROBLEMA

O problema de investigação do qual parte a presente pesquisa é: verificar quais critérios de preservação digital, usados na construção sistemas de gestão da informação científica e tecnológica e reconhecidos internacionalmente, podem ser adotados pelas instituições de pesquisa do Ministério da Ciência e Tecnologia do Brasil.

Levar-se-á em conta a compreensão da necessidade de preservação dos responsáveis pelos acervos digitais nas unidades de pesquisa governamentais, identificando os riscos da ausência de soluções adequadas para a possibilidade de perda dos registros em formato digital.

Na construção dos princípios e práticas relacionadas com o gerenciamento da informação, as pesquisas na área de ciência da informação estão sempre procurando conhecer as novas tendências nos espaços de informação e assinalando particularidades sociais e culturais. A ciência da informação tem nas práticas de preservação de documentos digitais um objeto de estudo que engloba todas as tarefas envolvidas no fluxo informacional que, após serem analisadas pelos pesquisadores, podem chegar a ser pontos focais de originalidade da área.

A preservação de documentos em formato eletrônico é uma das principais preocupações de arquivistas e bibliotecários. A própria definição do que seria a preservação digital não tem sido amplamente disseminada (ver item 2.2.1). Isso obriga qualquer tentativa de trabalho sobre o tema a ter de identificar o que é realmente um objeto digital e o tipo de processo gerencial que ele sofre.

A complexidade do assunto leva consigo o tratamento de temas ligados a ele, como são os aspectos tecnológicos, organizacionais e gerenciais. A heterogeneidade do tema inclui as formas que o objeto digital pode chegar a apresentar, provocando uma mudança no gerenciamento de acervos tradicionalmente identificados como estáticos, para uma realidade multimídia, dinâmica e interativa. O futuro parece não vislumbrar ainda uma definição de como esses documentos virão a ser acessados.

Fala-se na preservação digital como um problema complexo, porque existem muitas variáveis a ser consideradas. Entre essas ramificações, estariam as

de ordem conceitual, as normativas, as jurídicas, as administrativas, as operacionais e as de segurança e integridade. Especificamente as operacionais envolveriam as relacionadas com o preparo do pessoal técnico e as tecnologias aplicadas, sem deixar de lado os usuários (SAYÃO, 2005).

Segundo Lyman (2002), 93% da informação que está sendo criada nos anos recentes já se encontra em forma digital ou “nasceu digital”. O contexto dessa informação está definido pela natureza do suporte: ele é inerentemente frágil, fácil de corromper ou alterar, requer para seu acesso tecnologias que mudam constantemente e tem o mesmo valor na preservação que os materiais tradicionais. No último aspecto, o tempo dedicado à preservação é menor.

Segundo Hanna (2001) e Wiggins (2001), existem várias maneiras de perder os registros digitais, entre elas, reorganizar o conteúdo (não usar identificadores de páginas da Internet permanentes), não valorização da história pelos responsáveis; manter uma mídia ultrapassada e com sistemas e *hardware* obsoletos; esquecer os *softwares* e os manuais; perder os metadados importantes com suas formas de indexação e dependências; não proteger a autenticidade; e ignorar os custos da preservação.

Os bibliotecários e arquivistas sentem a necessidade de preservar esses registros ao longo do tempo, mas eles se encontram indecisos em relação ao tipo de estratégia a adotar. O questionamento inicial de como preservar os documentos digitais e seu conteúdo continua vigente. No caso específico deste trabalho, o conteúdo é a informação científica e tecnológica. Pretende-se identificar as abordagens com relação à preservação digital, discutir suas propostas, documentar as perspectivas publicadas e reunir as opiniões dos responsáveis por acervos dessa natureza no Brasil.

Propor critérios para implementação de sistemas de gestão de preservação digital da produção técnico-científica no Brasil significa tentar responder às seguintes questões:

1. Como uma percepção da preservação digital se transforma em critério para um modelo/sistema de referência?
2. Como podem ser avaliados os benefícios conseguidos na aplicação dos critérios de preservação digital num sistema de gestão da informação científica e tecnológica?

3. Qual é o estado atual das atividades de preservação dos documentos em arquivos digitais no Brasil?
4. Como podem ser analisadas as percepções e as ações relacionadas às necessidades de preservação digital da informação científica?
5. Quais são as necessidades e expectativas dos responsáveis pelos acervos digitais de informação em ciência e tecnologia (IC&T)?
6. Quais seriam os critérios constituintes e o modelo de implementação de um sistema de preservação digital de informação técnico-científica?

Ao enfocar na implementação das práticas de preservação digital, esta pesquisa objetiva apresentar os critérios que identificam as percepções e as ações relacionadas com a sua implementação, assim como avaliar a arquitetura das soluções propostas, com as quais se tenta assegurar que os objetos digitais considerados relevantes para a C&T estejam armazenados de forma a garantir a perenidade dos seus conteúdos.

Uma avaliação dos benefícios conseguidos pelas principais iniciativas internacionais de preservação digital servirá na formulação dos critérios que sustentem a implantação de um sistema, identificando as variáveis que caracterizam este processo. Constituem interesse desta pesquisa as iniciativas nacionais e a visão de seus responsáveis que podem interferir ou não no entendimento do significado e propósito da preservação digital, assim como no efeito e extensão das metodologias adotadas e nas ações das instituições.

Com este estudo, objetiva-se mostrar o alcance das estratégias e propor os critérios apropriados para a preservação de objetos digitais, gerando, assim, subsídios para as iniciativas nacionais relacionadas à manutenção e controle dos acervos digitais que sejam de importância estratégica para o desenvolvimento científico e tecnológico do país.

Para fins desta pesquisa, a ser efetuada no âmbito da ciência da informação, o esforço investigativo a ser empreendido estará em reunir conhecimento sobre as necessidades de critérios de preservação digital no contexto do gerenciamento da informação científica e tecnológica.

1.3 OBJETIVOS

1.3.1 Objetivo geral

O objetivo geral da pesquisa é formular um conjunto de critérios necessários à implementação de sistemas de gestão da preservação de informação científica em documentos digitais. Pretende-se determinar as necessidades de preservação, as tecnologias mais adequadas, os processos de inserção, manutenção, recuperação e acesso aos documentos eletrônicos. A proposta tem como base os critérios descritos na literatura especializada da área da ciência da informação.

1.3.2 Objetivos específicos

A elaboração da proposta de critérios para a construção de um sistema de gestão de preservação digital da informação científica e tecnológica envolve os seguintes objetivos específicos:

- OE$_1$) caracterizar o estado das práticas de preservação digital da informação técnico-científica;
- OE$_2$) identificar na literatura da área os critérios para sistemas de gestão de preservação digital de informação científica e tecnológica;
- OE$_3$) identificar as necessidades de preservação digital da informação técnico-científica das unidades de pesquisa do Ministério da Ciência e Tecnologia;
- OE$_4$) resgatar os conceitos e aplicações usados nessas organizações que caracterizam suas práticas de preservação, tendo por parâmetro as condições locais que interfiram no entendimento do significado e propósito da implantação das metodologias de preservação de documentos digitais.

2 REVISÃO DE LITERATURA

Com o objetivo de caracterizar a preservação digital da informação técnico-científica, sob a perspectiva da ciência da informação, procurou-se mapear os fundamentos teóricos e metodológicos das principais iniciativas empreendidas. A comunidade da ciência da informação vem realizando esforços para introduzir e definir as práticas e princípios arquivísticos que sustentem práticas consolidadas de preservação e acesso de longo prazo[3] a documentos em acervos digitais.

A revisão da literatura apresenta uma análise descritiva das principais conclusões formuladas desde a segunda metade da década de 1990 sobre o tema preservação digital e o acesso à informação científica em documentos digitais de longo prazo. São reflexões desenvolvidas por pesquisadores da ciência da informação e de outras áreas, que levam em consideração algumas concepções que permitirão compreender a preservação digital dentro do ciclo de vida da informação, influenciadas por determinadas práticas sociais contextualizadas nas duas últimas décadas.

O desenvolvimento dos trabalhos que foram publicados desde 1996 tem como característica principal o fato de se ter como contexto a disponibilização *on-line* dos acervos de diferentes tipos de instituições de ensino e pesquisa. A internet permitiu que as práticas de arquivamento eletrônico se expandissem de tal forma, que os bancos de dados e coleções digitais ficassem acessíveis a um número maior de pessoas e métodos de arquivamento. A preservação digital começou a ser chamada como tal a partir desse ano.

A produção bibliográfica consultada se derivou principalmente de projetos e iniciativas internacionais. Chegou-se assim a um conjunto de fontes de informação composto pelos seguintes itens:

a) páginas *web* de projetos e iniciativas de instituições internacionais;
b) catálogos de bibliotecas e repositórios digitais;
c) bases de dados *on-line* de periódicos das ciências da informação e computação;

[3] Considera-se preservação digital de longo prazo o armazenamento, a manutenção e o acesso dos recursos digitais, usualmente conseqüência da aplicação de uma ou mais estratégias de preservação digital, incluindo a migração de dados ou a emulação tecnológica (RUSSELL; SERGEANT, 1999).

d) referências bibliográficas de trabalhos em eventos que tratavam de assuntos relacionados com a preservação digital.

Foram contemplados documentos oficiais tornados públicos em forma de manuais e relatórios, assim como livros, capítulos de livros, artigos avaliados pelos pares e trabalhos apresentados em eventos. Foi constatado que a produção científica sobre o tema no Brasil é reduzida[4]. Especificamente neste capítulo, as fontes citadas são algumas das analisadas no levantamento bibliométrico da produção científica sobre o tema a ser descrito no capítulo 4.

Neste capítulo serão mencionadas as principais definições usadas para se referir às atividades que envolvem os documentos, após a digitalização do objeto físico ou com objetos nascidos já em formato digital; as estratégias adotadas para atender às necessidades dos responsáveis pelos acervos digitais, aquelas que garantem o acesso constante e a preservação de materiais digitais; passando em seguida, a mencionar a variedade de sistemas e soluções tecnológicas que têm sido testados e que estão sendo considerados pela comunidade científica como padrões de preservação digital.

Por último, serão citados os projetos e políticas que estão sendo implantados em instituições que trabalham com informação científica. Essa identificação das tendências das atividades de preservação digital contribuirá na estruturação do conjunto de critérios necessários para a construção de sistemas de preservação digital aplicáveis à realidade brasileira.

2.1 CIÊNCIA E PRESERVAÇÃO DIGITAL

O conhecimento científico pode ser obtido por meio de vários métodos. O objetivo de todos eles é construir teorias que permitam a explicação, predição e controle dos eventos de seu domínio. É o cruzamento entre conceitos teóricos que pretendem explicar fatos conhecidos e, ocasionalmente, a descoberta de fatos novos e suas regularidades.

[4] A análise de alguns trabalhos que condensavam as idéias gerais sobre o assunto deu como resultado a publicação de artigos do autor em algumas revistas da área (MÁRDERO ARELLANO, 2004; MÁRDERO ARELLANO; ANDRADE, 2006), trabalhos em eventos (MÁRDERO ARELLANO; BOERES, 2005; MÁRDERO ARELLANO, 2007) e em trabalhos para disciplinas do programa de pós-graduação da UnB, entre 2004 e 2006.

O progresso técnico-científico de uma sociedade pode ser experimentado nas atividades diárias e nas necessidades humanas às quais ele serve. A prática científica unida ao avanço tecnológico confere às sociedades graus de controle sobre o ambiente. Assim como a tecnologia, a ciência se desenvolveu para atender às necessidades humanas (GRAY, 2005).

A tecnologia tem estado presente na vida humana de várias maneiras: ela transforma e materializa percepções do mundo e desejos individuais. No contexto atual, a tecnologia chega a níveis de ingovernabilidade, e novos conhecimentos são requeridos para solucionar os problemas decorrentes do seu uso. O antropólogo Darcy Ribeiro em sua obra *O processo civilizatório,* menciona o conceito de "revolução tecnológica" para indicar que "[...] certas transformações prodigiosas no equipamento da ação humana sobre a natureza correspondem a alterações qualitativas em todo o modo de ser das sociedades [...]" (RIBEIRO, 1975, p. 34). As revoluções tecnológicas modelam um novo espaço-tempo, estruturando inovadores conceitos culturais e sociais.

Nesse avanço vertiginoso da tecnologia, propiciado pelo crescente conhecimento científico, as sociedades contemporâneas esperam contar com o domínio de recursos digitais para garantir o acesso de longo prazo a grandes estoques informacionais. A nossa dependência da tecnologia moderna nos conduz ao risco da perda dos registros materiais da investigação racional.

Também, esse avanço tecnológico e a produção do conhecimento científico na fase da industrialização da pesquisa, que se inicia na primeira metade do século XX, colocaram em evidência as instituições encarregadas pela preservação e disseminação da informação científica e tecnológica (IC&T)[5].

Com o advento da internet e o desenvolvimento de tecnologias para o acesso à informação científica, começaram a ser testadas soluções mais orientadas para o gerenciamento e armazenamento da informação em formato digital. Uma das questões que têm surgido em torno das novas tecnologias de informação e comunicação é a preservação digital de longo prazo, tema que tem sido tratado sob

[5] IC&T, Informação Científica e Tecnológica: é a informação que os cientistas e as organizações de pesquisa e desenvolvimento (P&D) precisam para desenvolver suas atividades, sendo importante no estabelecimento de elos entre a geração de conhecimentos, seu uso e absorção nas diferentes esferas da economia e da sociedade. Configura ainda as demandas pelas interfaces da produção científico-tecnológica com o Estado e suas instâncias decisórias, no planejamento e gestão de C&T, destinadas a ampliar a participação da cidadania e suas expressões organizadas nos processos de elaboração de políticas públicas (IBICT, 2001).

diferentes enfoques. Os estudos sobre essa temática têm avaliado os obstáculos e benefícios conseguidos na manutenção de serviços e controle de recursos de informação digitais.

Para Chilvers (2000), a confiança que a sociedade acadêmica tem depositado em certos recursos digitais reconhecidos e qualificados pelos pares leva a certa urgência na busca por estratégias para desenvolver, gerenciar e preservar conteúdos digitais. Ignorar isso, segundo a autora, pode levar à perda de dados únicos e criar grande dispêndio financeiro, de tempo e recursos humanos para recriar tais dados, sem falar dos que já "nasceram digitais", que podem também ser perdidos.

A principal meta de quem deseja fazer com que os documentos da ciência consigam permanecer no tempo é colocar os conteúdos em lugares com algum tipo de garantia de acesso de longo prazo. Essa premissa é mais urgente no caso das versões eletrônicas da chamada ciência eletrônica (*e-science*).

A ciência eletrônica tem como meta aproveitar a rapidez das redes computacionais permitindo o uso de laboratórios virtuais, novos métodos de trabalho colaborativo em rede para permitir o avanço da ciência. As atividades relacionadas com a ciência eletrônica requerem um gerenciamento digital de entrada e saída de dados via simulação de testes e grande volume de informações sendo distribuídas e usadas massivamente.

Disponibilizar resultados de pesquisas científicas na *web* tem levado à criação de iniciativas de arquivamento financiadas por governos no Japão, Canadá, Estados Unidos e Inglaterra (NATIONAL SCIENCE FOUNDATION, 2005). Especialistas em preservação digital esperam que, com a adesão a esse tipo de iniciativa, possa construir-se a base de um sistema distribuído de arquivamento e preservação de dados, chegando à criação de serviços de curadoria digital (POTHEN, 2001). O processo de conexão entre a "*e-science*" e a necessidade de preservação dos dados digitais estariam baseados nos repositórios institucionais, que proveriam o suporte para as atividades de preservação e curadoria (MESSERSCHMITT, 2003).

A criação de repositórios institucionais dentro dos princípios do Movimento do Acesso Livre[6] serve para facilitar o autodepósito das versões dos trabalhos (avaliados ou não) dos autores, permitindo um acesso rápido para aqueles que não possuem condições de pagar uma assinatura. O objetivo principal dos periódicos e repositórios de acesso livre é aumentar o impacto e permitir o acesso imediato aos resultados de pesquisa.

A indicação do Movimento do Acesso Livre é não induzir os autores e instituições a construir repositórios com a finalidade de preservação de longo prazo dos originais, mas apenas para garantir o acesso atual e no breve futuro aos conteúdos das pesquisas. Aos editores e responsáveis pelas bibliotecas assinantes ou depositárias, corresponderia a construção de revistas de acesso livre, com políticas claras de autodepósito para os autores, bem como dedicar-se à preservação dos originais e à construção de repositórios digitais confiáveis.

O Movimento do Acesso Livre à informação científica propõe que a preservação dos originais não está sob a responsabilidade dos autores nem das instituições de ensino e pesquisa. A única opção que esses autores têm é a de permitir que as versões dos seus originais sejam autodepositados nos seus repositórios institucionais, quer dizer, a sua preocupação de preservação é o acesso de longo prazo, mas não a preservação dos documentos originais.

Para os idealizadores do Movimento do Acesso Livre, os cientistas e as instituições a que eles pertencem precisam promulgar mandatos de autodepósito, para permitir o acesso imediato às versões e cópias dos originais aprovados e publicados em periódicos científicos. Esse trabalho de autodepósito seria o resultado da consciência da importância do acesso livre à informação científica imediata e sem fronteiras financeiras.

Para Harnad (2003), o problema da preservação dos documentos da ciência tanto em papel como em outros formatos está nas mãos dos editores e dos responsáveis pelas bibliotecas depositárias, quem têm a guarda das versões originais. Os originais em PDF ou *Extensible Markup Language* (XML) são de propriedade das editoras e das bibliotecas que assinam as revistas ou que foram designadas como depositárias das publicações científicas. Similarmente, o problema

[6] Movimento do Acesso Livre: surgiu no final da década com o propósito de derrubar barreiras de ordem econômica, devido ao elevado custo das assinaturas dos periódicos científicos. Mais informações podem ser obtidas por meio do blog http://kuramoto.wordpress.com

da preservação digital não inclui versões que os autores e leitores guardem em suas instituições e *sites* na Internet. Elas não têm o mesmo *status* das versões publicadas pelas revistas científicas. No caso das bibliotecas, se elas assinaram uma dessas revistas, podem ser consideradas bibliotecas assinantes ou depositárias.

A partir desse principio, o Movimento do Acesso Livre deixa clara a orientação de que a preservação digital deve estar dirigida aos originais dos autores publicados nas revistas científicas. A preservação digital dos originais não é um dever dos autores e das instituições, mas uma ação dos editores de periódicos científicos.

Temos assim a separação entre os propósitos do Movimento do Acesso Livre e os da preservação de longo prazo dos artigos publicados em periódicos científicos. O Movimento do Acesso Livre preocupa-se com os materiais suplementares dos documentos publicados, auxiliando os pesquisadores a usar e citar as versões oficiais que são propriedade das editoras. Para os fundadores desse movimento seria um desserviço induzir o autodepósito pelo argumento da preservação[7].

Em 2006, estatísticas mostraram que 85% dos pesquisadores ainda não depositavam nenhuma versão de seus trabalhos em repositórios institucionais. Isso poderia ser uma evidência de que os autores não estão preocupados com o acesso de longo prazo aos seus resultados. Para muitos deles, as editoras e as bibliotecas continuam quase que implicitamente responsáveis pela guarda desses documentos. Segundo Harnad (2006), esses pesquisadores não estão preocupados nem com o acesso de longo prazo nem com a preservação dos originais.

Nos últimos 10 anos, tem sido grande o desenvolvimento de *softwares* para a construção de repositórios digitais na preservação digital dos demais tipos de documentos produzidos por cientistas. Algumas dessas soluções integram funcionalidades específicas para a preservação digital. Desde 2001 têm sido propostos atributos necessários para que esses repositórios possam ser considerados “confiáveis” na preservação da produção institucional (RESEARCH LIBRARIES GROUP, 2001; RLG/OCLC, 2002).

Quem tradicionalmente tem assumido a responsabilidade de preservar dados científicos em documentos textuais são os editores de periódicos,

[7] *Self-Archiving FAQ on Preservation*: http://www.eprints.org/openaccess/self-faq/#1.Preservation

bibliotecários e administradores de documentos científicos, que possuem forte compromisso institucional e conhecimento técnico, nas tomadas de decisões gerenciais que envolvam a preservação da produção científica.

Nas últimas décadas esses documentos não estão apenas em papel, encontram-se em forma de correspondência eletrônica, arquivos elaborados em processadores de texto e arquivos multimídia. O material digital, seja aquele que foi criado em um computador ou digitalizado, está presente na maioria dos serviços de informação (WEBB, 2000).

A informação científica em documentos digitais pode ser perdida quando ela não recebe a atenção adequada, enquanto os *hardware*, *software,* as mídias e suportes continuam a se desenvolver. No caso da preservação dos documentos digitais autênticos, é necessário que seja mantido também o registro do conjunto de processos que garantem o seu acesso contínuo, confiabilidade e integridade.

2.1.1 Tecnologia e evolução dos documentos

A ciência não ocorreria sem a documentação da sua prática. Não existiria nenhum acesso a essa informação sem a mediação documental. Obter informação significa ter acesso a uma entidade material ou suporte de armazenamento. Uma conquista inestimável da humanidade foi a descoberta do papel, que é considerado, até hoje, um dos suportes mais nobres do conhecimento humano. Depois da invenção do papel, o registro das descobertas sobre a realidade aumentou e permitiu um uso para fins que não eram os que motivaram sua criação.

Os novos suportes possibilitaram a transferência dos conteúdos informacionais para outras mídias totalmente diferentes da original. A dissolução da relação entre a informação e o suporte é cada vez mais acentuada. Essa separação facilita o desaparecimento da informação, que os suportes sejam reutilizados, proporcionando uma reprodução documental dos originais em inúmeras cópias. Na prática documental observa-se, também, uma ruptura cada vez maior entre a mensagem e o suporte original. Toda a parte de codificação da mensagem original exibe a capacidade de reordenamento a cada nova cópia.

A chegada das versões eletrônicas de textos publicados em periódicos científicos alcançou a mesma façanha, dirigindo o desmembramento do documento em inúmeras versões e formatos de apresentação. Os chamados *e-prints* foram definidos como sendo a duplicata digital de um trabalho de pesquisa científica, disponibilizados na internet (*on-line*) para aumentar o acesso ao documento. Os *e-prints* são cópias eletrônicas de trabalhos de pesquisa acadêmica, teses ou dissertações. Mas a característica fundamental deles é que mesmo sendo pré-*prints* ou pós-*prints*, eles foram avaliados pelos pares.

Os *e-prints* estão disponibilizados, em sua maioria, no formato texto criado ou convertido por um *software* processador de textos fácil de usar; são documentos predominantemente armazenados em formatos PDF, HTML, MS Word, Postscript, Rich Text Format, ou Latex/Tex. As propriedades desses documentos estão determinadas pelas possibilidades que o formato digital permitir.

No caso da informação técnico-científica, as novas tecnologias de informação colocaram em evidência o papel dos meios e técnicas digitais de gerenciamento da informação digital. Os dados científicos em formatos digitais são mutáveis devido, principalmente, à constante aparição de novos *softwares*. Alguns dos desenvolvedores de *software* mudam sua arquitetura e perdem algum tipo de compatibilidade com versões anteriores. Seu efeito nas políticas de preservação, que exigem a manutenção de formatos digitais acessíveis de longo prazo, influenciará as ações futuras de seus responsáveis.

Os problemas já começaram a ser enfrentados pelos responsáveis por acervos digitais. São bibliotecários e arquivistas que estão aplicando padrões e normas, estratégias de migração e conversão. São os profissionais da informação que analisam a preservação a partir da necessidade de assegurar seu acesso e recuperação, como fundamento para a pesquisa acadêmica.

Dentro das práticas arquivística e bibliotecária, a preservação digital é um fenômeno do maior interesse. Na perspectiva da biblioteconomia, a preservação da informação é uma atividade contínua, permitindo que o seu acesso seja permanente de longo prazo. Já na perspectiva arquivística, a informação é retida segundo critérios de permanência motivados pelas necessidades do uso para o qual ela foi criada. A diferença de abordagem também inclui uma discussão sobre a importância de arquivar apenas parte da informação, mesmo que ela não tenha um valor de uso comprovado.

Apesar das diferenças metodológicas de tratamento documental entre as subáreas da ciência da Informação (arquivologia e biblioteconomia), muitos conceitos são convergentes, seja pelo fato de a informação ser o objeto tratado, ou pela mudança do ambiente (do físico para o digital) e suporte informacional em questão.

Trata-se de duas perspectivas profissionais diferentes, que determinam a forma como os especialistas são treinados para enfrentar problemas. Bibliotecários pensam em termos de coleções como uma reunião de objetos: um conjunto de livros que estão se deteriorando. A meta é manter os livros ou o conteúdo dos livros disponíveis no futuro.

Os arquivistas procuram mais por atividades que agregam e processos que documentam a tomada de decisão de uma instituição. Suas preocupações são menos lidar com o objeto individual e mais com a captura da agregação da documentação e contexto ao redor dos processos, não de documentos isolados, mas sim de um grupo de evidências que provêem *insight* das decisões e ações. Particularmente no mundo digital, partes do contexto do arquivamento de longo prazo ficam escondidas nos sistemas burocráticos internos e na indefinição de políticas da infra-estrutura técnica das instituições.

2.1.2 Arquivologia e preservação digital

Em 1997, Levacov estimava que apenas cerca de 1% da informação arquivada no mundo encontrava-se em formato digital. Com a digitalização e a crescente produção de materiais eletrônicos, as instituições estão enfrentando as exigências de manter a informação digital facilmente atualizada e sempre disponível (ARMS, 2000). Verifica-se em todo o mundo crescente interesse por parte dos responsáveis por sistemas de informação pelo problema da preservação dos arquivos eletrônicos.

Segundo Santa'Anna (2001) os documentos digitais devem ser preservados pelo organismo que os criou ou por alguma instituição arquivística responsável pela sua guarda permanente. Para Hedstrom (1996), o planejamento

das organizações deve alocar recursos e aplicar métodos e tecnologias para garantir a preservação de seus arquivos.

A perspectiva arquivística da preservação digital parte da compreensão dos limites e significados dos documentos (autenticidade, capacidade probatória, integridade das informações, contexto de produção e manutenção), dando ênfase às tarefas que as organizações e instituições arquivísticas que criam e são responsáveis pela guarda permanente desses documentos devem observar para lidar com objetos digitais autênticos.

Os arquivos administrativos requerem o desenvolvimento de parâmetros de certificação da qualidade dos seus processos e serviços. O chamado "arquivamento" relacionado com o gerenciamento do armazenamento de registros passou a estar unido ao termo mais comum "arquivo", que atualmente se encontra mais identificado com o formato eletrônico de documentos (*e-print, pre-print* e *pos-print*). Para Sant'Anna (2001), é responsabilidade dos arquivos adotar medidas preventivas e corretivas objetivando minimizar a ação do tempo sobre o suporte físico da informação, assegurando sua disponibilidade.

As primeiras considerações sobre a preservação digital partiram da arquivologia. Fatores como tempo e custo de armazenamento de longo prazo foram adicionados a decisões sobre o volume e formato do material a ser preservado. Também foram os arquivistas que propuseram algumas das características das ferramentas necessárias para a proteção, cuidado e manutenção de coleções digitais. Com base no princípio arquivístico da "custódia responsável", essa comunidade foi a primeira a apoiar o desenvolvimento dos repositórios digitais como sendo aqueles que reproduzem os espaços físicos dos arquivos tradicionais (LAZORCHAK, 2005).

A informação arquivística tem sido produzida, distribuída e acessada por meio da infra-estrutura de novas tecnologias de informação e comunicação, há algum tempo. A "materialização" (impressão) dessas informações, quando textuais, dá-se, na maioria das vezes, para que o documento se torne legalmente válido ou para que sua leitura ocorra de forma facilitada para aqueles que se incomodam com longos períodos à frente do monitor.

A validade legal de documentos eletrônicos depende de sistemas que garantam a autenticidade e integridade das informações neles contidos. Uma vez assegurada, os legisladores deverão começar a regularizar esse aspecto. A leitura

será mais confortável, à medida que a tecnologia se desenvolve, considerando que protótipos de “papéis inteligentes” vez por outra são anunciados.

Alguns aspectos diferenciam a preservação digital de documentos bibliográficos dos documentos arquivísticos. Os arquivos não podem ser montados. São criados naturalmente pela própria dinâmica de atividades do seu produtor. As relações existentes entre os documentos são, portanto, essenciais para a recuperação e entendimento das informações que foram registradas e armazenadas.

Em arquivos, a necessidade de preservar indefinidamente determinado tipo de documento está fortemente vinculada ao potencial de esse material ser utilizado posteriormente para uma pesquisa histórica e ou cultural. A arquivologia retomou o princípio de que nem tudo precisa ser preservado indefinidamente. Para Dollar (1994), o conceito de “documento permanente” é relativo, quando os custos de preservação da vida útil dos documentos digitais se sobrepõem aos benefícios da sua retenção definitiva.

Na arquivologia, quando se planeja o gerenciamento de um registro, é importante considerar o contexto no qual ele foi criado:

> A perspectiva arquivística traz um sistema de gerenciamento de conhecimentos registrados que se baseia em evidência. O sistema é fundamentalmente vinculado aos processos e contextos organizacionais e pessoais pelos quais registros e conhecimentos são criados, assim como as maneiras em que os registros refletem estes processos individualmente e coletivamente. (GILLILAND-SWETLAND, 2000, p. 5).

O que é preservado deve ser uma amostra representativa do que foi produzido, de modo a ser possível inferir de forma confiável o todo em relação à parte. Em contrapartida, preservar documentos que ainda possuam valor corrente, isto é, administrativo e ou jurídico, também é uma necessidade de longo prazo baseada numa tabela de temporalidade. Alguns documentos que devem ser preservados por causa deste valor revelam-se destinados a estar disponíveis por décadas.

Desde a década de 1990, alguns projetos de gerenciamento de documentos eletrônicos têm surgido na comunidade arquivística internacional. Segundo Rondinelli (2002), essas iniciativas têm produzido conhecimento para o “bom gerenciamento arquivístico” no contexto das novas tecnologias da informação. Segundo esse autor, contribuições de projetos como os das Universidades de

Pittsburgh e da British Columbia têm proposto um conjunto de requisitos funcionais para o gerenciamento de documentos arquivísticos “contemporâneos”, sem deixar de lado processos arquivísticos tradicionais, como a observância da fidedignidade e da autenticidade.

Na arquivologia, a gestão dos documentos eletrônicos constitui uma área de pesquisa que tem produzido contribuições relevantes relacionadas com a preservação de documentos digitais de “valor permanente” para instituições públicas. Alguns exemplos são a criação do padrão 5015.2 do Departamento de Defesa dos Estados Unidos e o *Australian Standard* 4390-1996 do Arquivo Nacional da Austrália, que evoluiria para o padrão ISO 15489.

A informação arquivística, quando possui valor histórico, deve ser preservada por tempo indeterminado; idealmente deve receber uma ordenação baseada na estrutura organizacional que a produziu, criando fundos, que muito podem revelar sobre o funcionamento interno da instituição produtora (BELLOTTO, 2004). Essa organização difere daquela aplicada aos documentos bibliográficos. Por isso, há a necessidade de observância da natureza da informação digitalizada para que não se cometam equívocos metodológicos.

2.1.3 Ciência da informação e preservação digital

O desenvolvimento de atividades de preservação digital requer um conhecimento refinado das atividades gerenciais das bibliotecas, já que elas envolvem diversos processos, tais como:

a) aquisição de objetos digitais, publicados ou não, em vários formatos;
b) armazenamento de muitos objetos digitais comprimidos em *gigabytes* ou *terabytes* de dados;
c) processamento de grande volume de material, por exemplo, migração de objetos múltiplos para evitar a obsolescência dos formatos;
d) disseminação de objetos digitais para usuários de forma fácil, segura e significativa.

Nos últimos anos estão sendo testados esquemas conceituais que permitem identificar as partes integrantes do processo de preservação digital, assim como sua

padronização em modelos de referência para iniciativas de repositórios e de bibliotecas digitais.

A preservação digital assume diferentes significados, conforme o contexto. Para os profissionais da ciência da informação, por exemplo, ela se refere à infra-estrutura e ao comprometimento institucional necessário para proteger a informação representada digitalmente enquanto para os especialistas da ciência da computação ela seria uma maneira de atenuar a obsolescência tecnológica e aumentar a memória humana.

A compreensão sobre os desafios associados a esse tipo de preservação parece começar a ser mais bem focalizada, mudando da estratégia tecnológica para o tema do "gerenciamento digital" (LAVOIE; DEMPSEY, 2004). Novas questões surgem dessa atual visão, a qual não considera a preservação digital apenas como um processo isolado, mas também como componente de um conjunto agregado de serviços, políticas e especialistas que constituem o contexto do ciclo de vida da informação digital.

Na área da ciência da informação, o uso da tecnologia digital que toma o lugar dos tradicionais meios de preservação, como a microfilmagem, trouxe consigo a preocupação com as normas para o uso das técnicas digitais e sua prontidão na tarefa da preservação de longo prazo (CHEPESUIK, 1997).

A pesquisa sobre sistemas de preservação digital é mais um campo de estudo da ciência da informação. Nele estão refletidas, mais uma vez, as ligações interdisciplinares dessa ciência. Por se tratar de experiências iniciais, esses sistemas demandam abordagens e soluções multidisciplinares. A ciência da informação como protótipo de uma nova ciência está sendo construída por abordagens estratégicas voltadas para a solução ou trato de problemas (PINHEIRO; LOUREIRO, 1995). Ela trata da informação que é transportada, conservada no tempo e afetada pelo "imperativo tecnológico" (SARACEVIC, 1992).

A informação que é representada, transmitida e recuperada por meio de mecanismos de comunicação seletiva é objeto da pesquisa aplicada na ciência da informação, assim como o é, também, o tratamento dos problemas decorrentes do uso das novas tecnologias de informação e comunicação, nos contextos individual ou institucional.

Foi Belkin (1978) quem afirmou que o objeto de investigação da ciência da informação é exclusivamente a informação científica formalmente documentada.

Com o impacto da tecnologia no processamento da informação técnico-científica, foram levantadas novas questões relacionadas com o gerenciamento, a perenidade dos registros do conhecimento científico e sua acessibilidade.

2.2 PRESERVAÇÃO DIGITAL

2.2.1 Preservação da informação em objetos digitais

A preservação é um dos grandes desafios do século XXI. Durante os últimos anos do século XX, as bibliotecas, os arquivos e os centros e institutos de pesquisa e organismos governamentais criavam conteúdo digital relevante. Segundo Cunha (1999, p. 266), duas das funções básicas das bibliotecas estão sofrendo "[...] perigo de extinção: a provisão de acesso à informação e a preservação do conhecimento para futuras gerações [...]".

Muitas coleções digitais importantes estão sendo construídas fora das bibliotecas e dos arquivos por diferentes organizações, ou sendo publicadas diretamente na Internet. Com o aumento da produção de informação em formato digital, tem sido questionada cada vez mais a importância de garantir a sua disponibilização e preservação por grandes períodos de tempo. Os objetos digitais não podem ser deixados em formatos obsoletos para serem transferidos, depois de longos períodos de esquecimento, para repositórios digitais, pois existe o risco de a tecnologia não conseguir recuperar as informações. Esta preocupação envolve tanto os produtores dos dados quanto os órgãos detentores dessa informação.

Para Hedstrom (1996, p. 189), a preservação digital se refere ao "[...] planejamento, alocação de recursos e aplicação de métodos e tecnologias para assegurar que a informação digital de valor contínuo permaneça acessível e utilizável [...]". A preservação digital compreende os mecanismos que permitem o armazenamento em repositórios de dados digitais que garantem a perenidade dos seus conteúdos. Para atingir esse fim, os objetos digitais devem ser compreendidos e gerenciados em vários níveis: como um objeto físico, como uma codificação lógica, como objetos conceituais ou possuidores de significado para os humanos e como

um conjunto de elementos essenciais que devem ser preservados para oferecer aos futuros usuários a essência do objeto (UNESCO, 2003).

A condição básica à preservação digital seria, então, a adoção desses métodos e tecnologias que integrariam a preservação física, lógica e intelectual dos objetos digitais. A preservação física está centrada nos conteúdos armazenados em mídia magnética (fitas cassete de áudio e de rolo, fitas VHS e DAT) e discos óticos (CD-ROM, WORM e discos óticos regraváveis). A preservação lógica procura, na tecnologia, formatos atualizados para inserção dos dados (correio eletrônico, material de áudio e audiovisual e material em rede), novos *software* e *hardware* que mantenham vigentes seus *bits,* para conservar sua capacidade de leitura.

A preservação digital é a parte mais longa e também a última do ciclo de gerenciamento de objetos digitais. Ela permite o emprego de mecanismos que viabilizam o armazenamento em repositórios de objetos digitais e que garantem a autenticidade e perenidade dos seus conteúdos. São necessários, não apenas, procedimentos de manutenção e recuperação de dados, no caso de perdas acidentais para resguardar a mídia e seu conteúdo, mas também estratégias e procedimentos para manter sua acessibilidade e autenticidade através do tempo, podendo requerer colaboração entre diferentes financiadoras e boa prática de licenciamento, metadados e documentação, antes de aplicar ações técnicas.

Ao analisar o ciclo da informação (geração, tratamento, preservação e conservação e, finalmente, difusão da informação), percebe-se que a preservação e conservação necessitam ser revistas no âmbito da informação digital. Até pouco tempo, entendia-se a preservação no sentido de conservar e prevenir os documentos do risco de deterioração. É importante analisar a preservação a partir da necessidade de assegurar o acesso e recuperação da informação científica como fundamento para a pesquisa acadêmica. Aliado a isso, deve-se considerar o crescente uso dos computadores e a quantidade de informações disponíveis em meio eletrônico (algumas, inclusive, somente neste meio).

A natureza dos documentos digitais está permitindo ampla produção e disseminação de informação no mundo atual. É fato que, na era da informação digital, muita ênfase é dada à geração e/ou aquisição de material digital, em vez de manter a preservação e o acesso de longo prazo aos acervos eletrônicos existentes.

Outro conceito associado ao da preservação digital, que surgiu da comunidade arquivística, é o de ciclo de vida do material digital. Nos primeiros

estudos do grupo de trabalho do National Preservation Office e do Joint Information Systems Committee (JISC/NPO) (BENNETT, 1997), encarregado de desenvolver um arcabouço para uma política estratégica de criação e preservação de objetos digitais, o ciclo de vida foi dividido em cinco partes: criação de dados; gerenciamento de coleções e preservação; aquisição, retenção e descarte; gerenciamento de dados; dados de uso. Foi graças a esses trabalhos que ficou esclarecido que os envolvidos (*stake-holders)* têm variados interesses em estágios diferentes do ciclo, assim como a importância que há em considerar o assunto preservação em todas as etapas, e não apenas no final do ciclo, exigindo a cooperação entre as grandes iniciativas.

A partir da formulação dessa estrutura, começaram a ser desenvolvidos guias estratégicos para os responsáveis por todos e cada um dos estágios do ciclo de vida dos objetos digitais. Neles, existem recomendações sobre a valorização de um estágio do ciclo em particular, assim como a relação entre eles, e como as decisões podem afetar as tarefas dos outros responsáveis. Em geral, as tarefas relacionadas com a preservação digital seriam os procedimentos de manuseio e armazenamento da mídia digital, a cópia da informação contida, a migração para novas mídias e a preservação da integridade da informação digital.

Como mostra o Quadro 1, a preservação física continua relevante na mídia eletrônica, ainda que o seu armazenamento haja demonstrado a necessidade de constante atividade de migração para novos materiais que favoreçam sua disseminação e uso (disquete, fita VHS, CD-ROM, DVD).

Requisitos	Documentos Impressos	Documentos Digitais
Preservação física	Relevante	Relevante
Preservação lógica	Pouco relevante	Relevante
Preservação intelectual	Não relevante	Relevante

Quadro 1 – Relevância dos requisitos de preservação

Na preservação intelectual, o foco são os mecanismos que garantem a integridade e a autenticidade da informação nos documentos eletrônicos. Uma

solução tem sido a autenticação digital (*digital time-stamping*[8]*,* DTS). Essa técnica permite que os documentos sejam criptografados no momento em que são depositados (GRAHAM, 1998).

No caso dos materiais impressos, a preservação lógica é pouco relevante, por estar garantida no formato específico em que foram publicados (periódico, livro, entre outros formatos). Na publicação digital, a preservação lógica está associada à necessidade de garantir a conversão dos formatos originais que se tornaram obsoletos ou de custosa manutenção. O Quadro 1 mostra, também, que a importância da preservação intelectual é maior no caso dos materiais digitais devido, principalmente, à capacidade de o objeto digital ser passível de modificação no seu leiaute, apresentação ou interação no formato de publicação. Com isso, a perda do conteúdo intelectual original pode ser declarada inaceitável pelo autor.

Esse requisito envolve a preservação da propriedade intelectual (*Intellectual Propriety Rights*[9], IPR), que tem um significado mais argüível do que na mídia tradicional. O IPR deve considerar não apenas o conteúdo, mas também qualquer ação relacionada ao *software* (cópias, encapsulamento de conteúdo, emulação de *software,* migração de conteúdo) que envolva atividades que podem infringir permissões específicas daqueles que mantêm os direitos. Para Bullock (1999), o IPR é uma das principais barreiras que interferem na preservação dos objetos digitais. Ela enumera os aspectos técnicos da preservação dos documentos digitais da seguinte forma:

- limites do objeto digital;
- presença física;
- conteúdo básico (ex.: arquivo ASCII);
- apresentação;
- funcionalidades;
- autenticidade;
- localização e rastreamento do objeto digital no tempo;
- proveniência;
- contexto.

Esses requisitos especificam os elementos que deverão ser efetivamente mantidos de um documento digital, a cadeia de *bits* que deverá ser recuperada.

[8] Selo de tempo digital. Tradução nossa.
[9] Direitos de Propriedade Intelectual.

Referem-se à capacidade de acesso aos conteúdos dinâmicos independentemente da sua apresentação, componentes de multimídia, hipertextualidade e interatividade. Observar essas condições significa identificar o objeto digital na sua origem e pelas suas dependências de *hardware* e *software*.

Um documento digital pode estar representado em um ou vários objetos digitais, em diversos formatos lógicos, e pode ser suportado por grande variedade de representações físicas. A sua dependência de *software* tem origem em sua natureza binária, armazenada em forma codificada, apenas legível por programas. Sem esses programas, os objetos digitais não podem ser acessados, lidos ou impressos.

Salvar os *bits* de um objeto digital é necessário, mas não suficiente para preservá-lo. É necessário também conhecer os atributos da aplicação na qual ele foi criado e com o qual ele pode ser interpretado. O esforço da preservação engloba também o *software* e o *hardware* no qual o objeto digital pode ser executado, já que eles também podem ficar obsoletos.

A problemática dos objetos digitais aumenta a partir do momento no qual eles são inerentemente digitais, quando não podem ser representados em forma de imagem ou páginas de texto; tentar representar documentos audiovisuais dessa forma significa perder aspectos essenciais dos seus conteúdos e comportamento, como no caso de artefatos interativos e dinâmicos (multimídia, dados meteorológicos em imagens ou gerados automaticamente em JavaScript, cgi, páginas ASP, entre outros). Sem deixar de mencionar simulações de experimentos, bases de dados e realidade virtual interativa.

As soluções propostas para a preservação desse tipo de objeto digital vão desde a arqueologia digital, até soluções mais completas, como a formalização (substituição dos artefatos pelas descrições formais deles), a migração (converter os artefatos em novos formatos repetidamente) e a emulação (rodar os *softwares* originais em *hardware* recriados virtualmente); passando pela reprodução dos objetos em forma de imagens e salvando os componentes principais dos artefatos.

Em geral, um objeto digital é um componente digital necessário para reproduzir um documento eletrônico. Ele não é apenas um arquivo, dentro de uma biblioteca digital, mas está composto de várias partes:

- metadados (descritivos, administrativos, estruturais, de preservação);
- arquivos de preservação;

- arquivos de disseminação.

Para mantê-los organizados e conectados, é necessária uma prática correta de organização de diretórios e projetos de documentação. O lugar ao qual pertencem é o repositório de documentos digitais.

É preciso chamar a atenção para a importância de informar o contexto do objeto digital a ser registrado (e preservado) para que, assim, futuros usuários possam entender o ambiente tecnológico no qual ele foi criado[10]. A preservação dos documentos continua a ser determinada pela capacidade de o objeto informacional servir às utilizações que lhe são imputadas, às suas atribuições que garantem que ele continue a ser satisfatório às utilizações posteriores. Porém no caso específico dos documentos em formato digital, a preservação dependerá principalmente da solução tecnológica adotada e dos custos que ela envolve.

Tanto a arquivologia quanto a ciência da informação prevêem a descrição de documentos digitais como uma maneira de possibilitar o armazenamento e a recuperação da informação. A descrição arquivística e a descrição bibliográfica compreendem o contexto e o conteúdo dos documentos digitais (os padrões de metadados resultantes começam a homogeneizar-se, e espera-se a criação de redes interoperáveis de arquivos e bibliotecas digitais).

Na preservação de documentos digitais, assim como em papel, é necessária a adoção de ferramentas que protejam e garantam a sua manutenção. Essas ferramentas deverão servir para restaurar registros protegidos, prevendo os danos e reduzindo os riscos dos efeitos naturais (preservação prospectiva), ou para restaurar os documentos já danificados (preservação retrospectiva).

O suporte físico da informação, o papel e a superfície metálica magnetizada se desintegram ou podem se tornar irrecuperáveis. Existem, ademais, os efeitos da temperatura, umidade, nível de poluição do ar e das ameaças biológicas; os danos provocados pelo seu uso indevido e o uso regular; as catástrofes naturais e a obsolescência tecnológica. A aplicação de estratégias de preservação para documentos digitais é uma prioridade, pois sem elas não existiria nenhuma garantia de acesso, confiabilidade e integridade dos documentos de longo prazo.

---

[10] Um documento eletrônico é um registro que foi criado (elaborado ou recebido) de forma eletrônica. Um documento digital é aquele que existe em formato eletrônico, mas que pode ou não ter sido criado em formato eletrônico - pode ter sido criado em papel e digitalizado. (CONSELHO NACIONAL DE ARQUIVOS, 2004).

Quando a primeira versão eletrônica de um documento é criada, inicia-se uma longa lista de ameaças inerentes aos documentos digitais. Segundo Greenstein e George (1998), para proteger a informação digital da destruição causada pelo tempo, a obsolescência tecnológica e a deterioração da mídia magnética, precisa-se de métodos apropriados para diferentes tipos de dados e estruturas.

Uma das principais ameaças é o problema da obsolescência tecnológica. *Hardware* e *software* são dispositivos para gerar e acessar informações gravadas que requerem estratégias técnicas diferenciadas para garantir seu funcionamento (FEENEY, 1999). Nesse contexto, os documentos multimídia são os mais comprometidos, pois não existem garantias de que a essência da informação possa ser preservada, se não houver dispositivos que leiam as mídias magnéticas usadas.

A busca por estratégias de preservação digital requer não apenas procedimentos de manutenção e recuperação de dados, no caso de perdas acidentais, para resguardar a mídia e seu conteúdo, mas também estratégias e procedimentos para manter sua acessibilidade e autenticidade através do tempo, podendo requerer colaboração entre diferentes organizações, boa prática de licenciamento, aplicação de padrões de metadados e documentação.

Todas as tarefas aplicadas às coleções digitais têm sido objeto de ações por parte da comunidade da área da ciência da informação, que tem entre seus desafios viabilizar a recuperação e o processamento da informação no futuro (FUNARI, 2004). Especificamente a preservação digital está sendo observada como o resultado do uso de normas nos processos de arquivamento dos documentos digitais. O uso de repositórios digitais que adotam essas normas está sendo apontado como uma melhoria na qualidade da produção da informação digital a ser preservada pelas organizações.

Os especialistas da área que trabalham com informação em formatos digitais estão elaborando normas necessárias para armazenar e compartilhar em repositórios digitais esses materiais, assim como buscam a formulação de políticas institucionais de preservação.

Segundo Jones (2003), existem três tipos de criadores de documentos digitais: 1) aqueles que não acreditam que os repositórios digitais tenham a habilidade de preservar seus materiais; 2) aqueles que desconhecem que esses repositórios existem; 3) aqueles que desejam que seus materiais digitais sejam preservados, mas para os quais tais recursos são inexistentes. Como Jones aponta,

existe relutância por parte desses criadores de documentos digitais em lidar com a responsabilidade da preservação desses materiais, além do desconhecimento das necessidades reais de preservação dos seus acervos digitais.

Se os criadores de materiais digitais originais não compreendem as implicações das suas ações em termos de viabilização a médio e longo prazo dos materiais criados, eles não serão capazes de proporcionar o suporte e o acompanhamento necessários com as infra-estruturas técnica e organizacional que possam dar assistência e facilite um gerenciamento eficiente da preservação digital.

2.2.2 Necessidades de preservação digital

O estudo da preservação digital tem passado do reconhecimento da sua necessidade e da sua definição, para abordagens mais técnicas e propostas de ações mais claras. O foco da preservação digital mudou da necessidade de uma ação imediata de “resgate” dos materiais ameaçados, para a conscientização de que perpetuar os materiais digitais ao longo do tempo envolve a observância de um conjunto de práticas gerenciais que integrem os sistemas de preservação digital.

No início, as práticas relacionadas com a preservação digital estavam baseadas na idéia de garantir a longevidade dos arquivos, entretanto essa preocupação está agora centralizada na ausência de conhecimento sobre as estratégias de preservação digital e o que isso poderá significar na necessidade de garantir a longevidade dos arquivos digitais.

O desafio é muito mais um problema social e institucional do que técnico, porque, principalmente para a preservação digital, depende-se de instituições que passam por mudanças de direção, missão, administração e fontes de financiamento. Muitos materiais publicados digitalmente são resultados de serviços de informação disponibilizados por organizações que adotam alguma infra-estrutura tecnológica. Essas instituições levam em consideração aspectos legais e culturais que atingem a oferta desses serviços orientados a atender às necessidades de determinados usuários. Porém muitas vezes essa preocupação não é a mesma que têm os produtores das novas tecnologias.

Em 1996, Waters e Garrett referiam-se àqueles que seriam os responsáveis pela preservação digital como “[...] a linha de frente na defesa contra a perda da informação digital valiosa [...]” (p. 21). Para esses detentores de acervos digitais, é cada vez mais imperiosa a necessidade de contar com mecanismos que garantam a preservação de seus documentos em formato digital. Especificamente, essa preocupação parte das comunidades responsáveis pelas bibliotecas e pelos arquivos, para as quais o desenvolvimento de padrões e de mecanismo legais para lidar com arquivos eletrônicos precisa de estratégias metodológicas bem definidas.

A necessidade de preservação digital, do modo como é percebida por vários autores, reforça a importância do planejamento na criação e gerenciamento de objetos digitais. A comunidade arquivística tem ressaltado a necessidade da preservação da integridade da informação digital. Para Elkington (1997), em citação de Chilvers (2000, p. 27), a necessidade de preservação está presente na motivação das ações empreendidas para garantir a integridade do conteúdo digital, tais como:

a) a percepção do uso corrente ou antecipado dos recursos que são criados;

b) o interesse da pesquisa;

c) as imposições legais;

d) as imposições de negócios;

e) a missão pública;

f) os propósitos para os quais os conteúdos são criados.

Chilvers (2000) menciona a prioridade da preservação digital como a necessidade de seleção e avaliação dos custos baseados no ciclo de vida de um objeto digital. Boeres (2004), após uma investigação junto às bibliotecas universitárias brasileiras, identificou a necessidade de preservação digital no momento da seleção das coleções. Para ela, a questão de seleção de documentos digitais deve ser observada na tomada de decisão, na qual seriam reavaliados os critérios da seleção e detectada a necessidade de uma política de seleção.

No caso das bibliotecas universitárias, a seleção do que será preservado deve estar de acordo com a missão da unidade de informação e a necessidade de seu corpo acadêmico. A necessidade de preservar tem um papel de destaque na garantia do acesso e recuperação de informação, a fim de fundamentar a pesquisa acadêmica.

Neste âmbito, Funari (2005, p. 8), ao mencionar a necessidade da gestão da preservação dos documentos eletrônicos gerados por uma instituição de ensino superior brasileira, enumera as ações a serem empreendidas como resposta a necessidades administrativas e de pesquisa acadêmico-científica. Segundo ele:

> Para além da manutenção da integridade física, é necessário prever uma série de procedimentos, em constante reavaliação, de transposição de dados de equipamentos e programas informáticos antigos para novos.

Verifica-se, nas instituições de ensino e pesquisa, um aumento na necessidade de modernização dos serviços de informação que elas prestam. A necessidade de digitalizar e ou copiar seus acervos é um indício de necessidade de preservação digital.

Para Boeres (2004), a necessidade de preservação digital é percebida também pelos gestores dos acervos, pois seu objeto de trabalho está em constante mudança e existe pouco investimento em cursos de atualização. A natureza dos documentos digitais permite a reprodução do documento em inúmeras versões, e a sua preservação precisa de novos conhecimentos e ferramentas.

Com base nos estudos dos autores consultados, as necessidades de preservação podem ser agrupadas da forma como mostra a Figura 1.

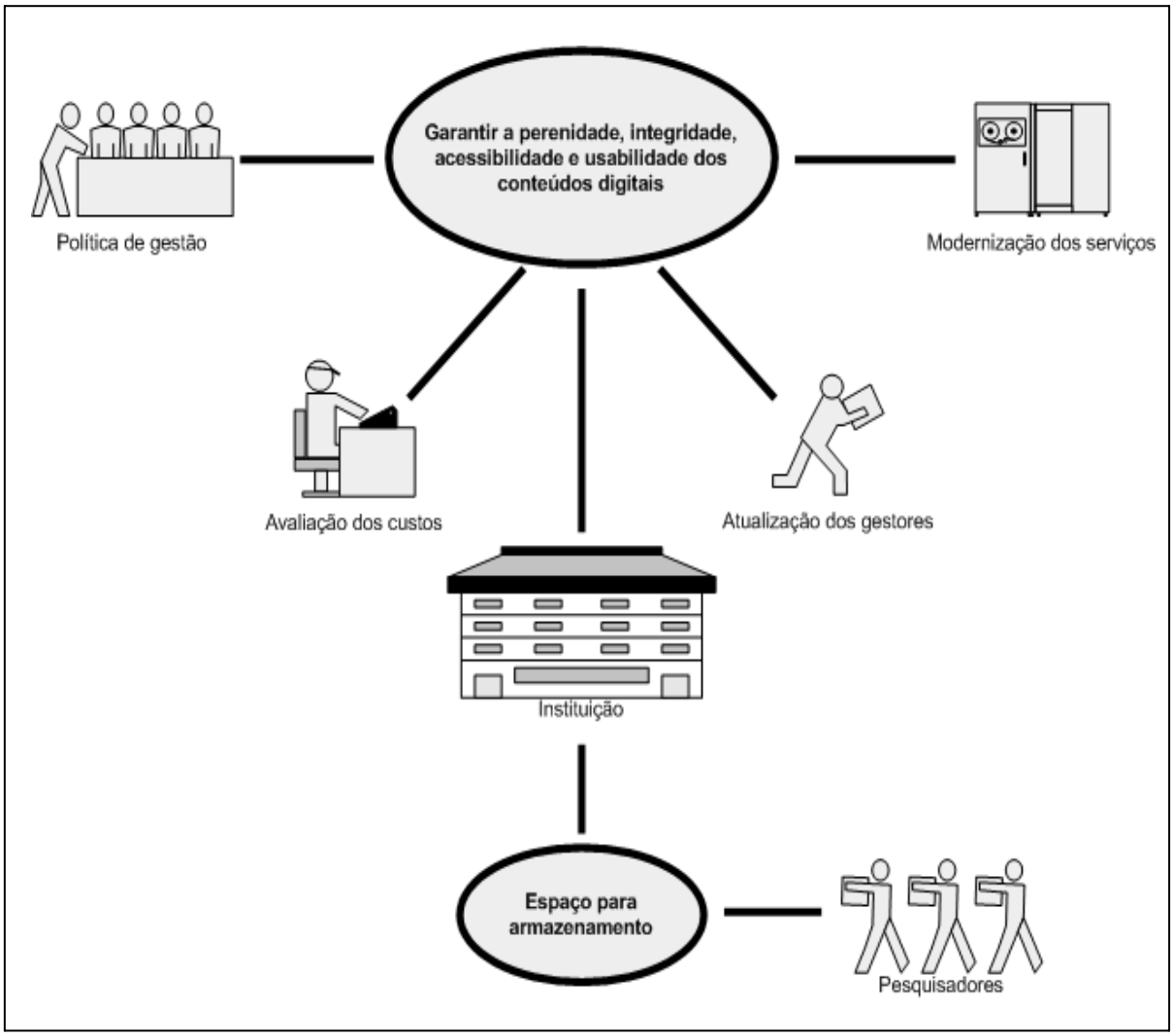


Figura 1 – Contexto das necessidades de preservação digital da informação científica

O ambiente onde a informação científica é criada e disponibilizada está sendo afetado pela rapidez dos meios de comunicação e de redes, pelos ambientes de aprendizagem virtuais e pelo trabalho colaborativo. Cada vez mais sistemas de informação para armazenamento e recuperação de informação são requeridos para conseguir acompanhar esse avanço na produção de conhecimento científico. Alguns aspectos do contexto em que ocorrem as necessidades de preservação digital incluem conceitos como novas tecnologias, comunicação científica, depósito legal e direitos autorais, entre outros.

Para atingir todo o potencial dos sistemas de preservação digital e aumentar a probabilidade de adoção nas instituições, é preciso uma efetiva identificação das necessidades de preservação. O desafio para os implementadores

desses sistemas é a adoção de políticas, modelos de negócios, estratégias, suporte tecnológico e guias para o desenvolvimento de repositórios digitais confiáveis.

2.2.3 Estratégias para a preservação digital

No início da era digital, a preocupação das organizações com a preservação concentrava-se na estratégia de fazer cópias de segurança (*backup).* Atualmente, em vários países, há urgência na definição das políticas, obrigações e metodologias mais apropriadas para a administração dos recursos digitais. Em nível mundial, as iniciativas se multiplicam (Apêndice C) e novas soluções são testadas por instituições detentoras de acervos de especial relevância para o desenvolvimento científico e tecnológico de seus países de origem.

Muitos projetos e iniciativas têm conseguido estabelecer os benefícios de algumas metodologias ou estratégias. A conclusão a que essas iniciativas chegam é que devem ser usados padrões e é preciso converter os documentos nos formatos abertos, para que eles sejam acessados após a obsolescência dos equipamentos e programas de informática em que foram criados. O uso de padrões abertos permitiria seu estudo e sua conversão para novos padrões.

Alguns dos tópicos relacionados com a preservação digital, citados por vários autores, estão relacionados com a aplicação de métodos e tecnologias para garantir a integridade e acessibilidade da informação (HODGE; FRANGAKIS, 2004; LEVACOV, 1997; THOMAZ; SOARES, 2004). A obsolescência da tecnologia coloca em questionamento todas as estratégias de preservação existentes. A execução de estratégias está condicionada aos conhecimentos do alcance das tecnologias e às limitações orçamentárias nas instituições. Por tais motivos, existe imperiosa necessidade de planejamento da preservação digital de baixo custo.

Rothenberg (1995) afirma que a vida média de uma mídia óptica está em torno de 30 anos. Seu estudo também apresentou outro dado o qual mostra que o equipamento de leitura estará obsoleto em 10 anos. Estudo realizado por Stinson, Ameli e Zaino (1995) apontou que 95% das mídias Kodak *Writable CD* e *Photo CD Media* têm seu tempo de vida estimado em 217 anos, quando conservadas a uma temperatura de 25º C e 40% de umidade relativa. Já no início das mídias digitais

alguns especialistas mencionavam a fragilidade dos suportes eletrônicos (ver Figura 2).

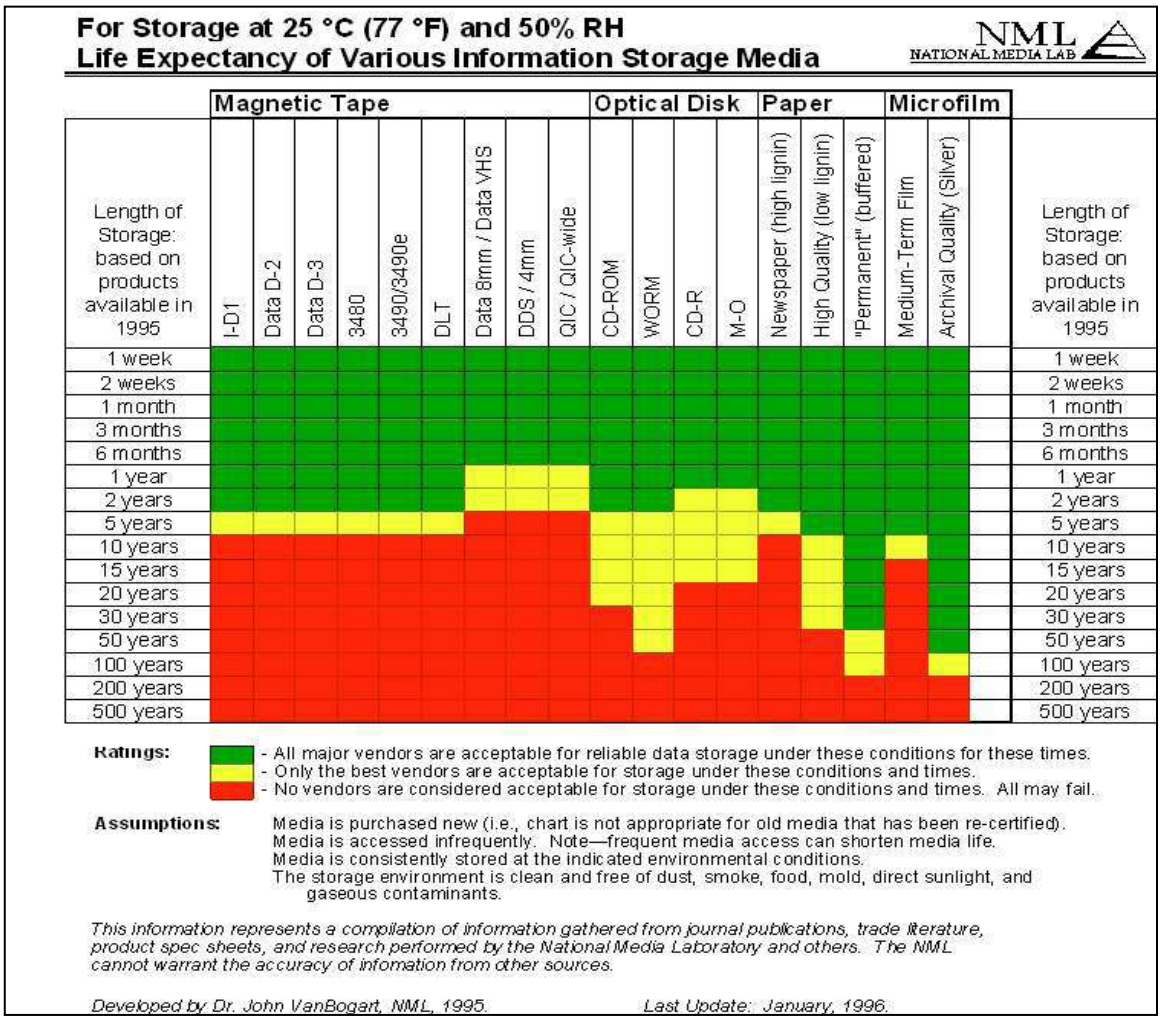


Figura 2 – Expectativa de vida de várias mídias para armazenamento de informação
Fonte: VanBogart (1995)

O acompanhamento das normas e padrões de preservação é cada vez mais difícil devido ao rápido avanço das tecnologias, especificamente no que se refere aos problemas trazidos pelos formatos digitais (MARCUM; FRIEDLANDER, 2003). Para Woods (1998), há três critérios que as novas mídias devem cumprir antes de serem escolhidas para a preservação digital: 1) a mídia deve proteger o documento original; 2) a mídia deve possuir mecanismos de verificação da integridade, de tal modo que o documento não possa ser alterado ou modificado, e 3) a mídia deve ser de acesso fácil. Por conseguinte, também seriam preservados a propriedade intelectual e o *hardware* e *software* usados para acessar o documento.

Autores como Thomas (2003) propõem que a preservação deveria ser apenas do conteúdo intelectual, e que separadamente seria desenvolvido o *software* para ler e acessar os dados, não sendo aplicável aos recursos em mídia interativa.

Dois aspectos importantes relacionados com a mídia digital são a expectativa de vida e a estabilidade. Para Stephens e Wallace (2003), a expectativa de vida de uma mídia é o espaço mínimo de tempo que uma informação está prevista a ser recuperada; estabilidade da mídia é a habilidade que têm várias mídias de reter o conteúdo para uso em determinado período de tempo (Quadro 2).

Mídia	Tipo	Expectativa de vida (anos)	Estabilidade
Papel	ANSI/NISO	200-300	Boa
Microformas	Microfilme	500	Ótima
	Microficha	100	Boa
Mídia Óptica	WORM (write once, read many)	25	Pobre
	CD-R	100	Boa
	CD-ROM	25	Pobre
Mídia Magnética	Disquetes	5	Pobre
	Fitas	10-20	Pobre

Quadro 2 - Expectativa de vida de algumas mídias para fins de preservação digital
Fonte: versão traduzida e adaptada de Stephens e Wallace (2003)

Os dados mostram que a mudança para uma mídia nova significa alguma perda na durabilidade e longevidade dos recursos. No seu estudo sobre as mídias, Stephens e Wallace (2003) citam como exemplo as fitas magnéticas usadas comumente como mídia para armazenamento de dados computacionais inativos, para pelo menos 10 a 20 anos, para mostrar como cada tipo de informação a ser preservada acrescenta algum elemento de avaliação da mídia escolhida. Esses autores também mencionam a durabilidade de uma mídia como outro dos aspectos da preservação digital que têm de ser levados em consideração, pois o *software* requerido para ler é menos durável do que a mídia na qual o registro foi gravado.

Os estudos atuais sobre os problemas relacionados com os *softwares* vêm mostrando que eles superam em número aqueles relacionados com o *hardware*. Para Dukart (2002), os formatos XML (*eXtensible Markup Language*) e o PDF

(*Portable Document Format*) são os mais promissórios, apesar das suas vantagens e desvantagens.

A linguagem XML pode ser considerada como um tipo particular de migração. Ela participa como uma linguagem de enriquecimento de informação sobre estruturas e significado, é um padrão aberto, independentemente da plataforma onde irá funcionar. Favorece a interoperabilidade e pode ser usada como formato de criação de documentos. Pode considerar-se como uma estratégia de preservação por si só. Junto com o formato PDF, a opção XML é complementar na área da preservação de longo prazo.

Nos Estados Unidos, o Goverment Printing Office adotou o PDF como o formato oficial para seus documentos eletrônicos por considerá-lo seguro para futuros projetos de preservação, mas nenhum dos formatos existentes até hoje pode ser considerado imune à obsolescência tecnológica (TEPER; KRAEMER, 2002).

Terrio (2002) defende um estudo mais profundo dos formatos não proprietários, para poder ter argumentos que mostrem o cenário diferente daquele onde as licenças não permitem o acesso ao *hardware* ou ao *software* necessário para representar os dados. Em contrapartida, sistemas como o DSpace do Massachussett Institute of Technology (MIT) disponibilizam uma lista de formatos que são suportados e que continuarão disponíveis e legíveis no futuro (FALK, 2003). Algumas estratégias implementadas continuam considerando o método de adotar *softwares* não proprietários e salvar o conteúdo intelectual dos documentos como o mais apropriado para resolver os problemas da preservação digital.

A publicação de relatórios por parte de grupos de estudo sobre materiais eletrônicos, como o *Studies on the Preservation of Electronic Material* da JISC/NPO *e* o da European Commission on Preservation and Access (ECPA) tem favorecido a construção de um corpo de guias e recomendações para o gerenciamento dos materiais digitais (BENNETT, 1997). As conclusões sugerem a manutenção desse tipo de acervos em formatos, sistemas e mídia padrão. As sugestões envolvem métodos operacionais para materiais específicos e métodos estruturais para atender às necessidades de comunidades específicas de usuários.

A Commission on Preservation & Access/Research Libraries Group (CPA/RLG) mencionou três das principais estratégias para a preservação digital: a preservação da tecnologia, a tecnologia de emulação e a migração da informação (WATERS; GARRETT, 1996). Desde o surgimento desses relatórios em 1995, as

sugestões apontam um ou vários métodos como o processo técnico que melhor garante a preservação. Para a ECPA, por exemplo, seria a emulação (ROTHEMBERG, 1995); já para os detentores do projeto Cedar (*CURL Exemplars in Digital Archives*), o foco não é uma mídia específica, mas o acesso de longo prazo ao conteúdo intelectual do recurso.

Em 1995, a JISC e a NPO (HAYNES, 1997), já afirmavam que a preservação digital precisa ser uma responsabilidade compartilhada, principalmente pelo grande volume de informação digital que está sendo produzido e pela natureza da tecnologia digital. A responsabilidade de que a informação científica e tecnológica seja preservada para benefício das futuras gerações está nas mãos dos criadores desses materiais. Para o desenvolvimento de uma infra-estrutura para a preservação de materiais digitais, é necessário esclarecer a importância dos criadores desses objetos e apontar as suas responsabilidades. Alguns dos criadores de materiais digitais de informação científica e tecnológica não possuem os conhecimentos arquivísticos necessários para realizar essa tarefa.

A literatura especializada indica que os responsáveis pelos acervos devem buscar uma visão sistêmica, segundo a qual a informação esteja resguardada pela preservação. Pesquisadores em várias partes do mundo estão desenvolvendo modelos do que seria a infra-estrutura para a preservação de longo prazo da informação em formato digital. Muitas das iniciativas propostas pelos grandes centros de pesquisa continuam sendo utópicas, mas proporcionam uma base para a discussão sobre a implantação de soluções futuras ao problema (GRANGER, 2002). Para Beagrie e Greenstein (1998), algumas precauções devem ser tomadas para reduzir o perigo da perda dos materiais digitais:

1) armazenar os materiais em ambiente estável e controlável;
2) implementar ciclos de atualização (*refreshment*) para cópia em nova mídia;
3) fazer cópias de preservação (assumindo licenças e permissões de *copyrights*);
4) implementar procedimentos apropriados de manuseio;
5) transferir materiais para uma mídia de armazenamento padrão.

2.2.3.1 Preservação da seqüência de *bits*

De acordo com Rothenberg (2003), o dado e a mídia que suportam a informação devem possuir um nível de funcionalidade representacional que permita a sua reprodução a qualquer momento em que a instituição mantenedora precise recuperar o dado. Cada mídia pode armazenar uma seqüência de *bits* de forma diferente, segundo as propriedades físicas da mídia. A cadeia de *bits* (a seqüência de zeros e de uns que compreende o dado) precisa então ser interpretada, pois toda seqüência significativa de *bits* pode representar qualquer coisa. Os objetos digitais são salvos como coleções de *bits* representando documentos específicos, significativos apenas para o programa que os criou.

A orientação básica com relação às cadeias de *bits* é que eles tenham sido transformados para um formato de mídia neutro ou padrão; que suas propriedades mais significativas tenham sido capturadas e descritas; que exista a possibilidade de retornar em qualquer momento para a cadeia de *bits* original, no caso de alguma solicitação de acesso ao conteúdo intelectual original. Para que a cadeia de *bits* continue tendo sentido, algumas opções têm sido apontadas:

a) manter a cadeia de *bits* original;
b) manter a ferramenta que interpreta a cadeia de *bits*;
c) focalizar a preservação para a ferramenta que interpreta;
d) usar técnicas para a longevidade do *software*.

Os motivos para escolher a estratégia de salvar a cadeia de *bits* vão desde o fato de ser uma escolha que deixa de lado o controle do desenho/criação do objeto a ser preservado, direcionando o foco para a ferramenta que interpreta, até o desenho de uma ferramenta que facilite a sua preservação.

A interpretação da cadeia de *bits* e do dado envolve definições de *hardware* (para cópia e visualização) e *software* ou aplicativos para as mais complexas entidades. Mas todo *software* é interpretado por um *hardware*. Ele, como um processo ativo, executa comandos especificados em uma linguagem formal. Assim como Levy (1998), Bearman já ressaltava que salvar o conteúdo é necessário, mas não o suficiente.

> O conteúdo, a estrutura e o contexto da informação digital devem ser enlaçados às funcionalidades do *software* que preserva suas conexões

> executáveis ou representações de suas relações, permitindo a sua reconstrução (BEARMAN, 1994, p. 302).

O motivo dessa limitação é que para salvar uma cadeia de *bits* de um registro é necessário salvar também o seu intérprete. Interpretação significa que a recuperação dos arquivos digitais, tais como textos, figuras, imagens e sons decodifica-os e transforma-os em representações inteligíveis. Rothenberg (2003) mostrou que isso desencadeia, em todas as estratégias aplicadas, um ciclo de atividades repetitivo (ver Figura 3).

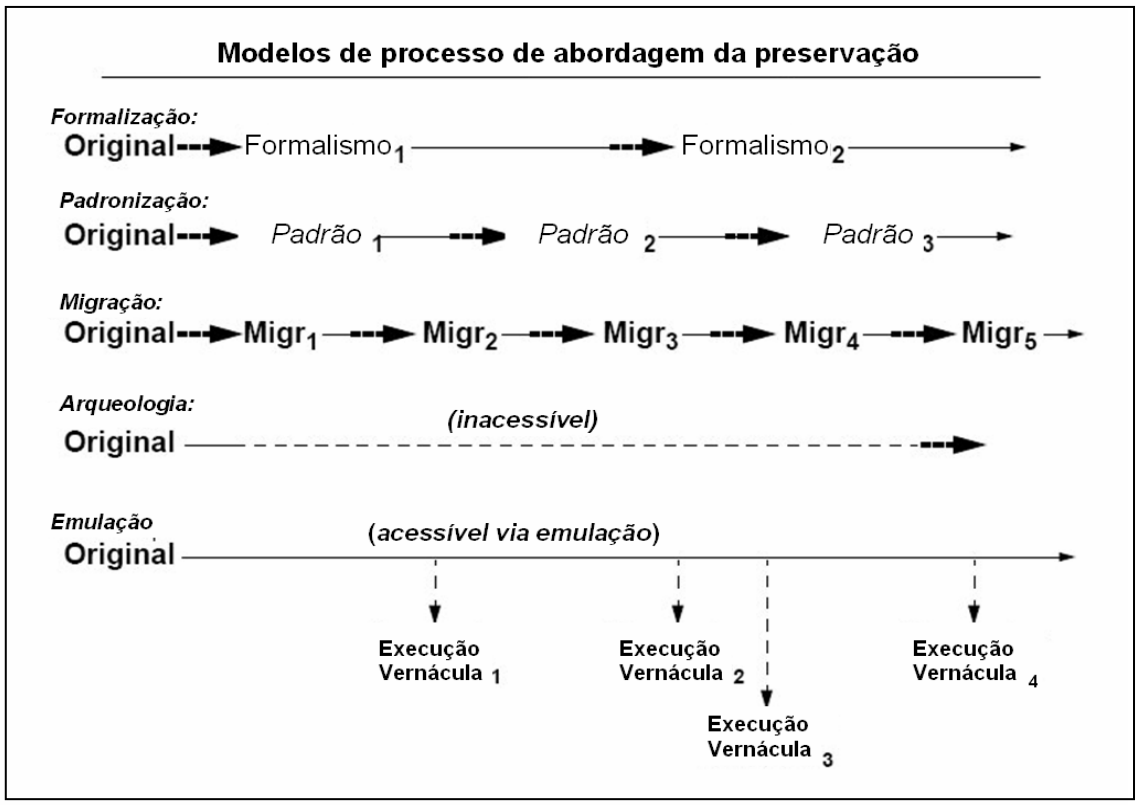


Figura 3 - Modelos do processamento nas perspectivas de preservação digital
Fonte: versão traduzida e adaptada de Rothenberg (2003)

Copiar a mídia garante que o original *bitstream* do documento digital será preservado. Mas um *bitstream* não pode ser auto-explanatório. Um *bitstream* pode representar qualquer coisa como um símbolo. Não é apenas texto, mas também dados, áudio, vídeo, gráficos animados e outras formas ou formatos, atuais ou futuros, únicos ou combinados em formatos de hipermídia. Cada *software* está tendo diferentes tipos de codificação. Por isso cada computador precisa de alguma especificação de *software* para ativar os documentos digitais. Um *bitstream* pode ser

inteligível apenas dentro de um *software* que o criou, ou algum *software* próximo que possa entendê-lo.

2.2.3.2 Preservação da tecnologia e do objeto digital

Atualmente, os principais métodos recomendados para a preservação dos objetos digitais podem ser agrupados em dois tipos: os estruturais e os operacionais. Os estruturais tratam dos investimentos iniciais por parte das instituições que estão se preparando para implementar algum processo de preservação e que adotam ou adaptam um dos modelos de metadados existentes ou seu próprio esquema. As atividades operacionais são as medidas concretas aplicadas aos objetos digitais. O Quadro 3 enumera os métodos de preservação mais freqüentemente usados.

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

Quadro 3 - Métodos de preservação digital

As estratégias operacionais que englobariam os novos requisitos de preservação seriam a migração de suporte e a atualização do meio (preservação física), a conversão dos formatos, a emulação (preservação lógica) e a preservação do conteúdo (intelectual). Para Thibodeu (2002), o que diferencia as estratégias de preservação é o grau de especificidade e o enfoque na tecnologia ou no objeto digital (ver Figura 4).

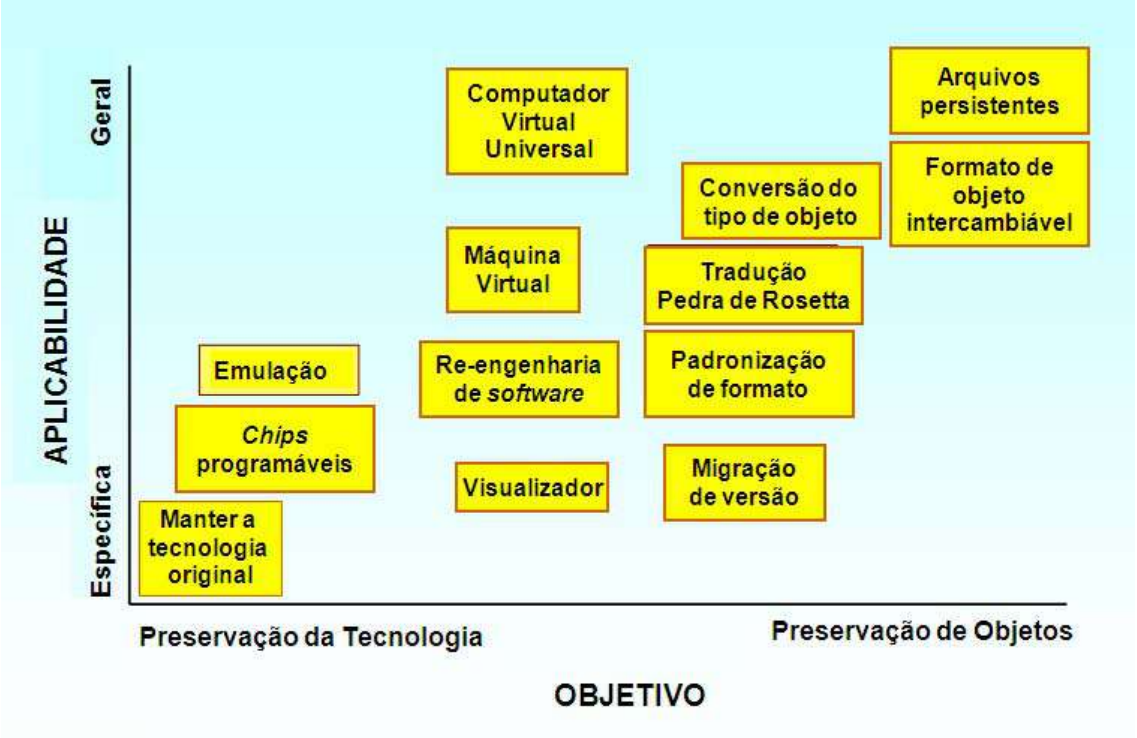


Figura 4 - Diferenças entre estratégias de preservação
Fonte: versão traduzida e adaptada de Thibodeu (2002)

Outra questão relativa à preservação digital é a autenticidade dos dados, pois os usuários precisam ter certeza de que a informação que estão utilizando é o que diz ser e não foi alterada nem por outros usuários, nem em alguma atualização de dados. Na manipulação dos objetos digitais existe o risco de corromper a originalidade, integridade e autenticidade da informação.

Serão mencionadas a seguir, as três metodologias mais usadas para enfrentar o problema da preservação digital: migração, emulação/encapsulamento e conservação de *hardware* e s*oftware*.

2.2.3.3 Migração

Para Martin e Coleman (2002), no ambiente tecnológico atual, todos os dados eletrônicos devem ser migrados a cada ano para que possam sobreviver. A migração periódica da informação digital a partir de um ambiente de *hardware* ou de um *software* para outro é a estratégia operacional para a preservação digital mais freqüentemente usada pelas instituições detentoras de grandes acervos.

Segundo a definição da Task Force on the Archiving of Digital Information da Commission on Preservation and Access e o Research Library Group (1996), migração é a transferência periódica de materiais digitais de uma configuração de *hardware/software* para outra, ou de uma geração de tecnologia computacional para a geração seguinte. O propósito da migração é preservar a integridade dos objetos digitais e assegurar a habilidade dos clientes para recuperá-los, expô-los e usá-los de outra maneira diante da constante mudança da tecnologia. A importância da migração é transferir para novos formatos enquanto for possível, preservando a integridade da informação.

Um arquivo digital pode converter objetos digitais para um reduzido número de formatos-padrão. Por exemplo, dados textuais podem ser estocados em um formato de *software* relativamente independente como ASCII, em formatos proprietários de grande difusão como PDF, ou em formatos baseados em aplicações de *Standard Generalized Markup Language* (SGML) como XML (COLEMAN; WILLIS, 1997).

Com o passar do tempo, os dados podem ser copiados, recriados e atualizados como migrações periódicas necessárias em novos formatos para seu uso em novas gerações de *hardware* e *software.* Na área de documentos digitais, alguns modelos começam a surgir com a finalidade de converter materiais de formatos proprietários para a linguagem XML.

Os metadados têm um papel importante em qualquer estratégia de migração bem-sucedida. Esse tipo de estratégia dependerá dos metadados criados para registrar a história da migração de um objeto digital. Também, existe a necessidade de informação do contexto para ser registrada (e preservada) para que, dessa maneira, futuros usuários possam entender o ambiente tecnológico no qual um objeto digital foi criado. A International Federation of Library Associations (IFLA) recomenda a utilização de formatos padrão nos objetos digitais e nos metadados. Esses formatos devem permitir a migração dos próprios metadados para outros padrões.

Vários estudos têm mostrado que a migração é a estratégia mais adotada pelas instituições que produzem informação científica (OCLC/RLG PREMIS WORKING GROUP, 2004). Mas como a Task Force on the Archiving of Digital Information apontou, “[...] apenas uma única estratégia aplicada para todos os

formatos de informação digital e nenhum dos métodos de preservação atuais são inteiramente satisfatórios [...]" (TASK FORCE, 1996, p. 38).

A metodologia de migração está evoluindo, as técnicas ligadas a essa estratégia estão sendo amplamente aceitas, mas a sua adoção está apenas começando a ser aplicada no caso dos objetos digitais complexos. Para alguns especialistas a migração continua sendo experimental e permanece como uma área de pesquisa interessante.

Para a Task Force a migração não é opcional, mas sim uma operação essencial:

> Existe uma variedade de estratégias de migração para a transformação de informação digital contida em sistemas obsoletos para sistemas de *hardware* *software* atuais, para que essa informação continue acessível e usável. Nenhuma única estratégia pode ser aplicada a todos os formatos de informação digital e nenhum dos métodos de preservação atuais é totalmente satisfatório. As estratégias de migração e seus custos variam dependendo dos contextos de aplicação, dos formatos, dos graus de preservação que as estratégias atendem e das suas possibilidades de recuperação. (TASK FORCE, 1996 p.27)

A migração continua sendo a estratégia mais adotada; entretanto, para muitas instituições ainda é cedo para pensar em uma grande mudança tecnológica. Além das grandes e antigas instituições, os arquivos e bibliotecas não têm iniciado nenhuma mudança tecnológica significativa. A prevalência da migração significa a prevalência de produtos comerciais como os da Microsoft Office e dos produtos da empresa Adobe na área científica (PDF/A). Os assuntos ligados com a atualização de *software* e *hardware* não estão sendo tratados pelas instituições, que manifestam estar mais preocupadas em coletar e inserir os conteúdos nos sistemas atuais.

Assim como aumenta o número de versões de *software* ou documentos que são movidos de uma aplicação para outra (por exemplo, de *WordStar* para *WordPerfect* ou *Microsoft Word* para o *Open Office*), o risco de corrupção e de perdas essenciais de informação também crescem.

O trabalho das bibliotecas na estratégia de migração envolve grande número de formatos, complexos e proprietários. Uma tarefa que diminui a sucessiva migração de formatos é a transferência de recursos digitais em formatos preserváveis no momento da submissão em um repositório. Um autor pode escrever seu trabalho em *Microsoft Word* e submetê-lo dessa forma no repositório. Como o repositório considera o *Word* não ajustável para preservação, os bibliotecários o

convertem para PDF, que possui especificações de formatos públicos e que promete uma representação imutável no futuro.

Como sendo um componente da estratégia de migração, várias iniciativas de preservação estão sistematicamente documentando formatos de dados, os *softwares* nos quais eles estão sendo criados, e coletando suas especificações dos formatos quando elas estão disponíveis. Nos Estados Unidos, a Library of Congress está construindo *sites* com informações sobre formatos de conteúdo digital. Na Inglaterra, a National Archives criou o primeiro registro operacional de formato, Pronom, e a comunidade internacional de preservação está trabalhando na criação de um formato de registro de dados internacional. Mesmo assim, um registro operacional de formato pode oferecer uma variedade de serviços, incluindo identificação de formato, verificação e conversão. Entretanto, isto ainda pode demorar alguns anos, devido principalmente a assuntos organizacionais.

O DSpace do MIT utiliza a migração como sua principal estratégia de preservação. O MIT mantém o original e permitirá o suporte através da migração de formatos classificada como “suportada” (ou não proprietária). Espera-se, mas sem nenhuma garantia, que serão migrados os arquivos de sistemas proprietários como *Word*, *Excel*, etc., para os quais as ferramentas de conversão tornam-se possíveis de serem disponibilizadas porque são formatos originais instalados pelos usuários.

Existe também a “migração a pedido”, na qual a versão original do material é retida, e quando necessário, são aplicadas ferramentas próprias para converter o original para os formatos requeridos pelo usuário (MELLOR, 2003). Com ela é possível poupar tempo, pelo fato de que os usuários não atualizam uma versão de *software* ou *hardware* para a seguinte com a mesma velocidade. Esse conceito foi testado no projeto CAMiLEON (CAMiLEON, 2001).

Nos últimos 10 anos, um grande número de estratégias de preservação surgiu, como projetos desenvolvidos em todo o mundo (LEE *et al.*, 2002). Todos os trabalhos analisados mostraram que a estratégia mais citada e aplicada é a de migração (WHEATLEY, 2001). Essa estratégia tem sido usada principalmente nos contextos em que não existem objetos digitais interativos, apenas imagens, bases de dados e documentos de texto.

O tipo e a quantidade de recursos devem ser levados em consideração na escolha de uma estratégia de preservação. A decisão deverá garantir que a melhor estratégia de preservação foi escolhida entre as opções disponíveis. Como todas as

estratégias usadas até hoje, a migração depende de fatores tais como: experiência técnica, expectativas de usuários, orçamento institucional, equipamento existente e tempo disponível.

Para Ferreira, Baptista e Ramalho (2006), é necessário entender todos os passos envolvidos no processo de migração, considerando uma seqüência de atividades:

1. Seleção da estratégia: as decisões a serem tomadas inicialmente envolvem os formatos e as aplicações. As instituições devem procurar otimizar a combinação do formato a ser tratado e o *software* para a conversão "[...] um que preserve o maior número das propriedades do objeto digital a um custo baixo [...]". Especificamente com relação ao custo, esses autores falam que ele deve ser considerado como uma variável multidimensional.
2. A conversão: a tarefa da conversão consiste na reorganização dos elementos que compreendem o objeto digital em estruturas lógicas definidas por formatos diferentes. Para o responsável pela preservação, a ação da conversão consiste na preparação de uma aplicação de conversão e sua execução em uma coleção de objetos digitais. O processo pode ser automatizado com a ajuda de *scripts* de programação.
3. Avaliação dos resultados: após o processo de conversão, os objetos resultantes deverão ser avaliados para determinar a quantidade de dados perdidos durante a migração. Para conseguir isso, as propriedades que compreendem o objeto original devem ser comparadas (chamadas de propriedades significativas) com as propriedades dos objetos convertidos. Se os resultados da avaliação não correspondem às expectativas (ex.: as propriedades do objeto foram degradadas para níveis inaceitáveis), deverá ser selecionada uma alternativa de migração diferente e o processo reiniciado na integra. O processo de avaliação requer de trabalho manual com relação aos documentos comprimidos de arquivos de imagem e a elementos gráficos nos documentos de texto.

Em Portugal, na Universidade do Minho estão sendo realizados estudos para automatizar as três atividades relacionadas com a migração (seleção das

opções de migração, conversão e avaliação). Os pesquisadores da Universidade do Minho esperam testar uma Arquitetura de Serviço Orientado (SOA) para combinar vários sistemas distribuídos ou repositórios digitais que permitam às instituições preservar suas coleções automaticamente. O sistema trabalharia na fase inicial da entrada de dados *(ingest processes)* oferecendo uma série de serviços. O exemplo que eles apresentam é o do Arquivo Nacional de Portugal: na instituição está sendo desenvolvido um sistema de repositório digital para preservar objetos digitais autênticos produzidos por instituições filiadas (Projeto RODA).

Ao mesmo tempo está sendo criado o processo de *ingest* e as políticas de preservação, para ajudar aos produtores na preparação do material antes de ele ser submetido ao repositório. Dessa forma, será realizado um processo de normalização para lidar com objetos em formatos não identificados. Na presença de um formato não reconhecido o repositório solicitará um serviço de identificação de formato ao SOA, para obter informação e conferir a sua integridade. Após essa tarefa, a instituição solicita ao SOA uma lista de formatos para os quais o documento pode ser convertido, informando suas preferências e requisitos de preservação (provenientes das políticas), por exemplo:

> - as intervenções de preservação deverão ser acessíveis e rápidas;
> - as intervenções deverão preservar o número máximo de propriedades do objeto original;
> - as intervenções não deverão recorrer a formatos que sejam dependentes de pagamentos ou direitos autorais (FERREIRA, BAPTISTA; RAMALHO, 2006).

Na estratégia de migração a habilidade de decodificar o formato atual deve estar sempre presente, garantindo a compatibilidade retroativa, a interoperabilidade dos programas e o uso de formatos padrão. Os recursos digitais podem ser transportados de uma plataforma para outra através de migrações múltiplas, considerando as corrupções acumulativas que podem alterar algumas dessas características (ROTHENBERG, 2000).

Uma das vantagens da migração é que ela permite o acesso rápido ao recurso, já que o documento estará sempre em formato padrão. Dessa forma os usuários encontraram como resultado da conversão, recursos digitais considerados suficientes apesar da perda de algum atributo visual (*look-and-feel*).

2.2.3.4 Emulação e encapsulamento

Com a emulação pretende-se preservar os documentos digitais e suas funcionalidades (ROTHENBERG, 1999). As técnicas de emulação sugerem a preservação do dado no seu formato original, por meio de programas emuladores que poderiam imitar o comportamento de uma plataforma de *hardware* obsoleta e emular o sistema operacional relevante. O processo consiste na preparação de um sistema que funcione da mesma forma que outro de tipo diferente, para conseguir processar programas. Essa estratégia está relacionada à preservação do dado original no seu formato original. No lugar de preservar o *software* e *hardware* hospedeiro, os engenheiros de sistemas poderiam construir programas emuladores (ROTHENBERG, 1995).

Na prática, os dados podem ser encapsulados junto com a aplicação de *software* utilizado na sua criação, assim como uma descrição do ambiente de *software* e *hardware* requerido para seu funcionamento. Para facilitar seu uso no futuro, Rothenberg (1996) sugere que seja anexada uma anotação de metadados na superfície de cada encapsulação que poderia explicar como decodificar os dados obsoletos contidos e prover qualquer informação desejada sobre esses registros. A emulação precisa do desenvolvimento de técnicas de encapsulamento de documentos, seus metadados, *software* e especificações de emulador de forma a assegurar sua coesão e prevenir sua corrupção (ver Figura 5).

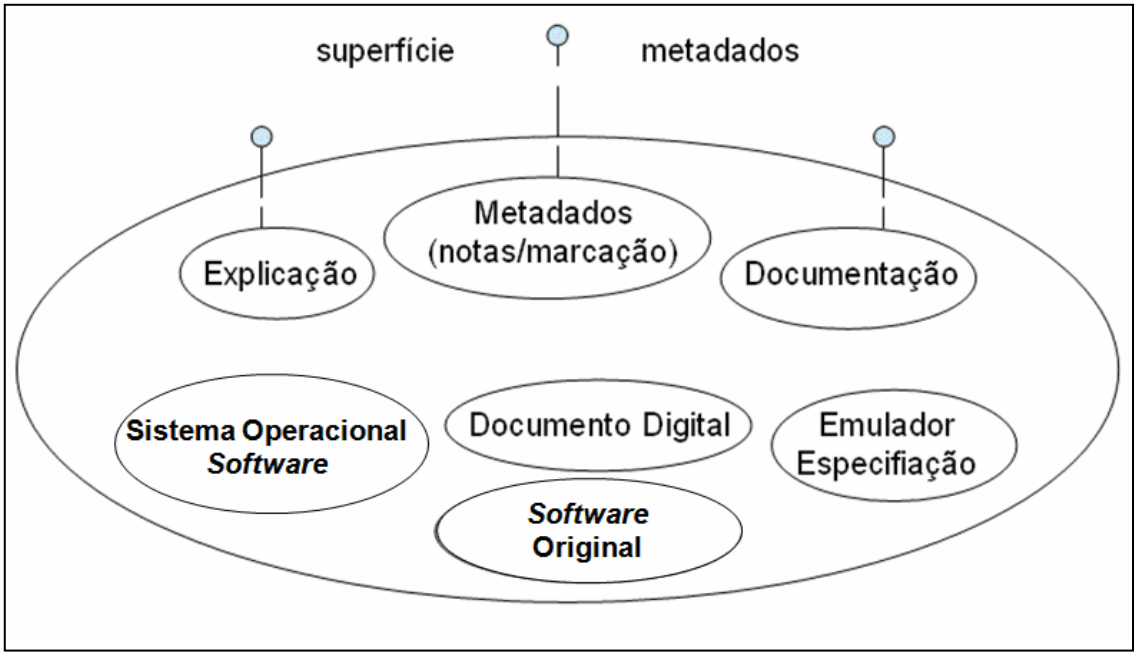


Figura 5 - Um documento digital encapsulado
Fonte: versão traduzida e adaptada de Rothenberg (1995)

A emulação é uma estratégia importante que possui aplicações úteis quando a aparência do recurso digital original é importante, desde que não seja aconselhável investir em uma tecnologia da informação de alto custo. Hendley (1998) pede precaução no caso de se usar apenas esta abordagem e comenta que os gerentes de coleções ficarão dependendo da habilidade técnica dos engenheiros de *software* para emular um ambiente específico e sustentá-lo, para processar em computadores futuros e desconhecidos e recriar o comportamento de documentos digitais.

Segundo Hendley (1998), a metodologia de emulação é a única que pode potencialmente preservar os objetos digitais originais e a capacidade deles serem executados como eles foram inicialmente. Para isto, ele prevê o entendimento do *software* necessário para a visualização do recurso, sem nenhum outro esforço necessário no decorrer do tempo. Esta metodologia retarda a necessidade da conversão dos registros em novos formatos, até o momento em que uma decisão gerencial determine que para acessar os documentos é necessário que o registro original seja acessado em formatos novos.

Rothenberg (1995) afirma que no futuro alguns programas como o MS DOS continuarão sendo ubíquos e que os gerentes das coleções digitais terão que continuar instruindo aos seus usuários a usarem programas como esse. Prevê-se que os programas proprietários ficarão obsoletos e que em conseqüência seu valor

comercial e restrições de direitos autorais expirem, ficando assim disponíveis para os futuros usuários.

Para alguns especialistas (GRANGER, 2000; LAWRENCE *et al.*, 2000), a estratégia de emulação é uma metodologia a curto e a médio prazo, especialmente para aqueles que precisam do recurso original pela sua importância para os usuários da coleção (ROTHENBERG, 1999). A emulação deve ser usada nos casos em que os recursos digitais não podem ser migrados dada a sua complexidade, nem convertidos para formatos de *software* independentes. A implementação desta estratégia é complexa e difícil, devido principalmente, ao tipo de necessidades efetivas a serem preenchidas.

A estratégia de emulação está sendo usada quando o recurso digital não pode ser convertido em formatos de *software* independentes, e migrados no futuro. Ela exige uma descrição da tecnologia usada durante a criação do recurso, e a encapsulação:

a) do recurso digital, acompanhado pelo sistema operacional e seus aplicativos e,
b) das especificações do emulador a ser usado e o histórico do recurso digital.

A emulação ocorre em nível de *software* e de *hardware*. No primeiro caso, a dificuldade está na descrição das aplicações que permita a reprodução posterior. Isto exige uma atualização constante dos conceitos usados devido principalmente às mudanças freqüentes de arquivos multimídia e hipertextuais (GRANGER, 2000). Já a emulação do *hardware* através de um software precisa da informação detalhada sobre suas características.

A estratégia de emulação tem algo em comum com a preservação de tecnologia e envolve os seguintes critérios:

a) uma mídia de informação digital estável deve ser armazenada em um sistema
b) a mídia digital será preservada enquanto o documento está sendo preservado como uma “*machine language*”;
c) dados serão representados como um formato de mídia novo através da conversão e reformatação;
d) a integridade da informação digital será incrementada através de um processo de cópias;

e) a aplicação original deve ser preservada e usada para criar ou acessar o recurso digital.

A linguagem XML está sendo considerada capaz de assegurar o encapsulamento dos metadados e das informações necessárias para a interpretação do objeto digital original. O projeto VERS (*Victorian Electronic Records Strategy*) na Austrália implementou a estratégia e recomenda que a estrutura dos dados encapsulados em um único objeto, em linguagem XML permita a inclusão de metadados e informações de identificação.

2.2.3.5 Conservação de *hardware* e *software*

A perspectiva de preservação digital que propõe a conservação do *software* e do *hardware* (*technology preservation*) sugere que os dados digitais sejam preservados em mídia estável e copiados para nova mídia, caso seja necessário, e associados às cópias preservadas de uma aplicação de *software* original. Esse sistema operacional será normalmente usado para ser lido sob uma plataforma de *software* relevante. Essa estratégia pode ter valor para casos particulares ou históricos de *software* e *hardware,* bem como ser útil para a comunidade dos museus (SWADE, 1992), de longo prazo, entretanto, é cara e pouco prática.

Segundo Hendley (1998), a perspectiva da tecnologia da preservação não pode ser apreciada como viável, mas apenas, para o curto e médio prazo, como uma medida de relativo desespero em casos em que valiosos recursos digitais não possam ser convertidos para formatos independentes de *software* e *hardware* e, assim, migrados para outros formatos.

A preservação do *hardware* e do *software* envolve algumas ações específicas, tais como: o armazenamento das cadeias de *bits* em uma mídia digital estável e a preservação dessa mídia; copiar ou atualizar os dados para uma nova mídia e preservar sua integridade e também, preservar a aplicação original, os programas e as plataformas para processar essas aplicações. Em conclusão, a aplicação desta metodologia tem sentido quando o objeto digital foi criado em formato proprietário e em um *software* obsoleto.

Esta estratégia envolve os seguintes critérios:

a) converter a informação através de uma máquina de linguagem como um méio estável;
b) a mídia digital deve ser preservada pela tecnologia;
c) informação pode ser atualizada e copiada como uma mídia de acordo com os requerimentos;
d) os programas de aplicação precisam criar ou acessar os documentos digitais;
e) preservar a integridade da informação digital durante o processo da cópia;
f) o *hardware* ou o sistema de *software* devem suportar a aplicação de *software*;
g) preservar a plataforma de *hardware* na qual opera o sistema de *software* no qual o documento pode ser processado.

A conservação de *software* e h*ardware* pode ser adotada quando um recurso digital valioso está sendo acessado por uma aplicação de *software* que deve ser operada em um sistema de *software*. Esta estratégia está sendo considerada em declínio.

2.2.4 Metadados de preservação

Na preservação digital, ao criar uma estrutura que guarde o conteúdo e a estrutura da informação, as informações devem ser preservadas de modo a evitar que sejam corrompidas. Para isso, várias técnicas estão sendo usadas, tais como a criação de metadados de preservação que são atributos especiais de dados ou documentos, geralmente descritivos. Para Hodge (2000), a preservação e o processo de arquivamento devem começar no momento da criação do objeto digital, indicando o valor que ele tem e enfatizando os assuntos ligados com a consistência dos metadados no início do seu ciclo de vida.

Harnad (2003) é enfático ao afirmar que os *e-prints* autodepositados, como versões para acesso livre, não precisam de nenhum trabalho técnico em relação aos metadados, e que o mínimo do protocolo de coleta de metadados da

Open Archive Initiative (OAI) é satisfatório. Para ele, os metadados usados para criar a interoperabilidade[11] entre repositórios têm mostrado sua eficácia desde 1999. Os teóricos do acesso livre concordam que para os originais é necessário esse gerenciamento de longo prazo.

Na área da ciência da informação, o papel dos metadados tem sido tratado amplamente, enfocando a sua necessidade e desafios que eles criam no gerenciamento de recursos digitais. Alemneh, Hastings e Hartman (2002) mencionam que o princípio fundamental dos metadados é integrar coleções digitais, heterogêneas, em múltiplos formatos e de várias instituições, permitindo que elas sejam acessíveis a qualquer pessoa, em qualquer lugar ou hora.

Segundo vários autores (Besser, 2000; Moen, 2001; Sutton, 1999; Waibel, 2000; Zeng, 1999), o uso de metadados está centralizado na identificação, autenticação e contextualização dos registros, das pessoas, dos processos e dos sistemas que criam, gerenciam, mantêm e usam esses objetos digitais. A função básica dos metadados é prover informações sobre o documento digital, dando subsídios aos processos de gestão, recuperação e reprodução. Com eles, a interoperabilidade dos estoques informacionais pode estar garantida, configurando-se em mais uma solução para a preservação digital (SANT'ANNA, 2001).

Os metadados têm várias funções e servem para vários propósitos. Eles dão suporte aos processos de registro e de negociação, das seguintes formas:

a) realizando a proteção dos registros como evidências e garantindo sua acessibilidade e usabilidade através do tempo;
b) facilitando a habilidade de entender os registros;
c) apoiando e garantindo o valor de ser evidências dos registros;
d) ajudando na comprovação da autenticidade, confiabilidade e integridade dos registros;
e) apoiando e gerenciando o acesso, a privacidade e os direitos autorais;
f) apoiando uma recuperação efetiva;
g) apoiando as estratégias de interoperabilidade, permitindo a captura dos registros criados em diversos ambientes técnicos e gerenciais e sua sustentabilidade;

[11] Capacidade de um sistema ou dispositivo comunicar de forma transparente com outro sistema ou dispositivo.

h) provendo e mantendo a lógica dos *links* entre registros e o contexto da sua criação.

Os metadados são expressos em grupos de elementos e atributos, sendo que o agrupamento de elementos depende das suas relações. Os elementos e atributos podem ser mandatórios ou opcionais. Estruturados em esquemas de metadados eles especificam e descrevem o conjunto padrão de elementos de metadados e suas inter-relações Os esquemas provêem uma sintaxe formal (ou estrutura) e semântica (ou definições) para os elementos de metadados (ex.: *ISO Metadata for Records*).

Baseados na experiência de vários projetos de implementação de metadados, autores como Rosenthal; Lipkis; Robertson e Morabito (2005) têm reforçado a visão de que os padrões atuais são prematuros e insuficientes em um futuro previsível. Assim também, o seu uso e evolução requereram sempre um tradutor. A tradução ou transformação de um padrão para outro não tem garantido, a inexistência de perda ou a cobertura de todos os dados, sendo, dessa forma, impossível a reconstrução de um objeto digital original.

Na visão desses autores, a evolução dos padrões de metadados sempre ficará atrás das novas versões que estarão sendo usadas considerando que a ciência da informação continua avançando e os novos objetos digitais não param de ser criados, permanecendo “órfãos” de padrões desde seu surgimento.

Na literatura da área podem ser encontrados comentários sobre a necessidade de fazer uma clara definição dos atributos de um documento através de metadados para que, além de minimizar os riscos de perda do conteúdo, possam ser feitas revisões da informação mantida em repositórios e identificar como ela pode ser aperfeiçoada (CAPLAN, 2004; DAY, 2004; McGOVERN, 2007; CANDÁS ROMERO, 2006). Um primeiro passo seria a revisão de todos os metadados existentes em repositórios em uso para, em seguida:

1) localizar os metadados para descoberta;
2) localizar os metadados de preservação mínimos e,
3) identificar os metadados adicionais requeridos para a preservação e métodos de captura, técnicos e de proveniência.

Uma parte importante do suporte que os metadados oferecem é permitir as estratégias de interoperabilidade, conseguindo capturar os registros criados e ambientes técnicos e de negócios diferentes e possibilitar sua sustentabilidade. Com

os metadados, os *links* entre registros e seu contexto de criação são mantidos e podem ser informados. Os atributos dos metadados podem ser identificados como tipos de metadados necessários para ser aplicados em diferentes áreas da organização, dependendo dos riscos e necessidades (Ex.: *ISO RM Standard*).

As comunidades que têm implementado algumas das iniciativas de metadados concordam que um único padrão de metadados não pode atender a todas as necessidades dos diferentes contextos digitais (OCLC/RLG, 2001). Nem mesmo o projeto *Dublin Core* tem conseguido prover uma solução que abranja todo tipo de recurso de informação digital. Por isso, alguns formatos de metadados têm sido desenvolvidos para ser usados em diferentes contextos de informação digitais. Entre eles estão os seguintes:

- *Government Information Locator Service* (GILS);
- *Visual Resources Association* (VRA);
- *Content Standard for Digital Geospatial Metadata* (CSDGM) e,
- *Encoding Archives Description* (EAD).

Também, tem sido foco da atenção o desenvolvimento de ferramentas para a descoberta de materiais digitais, para facilitar a interoperabilidade entre os recursos de diferentes áreas. Alguns desses esforços são o *Resource Description Framework* (RDF), o *Extensible Markup Language* (XML)[12] e o Z39.50. Entretanto, Mullen (2001) mostra que essas iniciativas não garantem o acesso de longo prazo dos recursos digitais, pois a abordagem que elas dão aos metadados enfatiza a descoberta de recursos de informação, e não sua preservação.

Os metadados de preservação são aqueles relacionados com o conteúdo do documento, seu contexto e estrutura no momento da sua criação, assim como das mudanças acontecidas em todo seu ciclo de vida. Essa perspectiva inclui sua estrutura física, técnica e lógica (a relação entre registros). Nos últimos anos têm surgido vários esquemas e padrões de metadados de preservação. A maioria originou-se da comunidade de profissionais da ciência da informação, entre os quais se encontram:

---

[12] *Extensible Markup Language* (XML) é um formato padrão para publicação, armazenamento e transferência de documentos por via eletrônica, independente dos fabricantes de *software,* facilmente interpretável por múltiplas aplicações e, em certa forma, auto-explicativo.

a) *National Library of Australia;*
b) *CEDARS;*
c) *California Digital Library;*
d) *National Library of New Zealand;*
e) *Open Archival Information System (OAIS) Reference Model;*
f) *OCLC Digital Archive Metadata;*
g) *VERS;*
h) *AUDIOMD e VIDEOMD;*
i) *Creative Commons Metadata.*

Alguns formatos de metadados estão sendo usados para a identificação, validação e extração de dados:

a) *GDFR/JHOVE;*
b) *PRONOM/DROID;*
c) *Toolkit da* National Library of New Zealand*;*
d) A base de dados de software *NIST;*
e) *CRiB.*

Também, arquivos nacionais e outras agências governamentais têm desenvolvido padrões para o registro de metadados:

j) Public Record Office [PRO];
k) National *Archives of Australia;*
l) Public Record Office Victoria*, VERS e Information Interchange Office of the e-Envoy;*
m) *METS: Metadata Encoding and Transmission Standard;*
n) *MIX ou NISO Z39.87 Technical Metadata for Digital Still Images;*
o) Library of Congress;
p) *NEDLIB;*
q) Arquivo Nacional do Brasil.

Os metadados de preservação são uma forma especializada de administrar metadados que podem ser usados como um meio de estocar a informação técnica que suporta a preservação dos objetos digitais. No seu documento sobre as diversas classes de informação que devem estar presentes num esquema de metadados de preservação, a iniciativa da OCLC/RLG (2002) afirma que existem três condições básicas para o uso dos metadados de preservação:

a) abrangência: metadados estendidos a todos os aspectos do processo de preservação digital;
b) estruturação: ser complemento aos componentes e processos funcionais de um sistema de arquivamento digital;
c) larga aplicabilidade: aplicáveis a uma extensa faixa de tipos de objetos digitais, atividades de preservação digital e instituições.

Para Besser (2000), os metadados de preservação têm como objetivo prover informação suficiente sobre os recursos digitais para que metodologias de preservação, tais como a migração e a emulação, possam ser aplicadas, e dessa forma contar com as informações necessárias sobre as versões, os esquemas de descompressão e conexões.

A criação e o uso de metadados representam uma parte importante em todas as estratégias operacionais de preservação digital, por não estarem baseados na conservação de *software* e *hardware,* emulação ou migração, como um meio para garantir a autenticidade, registrar o gerenciamento de direitos e coleções de dados. Além de assegurar a interação com recursos de busca (ROTHENBERG, 1996), eles também informam sobre a origem do material e os detalhes técnicos dos registros, como qual foi a versão do *software* usado, como foi construído, por exemplo, o registro.

Para Lavoie (2004), os metadados de preservação são empacotados (física ou logicamente) com os objetos digitais, acompanhando-os durante todo o período do arquivamento. As informações sobre o contexto contidas nesses metadados são essenciais para garantir a acessibilidade e usabilidade dos objetos. O desafio dos esquemas de metadados de preservação é antecipar a informação que será necessária para cumprir determinados objetivos de preservação. São informações sobre a proveniência, autenticidade, ações de preservação, o contexto tecnológico e condições legais (LAVOIE; GARTNER, 2005).

Em 1998, o Research Libraries Group (RLG) recomendou, a utilização de 16 elementos de metadados para garantir a preservação de longo prazo. Mas o consenso sobre o modelo único de metadados de preservação não foi ainda alcançado (McGOVERN, 2007).

Segundo Searle e Thompson (2003), a implementação de um esquema de metadados de preservação requer no mínimo a limitação do escopo dos metadados para preservação, a fim de maximizar o potencial para sua automação e garantir o

controle da mudança nos metadados. Na parte operacional do desenvolvimento do esquema de metadados para preservação, algumas atividades são necessárias:

1. implementação do padrão de metadados;
2. criação de repositórios de metadados de preservação integrados a outros repositórios de metadados usados pela instituição;
3. definição do *script* para extrair os metadados de preservação, que produza um relatório em XML daquilo que os metadados identificaram como importante para a preservação, para depois serem transferidos para o repositório de metadados.

Várias instituições internacionais têm construído marcos conceituais para identificar elementos específicos de metadados de preservação. Estão sendo desenvolvidos dicionários de metadados de preservação utilizados nesses projetos e que refletem o tipo de necessidades que as instituições enfrentam (CEDAR PROJECT TEAM, 2002; LUPOVICI; MASANÈS, 2000; NATIONAL ARCHIVES OF AUSTRALIA, 1999).

A National Library of New Zealand produziu um esquema de metadados para preservação, orientado para ser um ponto de equilíbrio entre os princípios expressos no modelo de referência *Open Archival Information* System (OAIS) para a preservação de metadados e a praticidade de implementação de um conjunto de metadados operacionais de preservação. Ele está dividido em quatro entidades: o objeto, o processo, o arquivo e a modificação de metadados (Figura 6).

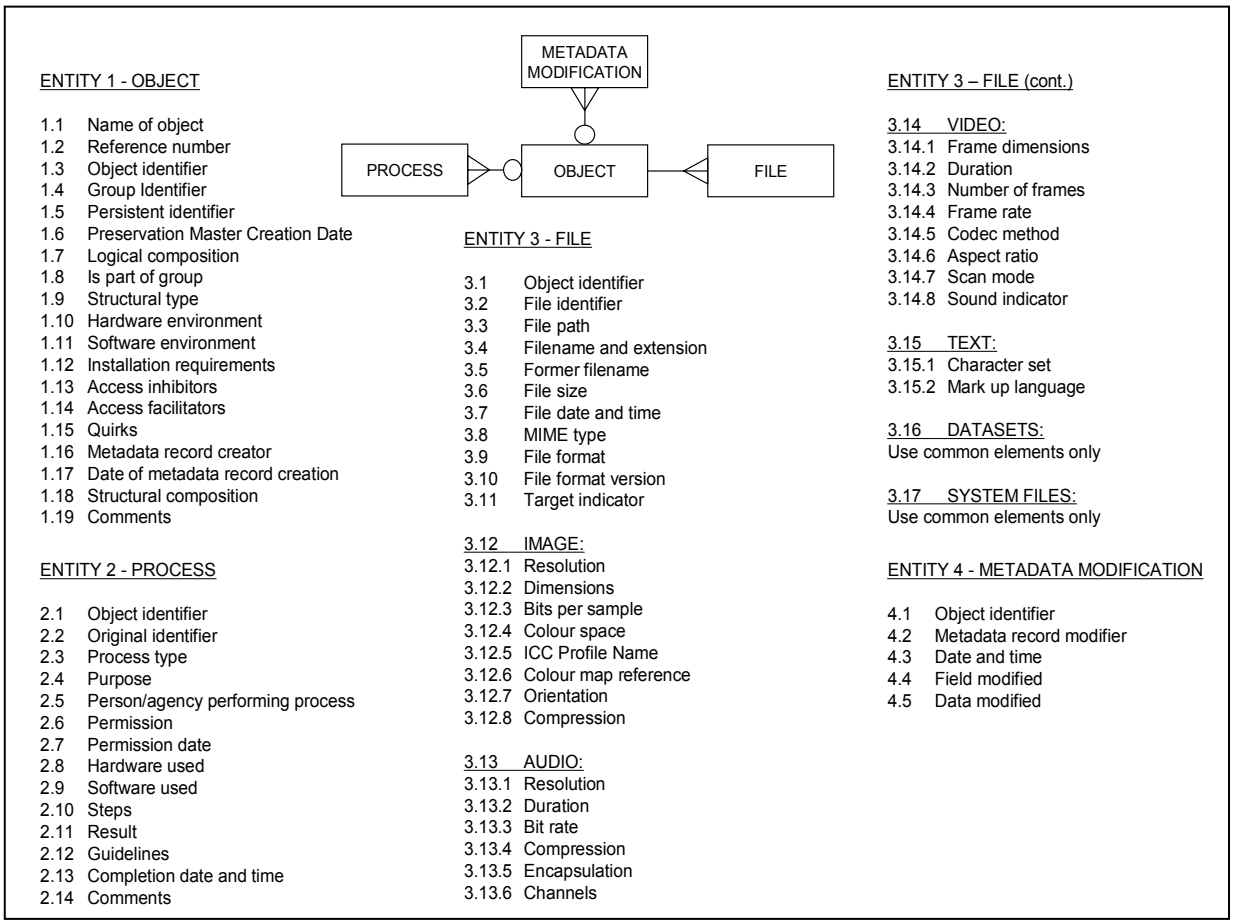


Figura 6 - Esquema de metadados da NZNL
Fonte: *National Library of New Zealand (*2003)

No artigo de Searle e Thompson (2003), o esquema da NLNZ é apresentado identificando-se as áreas nas quais ele pode ser desenvolvido: para a expansão dos dados de definição dos elementos do esquema; para criação de um repositório baseado nesses dados de definição, e para o desenvolvimento de ferramentas de extração automática de metadados para alimentar o repositório. O esquema identifica os dados que a biblioteca coletará e manterá. O arquivo original (*preservation master file)* deve ser preservado, gerenciado e disseminado através do tempo. Segundo a NLNZ (2003), os metadados de preservação contêm informações sobre:

a) políticas e técnicas de preservação adotadas;
b) efeitos da estratégia adotada;
c) gerenciamento de coleções;
d) gestão de direitos autorais;

e) autenticidade do recurso digital.

Também a NLNZ afirma que os metadados de preservação podem ser usados para:

1) armazenar informações que dão suporte às decisões e ações relacionadas com a preservação;
2) documentar processos de preservação, tais como migração, transformação e emulação de registros que afetam os processos de preservação;
3) garantir a autenticidade da versão "original" preservada ao longo do tempo;
4) permitir a identificação dos objetos que são responsabilidade da biblioteca.

A National Library of Australia (2003) tem mostrado como os metadados de preservação (considerados como uma forma especializada de metadados administrativos) podem ser usados como um meio de estocar a informação técnica que sustenta a preservação dos objetos digitas.

Em 2005, o *OCLC-RLG Metadata Implementation Strategies Working Group* (PREMIS) que trata dos "[...] aspectos práticos da implementação de metadados de preservação nos sistemas de preservação digital [...]" (OCLC/RLG, 2005) lançou a primeira versão do *Preservation Metadata Data Dictionary*[13]. O PREMIS é quase um padrão para metadados de preservação, mas muitos envolvidos com projetos de preservação digital não conseguem implementar metadados de preservação porque trabalham com apenas reduzida parte deles no gerenciamento da preservação digital.

O Dicionário de Dados PREMIS descreve o conjunto de elementos básico de metadados necessários para dar suporte à preservação digital e como eles devem ser usados na estruturação de um repositório digital (OCLC/RLG, 2005). O Dicionário classifica esses elementos em cinco entidades:

1) Intelectuais: descrevem a manifestação física do objeto digital.
2) Agentes: descrevem pessoas, organizações e aplicações envolvidas.
3) Eventos: descrevem ações de preservação direcionadas a um objeto digital.

[13] Dicionário de dados sobre metadados de preservação

4) Objetos: descrevem o conjunto de manifestações físicas de uma entidade intelectual.
5) Direitos: descrevem os direitos e as permissões relacionadas com o objeto digital.

As práticas de aplicação de metadados de preservação mencionada na literatura assinalam que, para cumprir com seu objetivo, elas devem ser mantidas consistentemente através do tempo. Para isto, é fundamental a criação da documentação que controle e informe cada passo na criação dos metadados. Alemneh, Hasting e Hartman (2002) recomendam o detalhamento do *workflow* e um guia para que os usuários possam prover toda a informação requerida sobre o processo de criação de metadados, com exemplos para diferentes formatos de arquivos. O guia, segundo esses autores, deverá informar as regras, sintaxe e descrição necessária para identificar a fonte de informação para cada elemento. Eles citam como exemplo disso o esboço do sumário da estrutura dos metadados de preservação da University of North Texas (Figura 7).

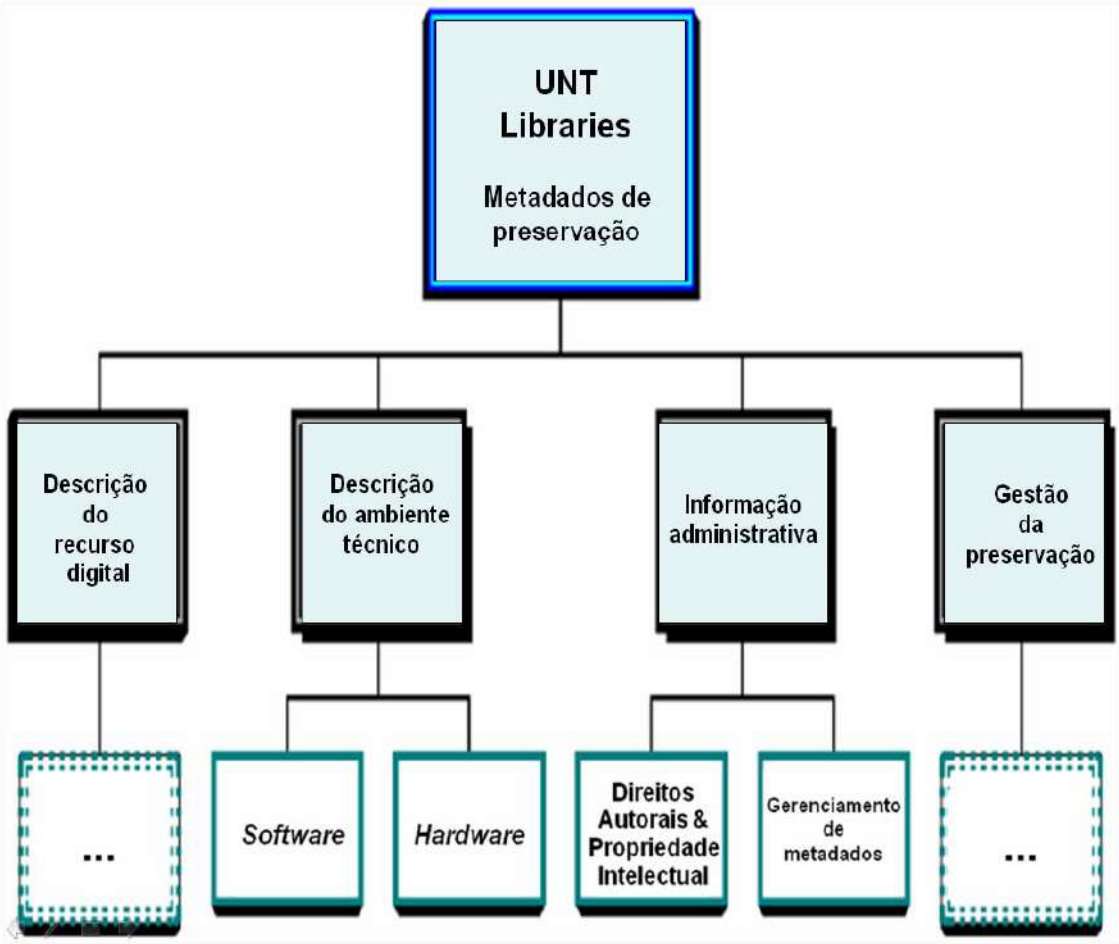


Figura 7 - Estrutura de metadados de preservação das UNT Libraries
Fonte: versão traduzida e adaptada de Alemneh; Hastings e Hartman (2002)

2.2.4.1 Iniciativas de implementação de metadados de preservação

A medida que os objetos digitais vão se convertendo no método dominante de registrar e disseminar informação científica, o gerenciamento dos metadados de preservação torna-se uma área de crescente interesse por parte dos responsáveis por tais recursos.

A área dos metadados de preservação continua a ser o foco das principais iniciativas de preservação digital devido à quantidade maciça de informação que está sendo colocada nos arquivos, bibliotecas e repositórios digitais, assim como também pela ênfase na criação de coleções digitais mais do que em documentos isolados. Internacionalmente, várias iniciativas estão desenvolvendo estruturas para incluir metadados de preservação. Entre elas conta-se com grupos de trabalho da

iniciativa conjunta da OCLC/RLG, o projeto CEDARS (Universidade de Leeds), PADI (National Library of Australia) e a NEDLIB (da Biblioteca Nacional da Holanda).

Não existe consenso sobre qual seria o padrão de metadados de preservação, mas todos esses grupos coincidem em alguns pontos, como é o caso do *CEDARS Guide to Preservation Metadata* e o *OCLC Preservation Metadata Framework*, sendo que esta última é considerada o mais avançado desenvolvimento no campo dos metadados de preservação. Outros modelos existentes são:

- *National Library of Australia's Guidelines for the Preservation of Digital Heritage*;
- Esquema de metadados para preservação da National Library of New Zealand;
- Elementos de metadados propostos pela Cornell University Library;
- *Library of Congress Metadata Encoding and Transmission Standard* (METS);
- *Open Archival Information System Reference Model,* publicado pelo Consultive Committee for Space Data Systems (CCSDS).

O trabalho da OCLC e a RLG vêm discutindo a necessidade de metadados de preservação e as diferenças entre os tipos de metadados e seus objetivos, sejam estes na descoberta ou na disseminação de conteúdos. Os especialistas envolvidos têm analisado os elementos de metadados de preservação identificados em outros grandes projetos, como o NEDLIB, CEDAR e HARVARD, tentando integrar todos eles.

Nessa mesma linha, está o trabalho realizado pelo Defense Technical Information Center (DTIC) dos Estados Unidos, que publicou um guia para metadados de preservação em 2002 (DEFENSE TECHNICAL INFORMATION CENTER, 2002), para apoiar a criação da sua *Defense Virtual Library*. Os metadados de preservação nesse guia abarcam vários tipos de objetos, como imagens, vídeos e relatórios técnicos. O conjunto de elementos chega a mais de 100, propostos pelo DTIC como os elementos necessários para dar início à preservação a longo prazo de objetos digitais e em concordância com o Modelo de Referência OAIS.

Uma experiência relevante foi a da Public Record Office Victoria na Austrália, que criou a *Victorian Electronic Records Strategy* (VERS) com registro de um número extenso de metadados de preservação identificados. Uma característica

especial é que o VERS inclui, também, metadados para descrever a própria estratégia. Dessa forma, no futuro, um usuário que acesse um objeto encapsulado sem documentação poderá extrair um documento em XML com as descrições textuais das técnicas usadas na construção do objeto encapsulado e dessa forma recriar o *software* no qual foi criado e até substituir o repositório (Figura 8).

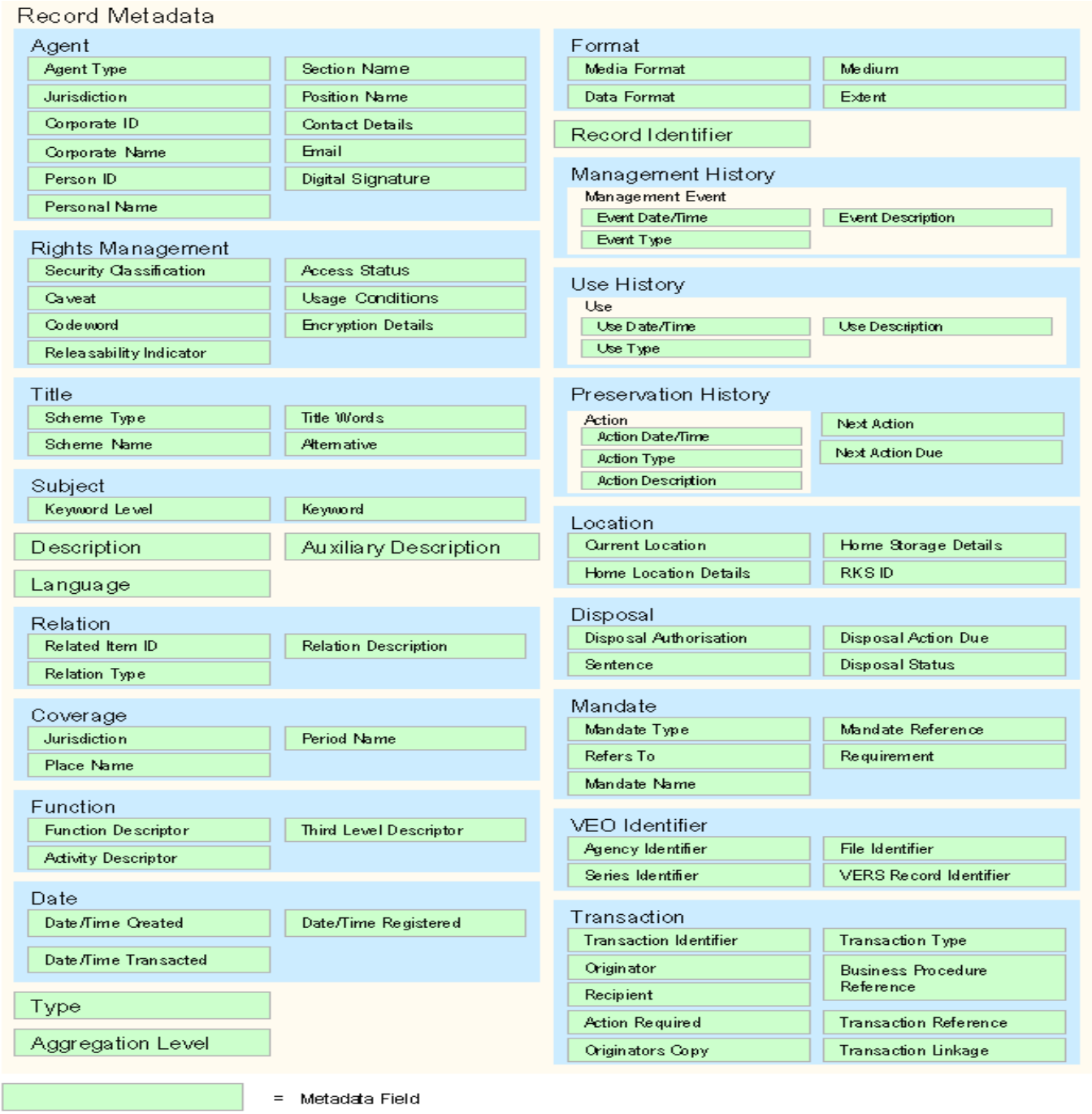


Figura 8 - Registro de metadados da Victorian Electronic Records Strategy
Fonte: Public Record Office Victoria (1999)

Além disso, cada registro leva incluída uma referência textual aos padrões de metadados como uma forma de documentar o formato de preservação usado. Essa estratégia converge com a linha de atuação proposta por Hodge e Carroll (1999), quando sugerem que, para melhor funcionamento dos objetos digitais, os metadados devem estar embutidos no próprio documento e serem extraídos pelos mecanismos de busca como etiquetas no próprio recurso *on-line*.

A estratégia VERS pressupõe que os objetos digitais provêm de sistemas de gerenciamento digital que contêm metadados. Esses metadados podem formar um conjunto simples ou um arranjo de metadados qualificados. Segundo os proponentes dos metadados VERS, estes podem chegar a servir como "língua franca" em que os metadados originais podem ser mapeados (QUENAULT, 2004). Há 141 elementos para a preservação e suporte do acesso de longo prazo. Destes, 34 são obrigatórios; 11 o sistema cria automaticamente; 8 vêm preenchidos (*defaulted*); e 2 são derivados de outros elementos. Dos restantes 13 elementos, 11 são selecionados de uma lista e com apenas dois o responsável deve criar seu conteúdo (SINCLAIR, 2003).

Os padrões continuam a ser revisados e melhorados. Com o desenvolvimento da tecnologia, o conhecimento sobre a estratégia de metadados de preservação precisa evoluir junto. Ferramentas relacionadas com a geração, a marcação e a editoração de metadados estão surgindo. Os *softwares* para construção de repositórios começam a integrar ferramentas direcionadas para a preservação digital. Entre eles está o *NLNZ metadata extractor* (RESEARCH LIBRARY GROUP, 2006), um dos projetos da iniciativa da NLNZ na criação de uma estrutura de padrões de metadados que começou no ano de 2000 (Figura 9)[14]

[14] *NLNZ's Metadata Standards Framework* 2000. Disponível em: http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta Acesso em: 22/out/2008

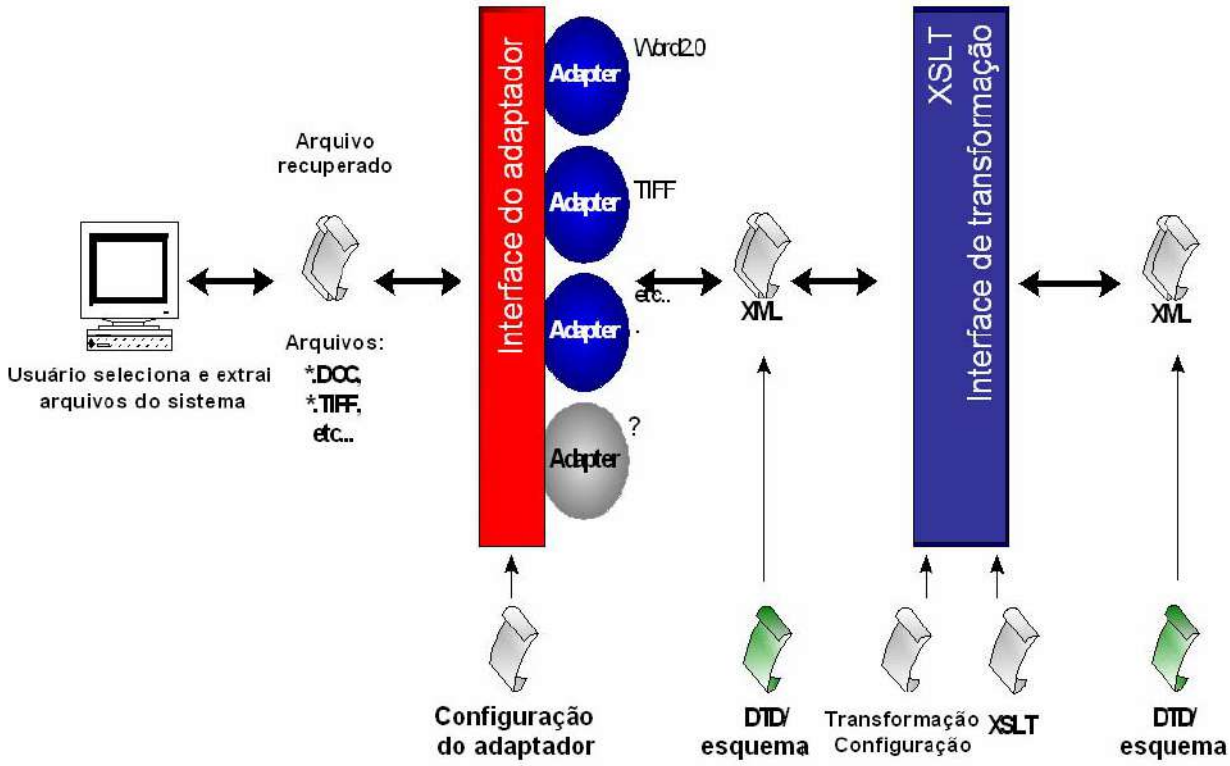


Figura 9 - Ferramenta para extração de metadados de preservação da NLNZ
Fonte: versão traduzida e adaptada de *National Library of New Zealand* (2007)

A ferramenta permite lidar com objetos digitais dentro de um processo de *workflow* identificado pela NLNZ, contendo especificações sobre o modelo de negócios, a infra-estrutura do material digital e as estratégias segundo o tipo de formato dos arquivos. Nesse desenvolvimento de padrões, modelos e representações, a ferramenta de extração de metadados (*Extract Tool*) está sendo considerada a concretização da experiência da NLNZ na manutenção de objetos digitais (ex.: *The New Zealand Locator Service* (NZGLS) e o *Metadata Standard and Reference Manual*).

As soluções propostas para enfrentar os desafios trazidos pelas mudanças constantes no ambiente digital estão sendo desenvolvidas e atualizadas em grandes projetos de preservação digital. Cada vez mais, informações sobre criação, entrega, operação e administração dos recursos digitais vêm sendo consideradas cruciais para determinar seu acesso e preservação (OCLC/RLG, 2001). Os metadados que informam aos usuários sobre o contexto tecnológico e sua

segurança estão ajudando a resolver questões que acompanham a vida desses recursos, tais como:

- que tipo de recurso é?
- como ele pode ser usado?
- como pode ser modificado? e,
- quem esteve envolvido na sua criação/modificação?

Considerando esses objetivos dos metadados de preservação, a National Library of Australia (1999) definiu as cinco funções principais desse tipo de metadados:

1) armazenar informação técnica que apóie as decisões e ações de preservação;
2) documentar as ações de preservação tomadas (ex.: migração ou emulação);
3) registrar os efeitos das estratégias de preservação;
4) garantir a autenticidade dos recursos digitais ao longo do tempo;
5) anotar informação sobre gerenciamento de coleções e direitos de autor.

É importante o consenso nos metadados de preservação e nos padrões para o gerenciamento dos documentos digitais ao longo do tempo, a fim de garantir sua longevidade, interoperabilidade, veracidade, consistência e, desta maneira, motivar que empresas criem soluções que possam dar suporte a eles. Além disso, deve-se apontar que existem elementos lógicos nos arquivos que precisam ser conservados, assim como o registro da evolução da estrutura lógica dos objetos digitais.

2.2.4.2 A norma OAIS

Vários autores mencionam a importância dos padrões na área da preservação digital. Nos últimos anos, têm sido formulados diversos modelos de referência, e algumas iniciativas de preservação afirmam já estar seguindo esses requerimentos e fatores condicionantes.

Em 1998, Fresko ressaltava que as normas aceitas para a preservação digital eram poucas e nenhuma cobria todos os aspectos relacionados com a informação digital. Em 2003, a Digital Preservation Coalition (DPC)[15] realizou uma pesquisa entre seus membros sobre necessidades de preservação digital. O resultado mostrou que, tanto as grandes organizações nacionais estavam começando a lidar com o problema, quanto as de pequeno porte não tinham elaborado nenhum projeto relacionado ao assunto. Nesse mesmo ano, a OCLC e a RLG criaram um grupo de trabalho para estudar os aspectos práticos da implementação de metadados de preservação e projetos de preservação digital (OCLC, 2003).

Como mencionado anteriormente, é fundamental o uso de padrões internacionais de arquivamento de metadados para proporcionar uma gestão mais efetiva de estratégias de preservação de documentos eletrônicos. A mais importante proposta de modelo de metadados de preservação foi produto de 10 anos de trabalho do Consultative Committee for Space Data System (CCSDS), ligado à NASA, quem propôs o modelo de referência *Open Archival Information System* (OAIS), cercado de políticas e procedimentos arquivísticos em sua arquitetura (CCSDS, 2002).

Considerado o modelo para repositórios de metadados de preservação mais usado atualmente, o modelo de referência *Open Archival Information System* (OAIS), traduzido para o português como Sistema Aberto de Arquivamento de Informação (SAAI) (THOMAZ; SOARES, 2004) é, desde fevereiro de 2003, uma iniciativa da International Organization for Standardization (ISO 1471). Ela define um alto nível de modelo de referência para organizações e pessoas que criam informações que podem necessitar de preservação de longo prazo, bem como das organizações que adquirem tais informações.

Originalmente, a norma era destinada a regular o armazenamento de longo prazo de informações digitais obtidas de observações de ambientes espaciais e da Terra, mas ela é aplicável a qualquer tipo de arquivo digital e compatível com os sistemas baseados no protocolo de coleta de metadados da *Open Archives Iniciative* (DAY, 2001).

[15] Disponível em: Digital Curation Centre http://www.dcc.ac.uk/ Acesso em: 23/set/2008

Um arquivo nos termos do OAIS significa considerar os arquivos como organizações de pessoas e sistemas, que aceitaram a responsabilidade de preservar a informação e torná-la disponível, melhorando a comunicação e produtividade entre diferentes comunidades (CCSDS, 2002).

O modelo de referência não especifica um projeto ou qualquer tipo de implementação, mas ele está dirigido para organizações que têm a responsabilidade de tornar a informação disponível de longo prazo. Cada vez mais, comunidades estão adotando o modelo OAIS como uma estrutura para reunir atividades cooperativas entre diferentes áreas de interesse. Várias instituições estão usando a estrutura conceitual para metadados de preservação do OAIS para identificar seus elementos de metadados específicos, os componentes funcionais de sistemas de informação dedicados à preservação digital: a Cornell University, as Bibliotecas Nacionais da Austrália e da Nova Zelândia, On-line Computer Library Center (OCLC), Massachussetts Institute of Technology (MIT), entre outras.

Mesmo proporcionando uma terminologia própria que facilita a comunicação entre os envolvidos na preservação, o modelo de referência não é uma implementação específica, mas proporciona uma lista de condições daquilo que deve ser considerado no estabelecimento de um projeto de preservação de qualquer tipo de documento, seja digital ou analógico.

No modelo de referência OAIS, o primeiro objeto é o pacote de informação (*packaging information)*; ele seria um recipiente que encapsula informação de conteúdo e de descrição para metadados de preservação e outros metadados. No esquema do OAIS, está incluído um modelo de informação para inserção dos metadados de preservação (ver Figura 10).

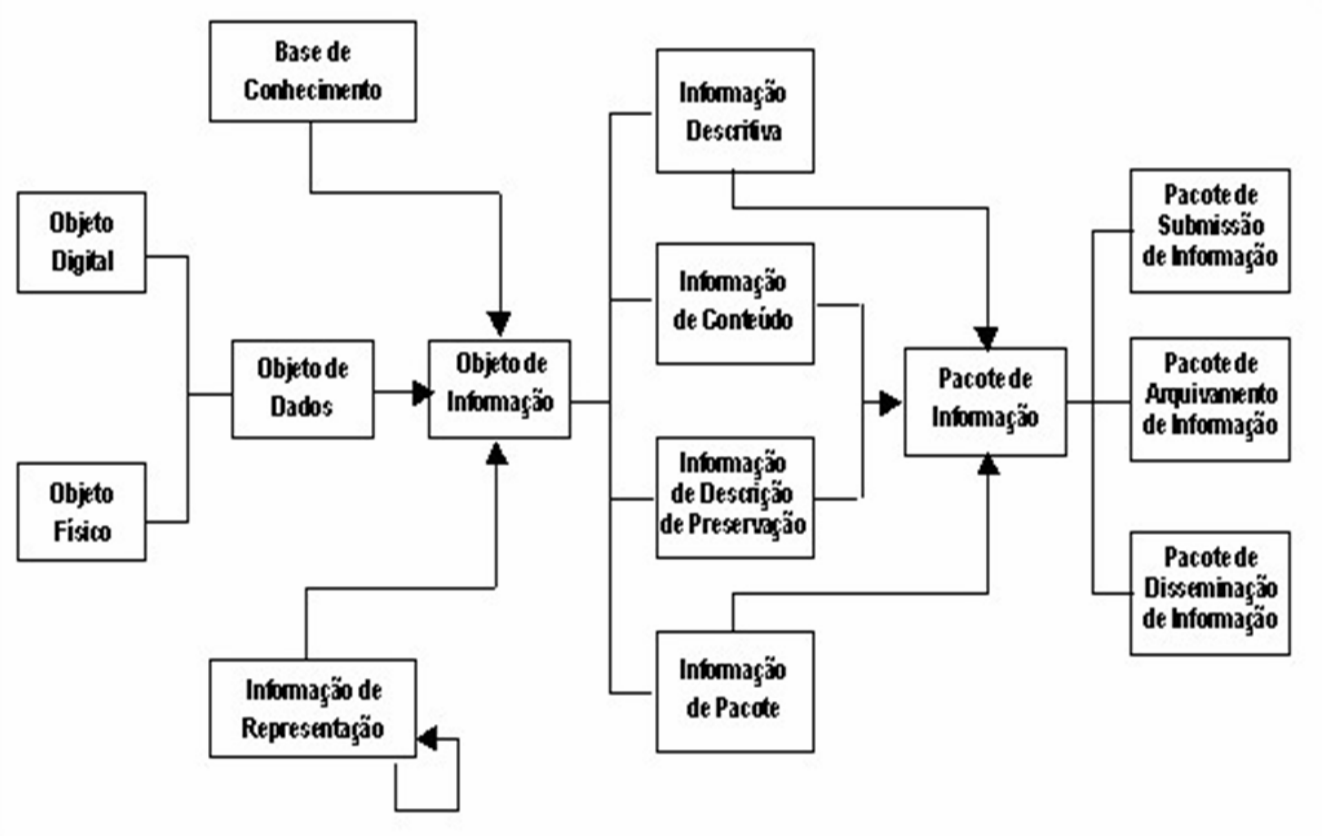


Figura 10 - Modelo de referência *Open Archival Information System* (OAIS) ou Sistema Aberto de Arquivamento de Informação (SAAI)
Fonte: Thomaz e Soares (2004)

No modelo, os objetos físicos e digitais podem ser coletivamente referenciados como objetos de dados. Eles são interpretados pela combinação da base de conhecimento da comunidade-alvo e a informação de representação associada ao objeto de dados. O objeto de dados pode ser complementado com informação de representação, para ser entendido pela comunidade-alvo. No modelo OAIS existem dois grupos de objetos de informação (metadados) unidos e identificados pelo pacote de informação, categorias já definidas no Relatório da CPA/RLG de 1996, são elas:

1. informação de conteúdo (*content information)*: uma informação que é o objeto primário da preservação. Ela contém o objeto digital primário e informação representada necessária para transformar esse objeto em informação com significado;
2. informação de descrição de preservação (*preservation description information)*: qualquer informação necessária para preservar

adequadamente a informação de conteúdo com a qual está associada, acrescentando significado em quatro aspectos:

1) informação de referência (*reference information*) (ex.: identificadores);
2) informação de contexto (*context information*) ex.: classificações por assunto);
3) informação de proveniência (*provenance information*) (ex.: *copyright*, histórico);
4) informação de integridade (*fixity information*): documenta os mecanismos de autenticação;

No esquema conceitual do OAIS atuam quatro entidades: produtores, consumidores, administração e o arquivo propriamente dito, e existem seis entidades funcionais: recepção, armazenamento, gerenciamento de dados, administração do sistema, planejamento de preservação e acesso (Figura 11)[16].

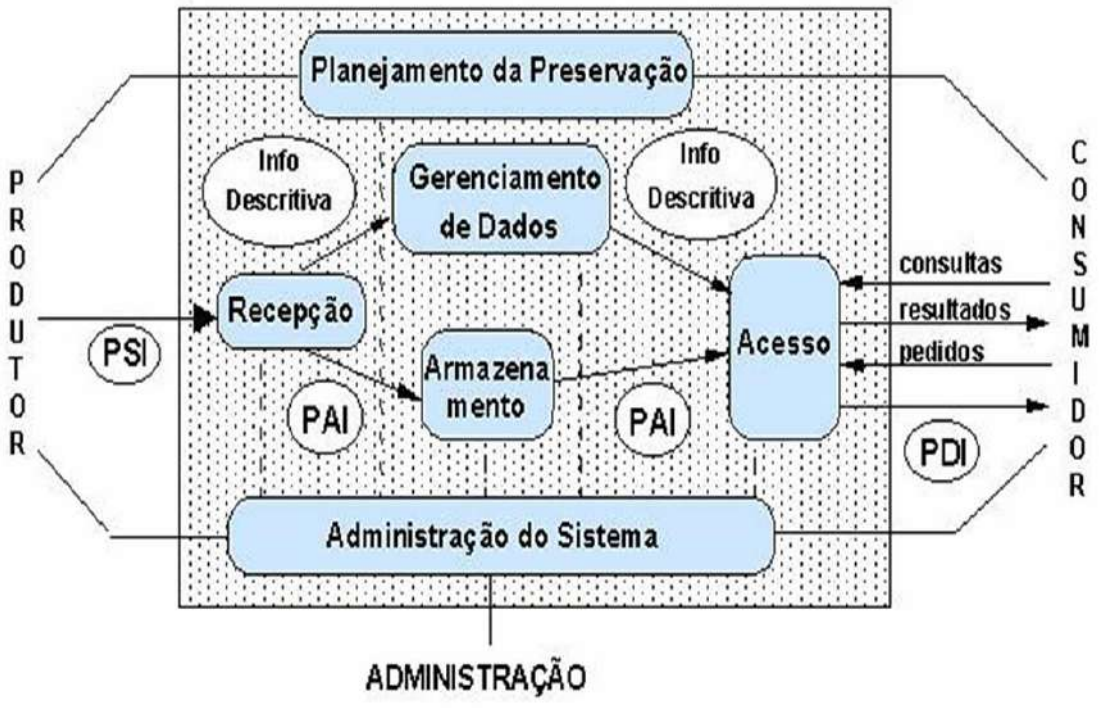


Figura 11 – Modelo funcional OAIS
Fonte: Thomaz; Soares (2004)

[16] PSI: Pacote de Submissão de Informação
PAI: Pacote de Arquivamento de Informação
PDI: Pacote de Disseminação de Informação

As entidades funcionais gerenciam o fluxo de informação entre as entidades que formam o ambiente OAIS e identificam os componentes funcionais dos arquivos relacionados com a preservação dos objetos digitais. No ambiente OAIS cada entidade é responsável por algum dos processos do sistema:

1) produtor: é a pessoa ou sistema cliente que fornece a informação;
2) consumidor: é a pessoa ou sistema cliente interessado no material arquivado ou por uma versão transformada da mesma;
3) comunidade-alvo: são os usuários potenciais, a serem reconhecidos para a realização da meta maior da preservação de longo prazo.

Com a norma OAIS, as instituições podem configurar sistemas responsáveis pela preservação da informação de longo prazo e torná-la disponível para uma comunidade de usuários. A definição dos componentes necessários para a criação desses sistemas que o modelo proporciona serve para suportar uma gama de serviços de preservação. Essa descrição está relacionada com as atividades de submissão, organização e gerenciamento e acesso contínuo de um repositório em funcionamento. Especificamente, o PAI (Pacote de Arquivamento de Informação)[17] contém o material a ser preservado e que será conservado no repositório, assim como a informação para sua representação.

A conceitualização do modelo OAIS serve para compreender os componentes necessários para preservar e acessar informação de longo prazo. Essa descrição do arquivamento de objetos digitais serve para que instituições não arquivísticas participem efetivamente no processo de preservação, permitindo também:

1) efetuar comparações entre arquiteturas, operações, estratégias e metodologias de preservação;
2) estabelecer o mesmo modelo para informações em formatos analógicos;
3) ampliar o consenso entre as partes comprometidas com a preservação digital, e,
4) guiar na identificação e produção de normas relacionadas com o modelo.

[17] O PAI é "[...] uma forma concisa de fazer referência a um conjunto de informação que tem em princípio, todas as qualidades necessárias para a permanente, ou indefinida preservação de longo prazo de um objeto designado de informação [..]" (CCSDS, 2002 p. 33). Tradução nossa.

Um repositório que segue as definições do modelo OAIS está obrigado a cumprir as seguintes responsabilidades:

1) negociar com os produtores de informação a aceitação dos critérios relacionados às características dos documentos (tipo, assunto, fonte, originalidade, singularidade, mídia, formatos, etc.) e a conformação com padrões e políticas;
2) obter controle das informações que garantem a preservação de longo prazo (propriedades e direitos autorais);
3) determinar as comunidades de usuários potenciais;
4) garantir a compreensão da informação fornecida;
5) seguir políticas de documentação de procedimentos para auxiliar na localização dos originais;
6) definir as condições de acesso, distribuição e preservação do acesso.

A administração de um sistema de preservação que siga o modelo do OAIS deve acompanhar o desenvolvimento de padrões para apoio, com isso ele terá a garantia de ser uma ferramenta na interoperabilidade entre bibliotecas e repositórios digitais. Suas funções principais de Recepção, Geração de Pacotes de Arquivamento e Acesso são fundamentais na determinação da existência, descrição, localização e disponibilidade da informação armazenada no OAIS.

No caso da Recepção dos PSI (Pacote de Submissão de Informação), a administração deve definir os formatos, a forma de transferência (*harvesting* ou convencional) e se ela será para custódia. Nessa primeira etapa é quando os PSI são transformados em PAI (Pacote de Arquivamento de Informação), reformatando os arquivos, convertendo os dados de representação e reorganizando as informações de conteúdo.

A administração do PSI é fundamental para a geração de informações descritivas (metadados) que facilitarão a recuperação dos PAI. Toda atualização deverá ser coordenada para uma correta transferência de informações novas dos PAI para a etapa de armazenamento de gerenciamento de dados. O armazenamento envolve a manutenção e recuperação dos PAI e sua incorporação na área de arquivamento permanente.

Na vasta lista de componentes, serviços e funções de um OAIS, o acesso auxilia os consumidores a determinar a existência, descrição, localização e disponibilidade da informação armazenada. A geração de um Pacote de

Disseminação de Informações (PDI) facilita a solicitação e recebimento dos produtos de informação contidos nas bibliotecas e repositórios digitais.

Coordenar um sistema OAIS para *e-prints,* de acesso livre, significa não apenas lidar com interfaces, pedidos de consulta, relatórios e pedidos de ajuda, mas representa lidar especialmente com as autorizações para usuários de documentos digitais. Para Harnad (2003), alcançar o modelo OAIS não é uma prioridade para os repositórios digitais de *e-prints*, mas sim, o fato de estar em concordância com o protocolo dos Arquivos Abertos (OAI-PMH).

Para a preservação digital de *e-prints* o modelo OAIS ilustra o tipo de atividades e complexidade de relações por detrás de um repositório digital. São as organizações que devem determinar o tipo de sistema, parcerias e políticas de acesso a serem implementados para o estabelecimento e crescimento dos repositórios. Um repositório que possui o protocolo de coleta de metadados da Iniciativa dos Arquivos Abertos e está em concordância com o modelo de referência OAIS pode oferecer garantias de acessibilidade de longo prazo, segurança e integridade, beneficiando o avanço da pesquisa científica (HIRTLE, 2001; HITCHCOCK *et al.*, 2007).

Em 2002, o relatório do OCLC/RLG apontava que um repositório digital que estiver fundamentado no modelo OAIS deve estar baseado no Pacote de Submissão de Informação (PSI) que inclui tanto o objeto digital como os metadados necessários. No primeiro momento, os objetos devem ser submetidos no repositório como um PSI*,* o qual contém os objetos digitais e sua respectiva informação. Em seguida, o PSI é reconhecido e encapsulado como um Pacote de Arquivamento de (PAI). O objeto digital original é armazenado como um *bitstream* no PAI, junto com os metadados necessários para que o *bitstream* seja entendível e útil como recurso digital.

Para os pesquisadores da iniciativa OCLC/RLG e outros autores (STEENBAKKERS, 2001), toda publicação eletrônica contém três elementos: o *bitstream*, o objeto lógico no *bitstream* e a funcionalidade necessária para decodificar esse formato lógico, sendo que cada um deles possui seus próprios critérios e requisitos de preservação. É o caso do *bitstream* que só pode ser preservado se os *bits* forem copiados e o ambiente de armazenamento atualizado. Steenbakkers (2001) identifica três normas principais para a preservação digital:

1. separar o documento do seu suporte original e armazenar o documento em um ambiente de arquivamento controlado;
2. o ambiente de arquivamento controlado deve ser construído em concordância com o modelo de referência OAIS;
3. o ambiente de arquivamento ou sistema de depósito deve estar separado do ambiente computacional para publicação da instituição, enfocado em arquivamento e não em recuperação ou autenticação. Isto faz com que o sistema de depósito dure mais e torne-se capaz de ser atualizado assim que novas tecnologias forem surgindo.

A estrutura conceitual do OAIS está construída para atender as necessidades desse ambiente de arquivamento digital, mas isso não evita que ela possa ser aplicada em qualquer tipo de objeto físico (impresso ou digital), assim como também, ela não é para um único tipo de implementação ou sistema. Qualquer instituição pode ter mais de um arquivo OAIS com diferentes missões, objetivos e comunidades, apenas se faz necessário que cada arquivo providencie todas as funções especificadas na norma, sendo possível que algumas dessas funções sejam compartilhadas entre os arquivos e até entre diferentes organizações, seguindo o modelo do OAIS.

O desenvolvimento de um sistema de preservação baseado no modelo OAIS apresenta a dificuldade de não haver na literatura muitos exemplos; a informação que o próprio modelo oferece não é suficiente, forçando as organizações a planejar como os componentes do sistema precisam ser incluídos.

Uma das características do OAIS é que ele não provê um conjunto de metadados, mas um modelo em que os padrões de metadados devem ser propostos. Nesse sentido ele é tanto um modelo funcional como um modelo de dados. Um exemplo disso é a National Library of New Zealand, que criou seu próprio esquema de metadados de preservação “[...] para encontrar um equilíbrio entre os princípios expressos no modelo de informação do OAIS e os aspectos práticos da implementação de um conjunto de metadados de preservação que podem ser trabalhados [...]” (NLNZ, 2003, p. 4).

Os desenvolvedores do esquema de metadados de preservação da NLNZ identificaram algumas áreas que devem ser estudadas:

a) desenvolvimento de definições de dados para os elementos no esquema;

b) *design* do repositório baseado nessas definições de dados, e

c) desenvolvimento de ferramentas para extração automática de metadados para incrementar o repositório.

Alguns conjuntos de metadados estão sendo propostos para comportar o modelo OAIS, tais como:

- padrão de metadados da National Library of Australia;
- especificação de Metadados CEDAR;
- metadados da NEDLIB (*Networked European Depository Library*);
- estrutura de metadados da OCLC/RLG (Online Computer Library Center/Research Libraries Group).

Entre as iniciativas e sistemas que abordam o problema da preservação digital e que de alguma maneira seguem o modelo OAIS, encontram-se as seguintes (ver Apêndice C):

1) CDPP (Centre de Données de la Physique des Plasmas) ;
2) DIAS (*Digital Information Archival System)*;
3) DIOnAS (*Data Ingest and Online Access Sub-System*);
4) DiVA Project (*Digitala Vetenskapliga Arkivet*);
5) IMAGE project (*Imager for Magnetopause-to-Aurora Global Exploration*);
6) Iniciativa DSpace do MIT *(*Massachusetts Institute of Technology)
7) JSTOR (*Journal Storage*);
8) LOCKSS (*Lots of Copies Keep Stuff Safe*);
9) LOTAR (*Long Term Archiving and Retrieval and Product Data within the Aerospace Industry*).
10) MoReq (*Model Requirements for the Management of Electronic Records*)
11) Digital Archive (OCLC)
12) PANDORA (National Library of Australia)

Com o número de iniciativas crescendo[18], a decisão de qual sistema adotar deve levar em consideração as necessidades da comunidade de usuários do sistema. Uma experiência mencionada na literatura é a da British Library. Em 2000,

[18] O CCSDS continua informando sobre as atividades que o centro desenvolve relacionadas com o modelo, e uma lista de requisitos para repositórios nele baseados (http://nssdc.gsfc.nasa.gov/nost/isoas/).

a biblioteca, que recebe materiais digitais como depósito legal, por aquisição e digitalização, precisou de um sistema que pudesse auxiliar no armazenamento, preservação e acesso de longo prazo a uma grande variedade de objetos digitais. Foi realizada uma definição inicial das especificações funcionais para que o sistema estivesse em concordância com os conceitos do modelo OAIS. Dessa forma, a British Library conseguiu apresentar às empresas um sistema adequado às suas necessidades. Houve dificuldades no início devido à ausência de guias de implementação.[19]

O modelo foi considerado perfeito pelo seu nível de estruturação e justificativa política para o desenvolvimento de sistemas desse tipo, assim como pela adaptabilidade da terminologia de comunicação semelhante ao sistema que a instituição queria construir. Mesmo que o sistema aparente ser uma unidade, ele precisou da definição de seus limites uma vez que alguns dos seus componentes já existiam na instituição. Para isso, os processos foram analisados, encontrando-se seus equivalentes.

Elaborou-se um glossário com a finalidade de entender o modelo e sua terminologia e publicou-se o conjunto de elementos de metadados na Internet. A British Library divulgou que alguns grupos de metadados foram inspirados nos da Biblioteca Nacional da Austrália, considerando os metadados do padrão Dublin Core limitantes para seus propósitos. O conjunto de metadados definitivo foi de 30 elementos.

Experiências, como a da British Library, mostram que algumas decisões independem do Modelo OAIS. Como é o caso do tipo de material a ser armazenado, sobre o local do armazenamento das informações descritivas, de relacionamento entre objetos e se outros componentes do gerenciamento devem ou não ser incluídos. Em conclusão, o modelo apresenta desafios para as instituições. Assim como ele pode ser considerado uma estratégia de preservação de longo prazo de informação digital, ele provê conceitos importantes e a adequação necessária dos componentes que podem chegar a ser desenvolvidos no futuro.

Outra aplicação do modelo OAIS é o projeto InterPARES[20] que surgiu dentro da comunidade arquivística. A característica principal desse modelo é que ele

[19] Disponível em: Library of Congress. National Digital Information Infrastructure and Information Preservation Program (NDIIPP). http:/www.digitalpreservation.gov/ Acesso em: 10/09/2008
[20] Disponível em: http://www.interpares.org Acesso em: 13/out/2008

leva em consideração aspectos relacionados com os elementos intelectuais e físicos, sob a perspectiva de quem os preserva (GREENAN, 2003). Tanto esses como outros projetos que usam o OAIS têm sua motivação nos interesses nas potenciais comunidades utilizadoras.

O CCSDS produziu o documento *Producer Archive Interface Abstract Methodology Standard* (CCSDS, 2003) para prover uma estrutura geral para a relação entre o produtor e o arquivo. A metodologia não especifica apenas uma implementação, mas uma lista de requisitos a serem considerados no momento de negociar um acordo entre as partes. A primeira fase envolve o contato inicial, a definição preliminar do projeto e se ele é praticável; ele também identifica os objetos digitais e os padrões que devem ser aplicados para que esses objetos possam ser identificados, assim como informações sobre o número de itens, segurança, aspectos legais e contratuais, operações de transferência e validação. A segunda fase ou fase de formalização inclui mais informações sobre os objetos a serem transferidos, metadados específicos e a criação de um dicionário de dados e um modelo formal. Após a formalização dos aspectos legais e as possíveis mudanças gerenciais especificadas inicia-se a fase da transferência, procurando-se testar e validar o sistema. Na seção final da Metodologia é descrito como o sistema será adaptado para criar um padrão para uma comunidade específica (SAWYER, 2002).

Todas as iniciativas mencionadas indicam a necessidade de mais orientações sobre a implementação do modelo OAIS. Não existem na literatura muitos exemplos de projetos ou relatos de problemas encontrados, na sua aplicação em diferentes processamentos de informação em bibliotecas e arquivos. Existe sim a recomendação de que profissionais das áreas da ciência da informação e arquivologia trabalhem junto com profissionais da informática na implementação do modelo (GREENAN, 2003).

Pelo fato de que o modelo OAIS pode atender a necessidades de comunidades individuais, não existem dois sistemas que sigam o modelo da mesma forma. Isso coloca o OAIS não como uma função de preservação, mas como parte da administração e do planejamento da preservação de documentos digitais.

2.3 PRESERVAÇÃO DIGITAL DA PRODUÇÃO CIENTÍFICA

2.3.1 Preservação digital e acesso livre

Acesso Livre (O*pen Access*), segundo entendimento amplamente difundido na comunicação científica, significa a disponibilização livre na Internet de literatura de caráter acadêmico ou científico, permitindo a qualquer pessoa ler, descarregar (*download*), copiar, distribuir, imprimir, pesquisar ou referenciar (*links*) o texto integral dos documentos. A expressão começou a ser utilizada a partir do momento em que a quantidade de periódicos eletrônicos aumentou significativamente, surgindo, assim, os primeiros títulos que permitiam o acesso gratuito ao seu conteúdo (ex.: PACS *Review*, *Postmodern Culture*, *Bryn Mawr Classical Review*). Entretando, Sally Morris, membro da Association of Learned and Professional Society Publishers, sugere uma definição mais simples, em que o acesso livre significa o acesso aberto a artigos científicos sem restrições para todos (MORRIS, 2004).

A comunidade de especialistas dedicada ao estudo dos fenômenos da comunicação na ciência particularmente relacionados com o acesso livre parece ser unânime em concordar que se trata efetivamente de um novo modelo de comunicação científica. Nesse modelo é permitido ao autor retomar o controle sobre o conhecimento que ele próprio gerou, sobre a integridade do seu trabalho e, paralelamente, a possibilidade de ser citado e reconhecido apropriadamente. As barreiras do acesso à informação científica ficam limitadas apenas ao preço e às permissões associadas com o uso e direitos autorais (BUDAPEST OPEN ACCESS INITIATIVE, 2002).

No Movimento do Acesso Livre à informação científica a questão da preservação deve ser uma ação específica orientada a dois tipos de objetos digitais diferentes, cujos requerimentos não podem ser considerados ao mesmo tempo que os da provisão do acesso. Com isso se quer dizer que cópias da versão final de artigos científicos depositadas em repositórios digitais respondem à necessidade primordial do acesso imediato aos resultados das pesquisas, diminuindo consideravelmente o tempo entre a produção do conhecimento e o seu uso por parte de outros pesquisadores.

Em contrapartida, a versão oficial publicada em periódicos científicos deve ser objeto das ações de preservação de longo prazo. Desse modo, o problema da preservação continua sendo como garantir o acesso para esses originais por parte dos que podem pagar hoje e no futuro. Esse problema, em princípio, não possui relação com a problemática das restrições de acesso impostas aos pesquisadores e instituições que não podem pagar as assinaturas; para estes estão sendo criadas versões duplicadas autodepositadas em repositórios digitais, uma vez que, para a maioria dos 2,5 milhões de artigos publicados anualmente, não existem versões autodepositadas (HARNAD, 2007).

Na perspectiva da OAI, o problema do acesso livre não é a preservação de longo prazo, mas a provisão do acesso hoje. Transpor o problema da provisão para o da preservação futura é outra forma de prolongar o fato de que 85% dos periódicos não estão provendo o acesso livre aos seus conteúdos.

2.3.1.1 OAI e a preservação do acesso de longo prazo

A *Open Archives Initiative* (OAI) possui relação com dois aspectos diferenciados. O primeiro diz respeito à iniciativa que deu origem a uma nova perspectiva dos conteúdos digitais conectados na Rede de forma simples e federada, e o outro, relacionado com um protocolo de comunicação, *OAI Protocol for Metadata Harvesting* (OAI-PMH). Os dois respondem à mesma missão, proposta em 1999 na propalada Convenção de Santa Fé (VAN DE SOMPEL; LAGOZE, 2000), cujos esforços eram desenvolver e promover padrões de interoperabilidade para facilitar a disseminação eficiente de conteúdos.

O OAI representa uma mudança de perspectiva dos padrões de estrutura de dados para aquela dos protocolos de intercâmbio de dados. A palavra "*Open*" refere-se à visão da arquitetura dos dados e ao conceito de uma interface que possibilita o acesso a uma variedade de materiais provenientes de diferentes provedores de dados. Portanto, nesse sentido, ser "livre" não significa que ele seja gratuito. O termo *Initiative*, por seu turno, refere-se ao aspecto da governança da instituição que está por trás dos provedores de dados, uma idéia que remete ao

governo de todas as partes constitutivas do corpo de instituições que suportam e promovem o intercâmbio de dados.

Com o surgimento de mais repositórios digitais, a OAI tem ganhado aceitação e contado com o apoio financeiro de várias agências internacionais (nos Estados Unidos sobressai-se a *IMLS* e na Inglaterra a *JISC*). Já a expressão "*Archives*" advém da origem histórica dos repositórios de *e-prints*, utilizada como sinônimo de repositórios e sem conotação arquivística.

O OAI-PMH, lançado em 2001, é o mecanismo que permite alcançar os objetivos dessa organização (atualmente na versão 2.0), usado por instituições de todo o mundo. Mesmo tendo surgido dentro da comunidade científica para promover e manter o acesso *on-line* aos resultados das pesquisas acadêmicas, o protocolo passou a ser de interesse de outras comunidades que lidam com arquivos digitais e redes de bibliotecas.

O problema que a OAI enfrenta é o compartilhamento de metadados de coleções digitais e objetos digitais. O protocolo OAI tem sido usado, ainda, para coletar dados sobre objetos análogos de bibliotecas e museus. Até o momento, o OAI centraliza seus esforços apenas nos dados que podem ser coletados, e não com a coleta ou interoperabilidade entre conteúdos. Para tanto, os dados a ser coletados devem estar descritos em esquemas de linguagem XML; o protocolo trabalha com os metadados descritivos expressados no padrão de metadados *Dublin Core*, não qualificado e simples como requisito. Outros padrões de metadados ainda estão sendo estudados para serem adaptados ao protocolo no futuro.

As comunidades que integram o OAI estão divididas em duas: as provedoras de dados interessadas em compartilhar dados e as provedoras de serviços, interessadas em usar esses dados compartilhados. Uma instituição pode cumprir os papéis das duas comunidades ao mesmo tempo. As solicitações que o protocolo OAI lança são feitas como um pedido em HTTP, trazendo como resposta um XML que pode ser formatado em um registro de metadados descritivo ou apenas dar as informações sobre determinado repositório.

Em contrapartida, o OAI ainda não permite o gerenciamento dos direitos autorais dos dados intercambiados, assim como também não provê o acesso aos conteúdos, apenas aponta para o local onde estão armazenados, e não suporta outros formatos de metadados que não sejam os do *Dublin Core*. A razão disso é a

intenção dos desenvolvedores em manter o OAI simples para poder alcançar as suas principais metas.

Para muitos especialistas, o esforço da OAI para que seja aberto o acesso aos conteúdos deveria ser o mesmo para que eles continuassem accessíveis de longo prazo. Porém, uma das lições aprendidas com o protocolo é que não é necessário pensar mais em grandes repositórios centrais (análogos a bibliotecas). O protocolo permite que se pense mais em termos de pequenos arquivos distribuídos unificados pelos mecanismos da interoperabilidade. Dentro da OAI alguns arquivos precisam de um tipo de metadados mais detalhados. No caso dos repositórios das versões, esse grau de detalhamento nos metadados apenas é necessário para depurar o acesso.

Segundo Harnad (2007), a necessidade atual é fazer com que os arquivos sejam interoperáveis, via protocolo OAI, incluindo outros arquivos da mesma instituição, sem tentar forçar todos a usar o mesmo esquema de metadados. Isso significa deixar que os esquemas detalhados de metadados trabalhem para a preservação digital, pois são eles os que precisam de mais tempo para implementação.

Os repositórios institucionais deparam-se com o problema de gerenciar suas próprias coleções *on-line* ou qualquer outro material que possuam, acarretando a sua preservação. Essa carga, no caso da literatura de periódicos por assinatura, é aquela que deve ser compartilhada com bibliotecas universitárias e as editoras de literatura seriada. Essa afirmação pode ser relativa, uma vez que muitas editoras acreditam que o trabalho da preservação deve ser de responsabilidade das bibliotecas.

2.3.1.2 Acesso livre e a comunidade científica

O movimento a favor do acesso livre à informação científica vem crescendo, os esforços empreendidos para tornar a informação científica acessível a todos deixou claro que a preocupação com a preservação dos documentos digitais pertence ao grupo de motivos que poderiam fazer com que os pesquisadores e as editoras autodepositassem seus manuscritos originais. Para alguns autores, o

acesso livre aparenta favorecer a preservação digital da informação científica, mesmo que as implementações da Iniciativa enfatizem e estejam focadas no “acesso imediato”. Nesse novo modelo de comunicação científica, a preservação é considerada mais um requisito para o arquivamento em repositórios digitais do que uma necessidade urgente na construção dos repositórios institucionais (SUBER, 2003).

A evolução do movimento a favor do acesso livre tem revelado algumas interações entre os papéis de várias comunidades envolvidas, entre elas:

1. a comunidade de pesquisadores (autores de artigos avaliados pelos pares, assim como seus leitores e usuários);
2. a comunidade de bibliotecários de bibliotecas especializadas (os compradores e curadores do acesso de artigos de periódicos da comunidade de pesquisadores);
3. a comunidade de editores de periódicos científicos (os administradores dos pareceristas, os certificadores desses produtos e os provedores de acesso a esses artigos, anteriormente em papel, mas agora também *on-line*);
4. os empregadores dos pesquisadores (universidades, instituições de pesquisa que se co-beneficiam da produtividade e impacto da pesquisa que eles financiam);
5. as agências de fomento (que respondem para os contribuintes sobre o uso dos fundos e pela produtividade e o impacto da pesquisa e pesquisadores que eles beneficiam).

Embora a comunidade científica tenha sido a primeira a descobrir o poder da mídia *on-line* para aumentar o uso e o impacto dos artigos avaliados, foi a comunidade das bibliotecas a primeira a direcionar a atenção para o escopo e a urgência do problema do acesso à pesquisa, resultado do problema dos custos/possibilidade financeira.

Constata-se que são os pesquisadores, as instituições (de ensino e pesquisa) e os financiadores os únicos que se encontram na posição de resolver o problema do acesso/impacto. As instituições e as agências de fomento são os que podem estabelecer mandatos de depósito compulsório para seus pesquisadores (pelo seu próprio interesse e o interesse da pesquisa em si) do mesmo modo que

têm obrigado aos pesquisadores publicar os resultados de suas atividades científicas.

A comunidade da biblioteconomia e ciência da informação percebeu que existia o problema do impacto dos resultados de pesquisa e acesso, e que ele deveria ser atendido. Mas ela está sendo lenta em perceber qual é o próprio papel na solução do problema, considerando que a questão da falta de recursos não diz respeito ao problema do acesso/impacto. Como resultado, essa comunidade inadvertidamente torna-se parte do problema do acesso e do impacto, no lugar da solução, questionando, no lugar de afirmando, a necessidade de prover o acesso imediato aos resultados das pesquisas. Nesse sentido, Harnad (2007) ressalta que há resistência por parte dos bibliotecários de que todos sejam apenas periódicos de acesso livre, mesmo que alcançar os 100% disso seja extremadamente lento, enquanto a “via verde” do acesso livre é mais rápida, direta e segura.

Harnad observa que a comunidade de bibliotecários não deve perceber o acesso livre como um assunto relacionado com a aquisição de conteúdo, como são as coleções das bibliotecas, mas deve encara-lo como assunto relacionado à “provisão”. Os próprios autores das instituições são os provedores de conteúdo, mesmo que eles publiquem em periódicos de acesso livre ou em periódicos com acesso restrito, mas que autodepositem seus artigos nos repositórios digitais de suas instituições para, dessa maneira, torná-los conteúdos de acesso livre.

Ao compreender que não se trata de um assunto relacionado com a aquisição e sim um tema de interesse da comunidade de pesquisadores, e, cientes desta necessidade, os bibliotecários não apenas podem contribuir na redução dos preços dos periódicos, mas, sobretudo, garantir que aquilo que já está em acesso livre continue permanentemente acessível.

Os bibliotecários não devem trabalhar junto aos autores persuadindo-os a prover o acesso livre à sua produção científica fazendo uso dos seguintes argumentos: a) isso diminuiria os custos com a aquisição dos periódicos; b) resolveria a crise dos periódicos, ou c) garantiria a permanência das versões autodepositadas (suas versões publicadas já estão permanentes como nunca tinham sido), nem para reformar a lei de direitos autorais, entre outros. Para Harnad (2003), os pesquisadores e bibliotecários não apresentam as mesmas motivações para o progresso do acesso livre, são os pesquisadores os únicos que podem prover esse acesso e, portanto, isto deve ser enfatizado.

> Mas a pior coisa que uma biblioteca pode fazer é no lugar de promover o autodepósito para o acesso livre, levantar a bandeira das "preocupações com a preservação" e falar das irrelevantes dúvidas dos seus possíveis depositantes. A preservação é uma preocupação fora de lugar quando se fala de aumentar o acesso e o impacto da pesquisa. Se a biblioteca não tem recursos para contribuir com o acesso livre melhor não fazer nada agora, pois, agora é melhor do que não ter nada agora, mesmo que o agora não seja para sempre (HARNAD, 2003).[21]

Para os defensores do acesso livre, as bibliotecas poderiam usar seus fundos para fazer melhores negócios com as bases de periódicos pagos, depois investir o restante dos recursos comprando assinaturas institucionais em periódicos de acesso livre, promovendo a publicação nesses periódicos por parte dos seus autores institucionais e desenhando um acesso fácil para os periódicos de acesso livre pelos seus usuários. Isso não precisa de financiamento, apenas tempo por parte da equipe da biblioteca. [22]

2.3.1.3 Autodepósito e preservação digital

Como foi mencionado, o objetivo do autodepósito não é a preservação digital, mas sim o acesso e uso. O autodepósito é compatível com o trabalho em paralelo de qualquer meta de longo prazo, incluindo preservação e transição para a publicação de acesso livre.

As versões de artigos científicos depositadas pelos autores em repositórios digitais são apenas suplementos do artigo científico; elas não são versões que apresentam o problema da preservação, do qual os especialistas em preservação mostram estar preocupados. A versão oficial dos editores em XML/PDF é a que apresenta esse problema.

Desde a perspectiva do Movimento do Acesso Livre, a atenção com a preservação de longo prazo é uma preocupação dos editores, das bibliotecas que assinam as bases de dados e das bibliotecas depositárias, e não dos autores. Apenas o acesso aos resultados das pesquisas, uso e impacto constituem as

---

[21] Documento eletrônico. Tradução nossa.
[22] Ex.: St Andrews University Library Disponível em: http://eprints.st-andrews.ac.uk/proxy_archive.html. Acesso em: 20/set/2008

preocupações dos autores, das instituições e agências que financiam a pesquisa. Para Harnad (1999), essa é a principal razão pela qual o autodepósito deve ser mandatório. Portanto, temas que distraem e confundem os autores, tais como direitos autorais, avaliação pelos pares, reforma editorial e preservação devem ser deixados de lado em benefício da promoção do acesso livre à informação científica de fato. Segundo ele, declarações e manifestos continuam declamando esses obstáculos como se eles fossem parte da problemática, quando, na realidade, o acesso livre significa apenas acesso *on-line* a artigos científicos.

Não existe uma preocupação de preservação por parte dos pesquisadores na hora de autodepositar seus documentos. As estatísticas revelam que apenas 15% dos pesquisadores autodepositam sua produção científica. A Iniciativa dos Arquivos Abertos tem como meta chegar a 100% dos pesquisadores autodepositando seus trabalhos.

Harnad (2007)[23] assegura que nenhuma sugestão que se faça para que os autores autodepositem seus documentos em repositórios institucionais conseguirá ajudar no aumento das taxas de autodepósito. Levantar o interesse na preservação pode chegar a ser uma razão negativa e exigir um trabalho extra para a concretização dessa ação. A prioridade para a Iniciativa dos Arquivos Abertos é o autodepósito, que não deve ser negligenciada por outras prioridades arquivísticas, como a preservação digital. Para o acesso livre o problema não é a preservação de longo prazo, mas a provisão para hoje.

A razão do autodepósito para o acesso livre não é preservar o artigo do editor nem substituí-lo. O propósito é prover um suplemento dele, para aqueles que poderiam ser usuários e que não podem pagar pelo aceso à versão proprietária do editor, com o fim de maximizar o uso da pesquisa e seu impacto.

Harnad sugere que não se deve induzir as instituições de pesquisa e as agências financiadoras a criar mandatos de autodepósito para fins de preservação da versão dos autores. Isto é compatível com a continuidade de um trabalho paralelo com metas de longo prazo, incluindo a preservação e a transição para a publicação em acesso livre. Trata-se de assuntos que ocorrem simultaneamente, mas que são distintos porque as soluções de alguns podem não ser adequadas para outros.

[23] Comunicação pessoal em 9 de julho de 2007, Belém, Pará.

A mudança de paradigma, segundo Harnad, será quando 100% dos artigos estiverem autodepositados, o que causará mudanças radicais no sistema de publicação atual. Tais transformações forçarão os editores a se tornarem apenas provedores de serviços de avaliação pelos pares e certificadores de conhecimento científico, no lugar de serem provedores de acesso a produtos analógicos e digitais, como se verifica atualmente. Desse modo, no momento em que os editores cederem, por força das circunstâncias, a provisão do acesso e o arquivamento para as instituições dos autores, estas herdarão a missão de preservar as versões originais e não apenas aquelas dos conteúdos suplementares.

2.3.1.4 “Preservação” e “preservação”

Harnad (2003) distingue dois tipos de preservação: *small-p preservation* e *large-P Preservation*. Para ele, os repositórios institucionais realizam a chamada *small-p preservation* e continuarão a fazê-lo. O motivo principal é porque o acesso livre é para conseguir o acesso imediato, portanto, a *small-p preservation* é um meio necessário para esse fim.

Harnad chama de *small-p preservation* aquela na qual os repositórios de acesso livre garantem que o acesso aos conteúdos continuará para um futuro próximo. Dessa maneira, o problema não é a preservação em si, mas sim a ausência de 85% dos periódicos científicos que podem não ser depositados porque os responsáveis pelos repositórios continuam tornando o processo complicado. Se não existe conteúdo a ser preservado, a preservação simplesmente não acontece.

Harnad cita, como exemplo, o caso do quadro La Gioconda. O original da obra de arte deve ser o objeto de preocupação pela preservação, e não as réplicas que são disponibilizadas para aqueles que não podem pagar para ir ao Museu do Louvre. Assim, enquanto os especialistas trabalham para resolver o problema de preservar o original, os visitantes de qualquer parte do mundo podem apreciar a réplica. O autor argumenta que

> “[...] o pressuposto é prover hoje acesso imediato aos possíveis usuários do que esperar até encontrar uma forma de garantir que o acesso será para sempre. Alguns autores acreditam que os pesquisadores precisam de um acesso permanente, mas não livre. O Movimento do Acesso Livre não é sobre esforços para a digitalização, é sobre aumentar o impacto da pesquisa,

> além do que os autores imaginam. As únicas universidades que não têm interesse em aumentar o impacto das pesquisas são aquelas que não têm produção científica, nesses casos não é um assunto que tenha a ver com dinheiro, mas com política. (HARNAD, 2003)[24]

Alguns dos principais repositórios digitais baseados em arquivos abertos, como o Arxiv (fundado em 1991) e o CogPrints (fundado em 1997), atendem às necessidades de preservação digital, classificada por Harnad de *small-p preservation*. O conteúdo desses repositórios permanece acessível sem nenhuma pretensão de que estejam atendendo a preservação denominada *large-P Preservation*, aplicada nos originais guardados pelas editoras. Isto ocorre, principalmente, porque a preservação não era o principal motivo para o autodepósito nesses repositórios.

Em contrapartida, a *large-P Preservation* é a Preservação como um fim em si mesma: como a motivação para o arquivamento em primeiro lugar; ou como uma necessidade urgente dos conteúdos efêmeros e de conteúdos frágeis nascidos digitais; ou como uma responsabilidade para os provedores de conteúdo (provedores de periódicos) ou compradores de conteúdo (bibliotecas assinantes) ou preservadores de conteúdo (depósitos/bibliotecas que registram conteúdo) que precisam garantir a perenidade dos produtos que vendem ou que adquirem. Mesmo assim, segundo Harnad,

> é absurdo imaginar que *software* como EPrints ignorem a *small-p preservation* ou que seu conteúdo é um pouco plausível de desaparecer amanhã assim como qualquer outro conteúdo digital que tem sido conscientemente preservado e migrado e atualizado hoje, ainda que atualizado segundo os desenvolvimentos das estratégias de preservação (HARNAD, 2003)[25].

A diferença entre preservação e Preservação está em que preservação não é um fim por si só, ela é apenas um meio para um fim (que é o acesso imediato, assim como a provisão do acesso e uso), enquanto Preservação é um fim por si mesmo. As versões e documentos suplementares autodepositados não requerem um tipo de preservação denominado de *large-P Preservation.*

A preservação com “p” minúsculo é assegurada pelos arquivos de acesso livre e para um previsível futuro. Harnad ressalta (2003) que quando todos os artigos avaliados pelos pares forem autodepositados, então o “p” minúsculo se converterá

[24] Documento eletrônico. Tradução nossa.
[25] Documento eletrônico. Tradução do autor.

em “P” maiúsculo, pois o problema da preservação com “p” minúsculo será idêntico ao problema da preservação com “P” maiúsculo. Ainda por algum tempo, a preservação digital, na perspectiva da OAI, deverá continuar a ser aplicada aos originais guardados pelas editoras.

2.3.1.5 Tecnologias para o Acesso Livre

Argumenta-se, portanto, que a fusão da missão de preservação e arquivamento com a missão do acesso e provisão está adiando e esvaziando a motivação para a provisão do acesso livre, que é a principal razão dos repositórios digitais, tais como aqueles construídos com plataformas, como o Eprints ou Dspace. Para o acesso livre o que está faltando é a motivação por parte dos autores. As evidências dos benefícios decorrentes do acesso livre, como o aumento do uso e do impacto dos resultados de pesquisa parece ainda não chegar até eles. Um dos fatores apontado como o principal estimulador do autodepósito nas instituições que criam seus repositórios digitais tem sido a elaboração e implementação de políticas mandatórias que obrigam o autodepósito por arte dos autores, e não unicamente o argumento de o acesso livre como mais uma camada na plataforma genérica institucional da preservação.

Mesmo que *softwares* como o EPrints, Dspace e Fedora possam prover longevidade e acompanhem o desenvolvimento na área da preservação digital, a motivação primeira não é a preservação de longo prazo. Dos 2,5 milhões de artigos publicados anualmente nos mais de 24 mil periódicos existentes no mundo, 85% não estão sendo autodepositados, assim como não têm sido as edições anteriores desses periódicos. Os conteúdos, no entanto, não estão perdidos, pois continuam preservados nas versões proprietárias pelos produtores (editores) juntamente com as bibliotecas, suas compradoras.

As ferramentas para repositórios institucionais não são originalmente destinadas a projetos de preservação digital. A motivação primordial não diz respeito à garantia da longevidade dos conteúdos digitais, mesmo que elas acompanhem o desenvolvimento nas áreas de preservação de longo prazo de conteúdos digitais. O conteúdo dos repositórios institucionais em ferramentas para repositórios digitais

pode ser preservado, mesmo que seu objetivo não seja a preservação, e principalmente, porque o conteúdo de acesso livre está em risco de não ser preservado devido ao fato de ele não estar sendo totalmente disponibilizado, sem restrições.

2.3.2 Preservação de periódicos eletrônicos

Os periódicos eletrônicos estão deixando de representar apenas uma extensão da versão impressa, para serem uma garantia do acesso à informação científica, e projetos de criação de arquivos de *e-journals* continuam aparecendo. As evidências demonstram que as publicações periódicas são mais facilmente acessíveis nas novas mídias eletrônicas (via *web*), assim como esses formatos podem oferecer algum tipo de garantia de que estarão disponíveis por muito mais tempo.

A opção pela publicação eletrônica, disponível desde os anos 1980, vem crescendo a cada dia. Em 2006 eram 14.338 títulos de periódicos registrados no *Ulrich's International Periodicals Directory,* representando 62% de todos os periódicos referenciados nesse diretório (os 38% restantes tinham versão apenas em papel). Os periódicos de acesso livre eletrônicos já eram 1.429. A Electronic Publishing Services, em estudo sobre o futuro das publicações científica elaborado em 2003, projetou que, em 2016, a metade dos periódicos já terá migrado para a versão unicamente eletrônica; predizendo que as áreas das ciências exatas, da saúde e das tecnologias serão as primeiras a fazê-lo (COUNCIL ON LIBRARY AND INFORMATION RESOURCES, 2006).

No caso das bibliotecas no mundo, a EBSCO estima que as assinaturas de títulos impressos das áreas das ciências exatas, saúde e das tecnologias, assinaturas de versões analógicas de periódicos serão superadas pelas assinaturas das versões *on-line* (BRUCE, 2005). Outros levantamentos mostram que 84% das bibliotecas universitárias do mundo cancelam a assinatura da versão impressa quando a versão eletrônica fica disponível (HAHN, 2006).

Pensar na necessidade de preservação dos periódicos eletrônicos significa captar o valor do artigo eletrônico quanto às ligações que ele possui:

gráficos e bases de dados multimídia. São elementos que têm se tornado partes intrínsecas da moderna literatura científica. As publicações científicas de universidades e instituições de pesquisa estão adotando os arquivos de acesso livre porque a informação que elas produzem está, na sua maioria, em formato digital, o que facilita a sua transferência, a transposição do problema da obsolescência da informação científica, assim como a criação de um espaço para a produção e armazenamento dentro da própria instituição. Trata-se da garantia da existência de um avançado conjunto de serviços no qual está incluída, principalmente, a questão da preservação digital (CAFÉ *et al.,* 2003).

No modelo atual de publicação científica, a preocupação dos responsáveis pelas bibliotecas e centros de informação está na centralização dos periódicos em poucas bases de dados, que detêm o acesso e o controle das publicações, sem permitir o direito a reter cópias digitais nas suas próprias instituições.

A preservação digital de periódicos eletrônicos está dirigida especialmente a assegurar que seus conteúdos poderão ser acessados sempre de alguma maneira. Os gerentes das bibliotecas precisam reconhecer que a preservação dos objetos digitais é uma forma de gerenciar riscos e um direito ante o possível descumprimento de acordo com as editoras e donos de bancos de dados. Para alguns editores, a preservação é considerada uma extensão da sua missão com vistas a preservar o conhecimento de uma área e justificar os recursos conseguidos para as suas atividades.

Em torno das discussões sobre preservação digital de longo prazo, os periódicos eletrônicos aparecem sempre em primeiro plano. Isso se deve ao seu papel na comunicação científica e pelos interesses comerciais envolvidos. A solução sugerida por vários grupos de pesquisadores é que sejam construídos repositórios digitais fora do controle dos editores e provedores de bases de dados comerciais para os periódicos, assim como para outros tipos de publicações científicas.

Preservar periódicos eletrônicos é uma responsabilidade dos editores e das bibliotecas depositárias. A aplicação de alguma das metodologias mencionadas dependerá da política de preservação que elas adotam. No caso das editoras que não permitem que nenhum dos seus autores e leitores copie e deposite versões dos artigos originais em repositórios institucionais, elas estão declarando a sua responsabilidade pela preservação e o acesso aos artigos de longo prazo. Alguns casos apresentam mais características restritivas ao acesso quando os editores

provêem acesso temporário apenas a arquivos criptografados, criando maiores dificuldades de recuperação, se porventura a editora fechar.

No caso específico de periódicos científicos de acesso livre, quando o editor permite aos autores o autodepósito da versão final dos autores como suplementos da versão final em repositórios digitais, essas versões não substituem os originais. Os suplementos autodepositados servem apenas para prover o acesso imediato e contínuo para aqueles que não têm como pagar o acesso à versão final. Eles não funcionam como mecanismo de preservação da versão oficial, mesmo que nesse aspecto cheguem um dia a servir para ajudar também.

As iniciativas de Acesso Livre parecem beneficiar a preservação digital dos periódicos científicos, mas a motivação principal delas é o acesso. Tanto para Suber (2003) como para Harnad (2007), a preservação de longo prazo, assim como o autodepósito em repositórios, é apenas um dos vários requisitos para o acesso livre. Na visão do Movimento da OAI, esses requisitos não estão inerentemente unidos, pois, ao colocá-los juntos, corre-se o risco de sobrecarregar o conceito principal (acesso livre) e retardar sua implementação com condições que a Iniciativa realizará.

A discussão continua sobre se a preservação é parte essencial ou não do acesso livre. Segundo Suber (2003), as grandes iniciativas de acesso livre diferem na inclusão ou não de medidas para garantir a preservação de longo prazo. Para esse autor, a Declaração de Bethesda e a da BioMed Central (BMC) incluem a preservação, enquanto as de Budapest Open Access Initiative (BOAI) e a da Public Library of Science (PLoS) não o fazem. Suber também afirma que, o acesso livre, apesar de estar enfocado na disseminação imediata, atua como catalisador da necessidade de preservação de longo prazo, sem as restrições dos direitos autorais.

Nas estatísticas do Movimento dos Arquivos Abertos, entre 10% e 20% de todos os periódicos científicos no mundo já possuem algum tipo de política de autodepósito para as versões suplementares, e mesmo esses periódicos precisam enfocar os objetivos reais da preservação para as suas versões proprietárias originais. Com o crescimento do número de soluções para a criação e gerenciamento de periódicos de acesso livre, os editores científicos não precisam se preocupar com a necessidade de aumento da visibilidade, uso e impacto dos seus conteúdos. A atenção desses editores com a preservação deve recair agora nas

potencialidades dessas soluções com relação ao arquivamento das versões originais.

2.3.2.1 Arquivamento de periódicos eletrônicos

No uso de programas proprietários e que podem tornar-se obsoletos, os editores enfrentam o problema de que a decisão de autorizar a migração para um novo formato não entre em conflito com os direitos autorais aceitos pelos autores dos artigos. Algumas grandes bases de revistas, como a Elsevier, já manifestaram a promessa de manter os arquivos originais preservados.

O arquivamento do periódico eletrônico é uma das tarefas editoriais que precisa de uma política clara para a criação de cópias arquiváveis. As cópias em papel estão diminuindo como opção de *back-up* e o fator da redundância não é considerado por muitos editores. As editoras de universidades americanas, sob auspícios de instituições de fomento, como a *Mellon Foundation*, têm trabalhado em projetos com editores individuais e por área de conhecimento, para estudar as principais características das mudanças que as versões eletrônicas envolvem e as ferramentas de *software* mais apropriadas para o arquivamento dos seus originais.

Algumas dessas iniciativas têm optado por metodologias para garantir a permanência dos formatos dos arquivos, como a do serviço *Electronic Journal Platform da* American Institute of Physics (AIP, 2003), que permite que vários documentos sejam criados a partir da marcação em SGML ou de uma DTD, colocando uma versão em formato ASCII para ser disponibilizada em processos de preservação no futuro. Outras iniciativas de preservação digital de periódicos científicos têm optado pela metodologia da migração cíclica. Como a American Geophysical Union, que em 1996 criou um fundo para a manutenção e atualização das versões eletrônicas dos periódicos da área. Também a American Astronomical Society usa parte dos recursos financeiros arrecadados com as assinaturas para apoiar os processos de migração da versão eletrônica da sua publicação, a cada cinco anos.

Várias instituições em países desenvolvidos têm financiado projetos para a criação de repositórios digitais de periódicos eletrônicos. Em 2001, um projeto da

Harvard University Library foi financiado pela Andrew W. Mellon Foundation, com o objetivo de analisar os mecanismos por meio dos quais os periódicos eletrônicos deverão ser selecionados para serem preservados, assim como outros assuntos relacionados com o acesso de longo prazo. A proposta desse projeto é mostrar a importância da integração das iniciativas em programas de preservação que operem em conjunto (HARVARD, 2000). Dentro desse mesmo projeto foi realizado um estudo que indicava o desenvolvimento de uma DTD comum para todos os periódicos eletrônicos, mas percebeu-se que ela teria limitações devido às especificações de cada área de conhecimento (INERA, 2001).

Em 2003, um estudo da Association of Learned and Professional Society Publishers mostrou que 52% dos editores comerciais e 45% dos editores não comerciais entrevistados já usavam algum tipo de solução para cumprir com a sua responsabilidade de preservação de longo prazo. Mesmo assim, para as editoras comerciais, os direitos autorais continuam nas mãos dos editores, e por serem eles os beneficiários comerciais, tornam-se os responsáveis pelo seu controle e preservação. Alguns sistemas como o *Journal Storage: The Schorarly Journal Archive* (JSTOR) e o *Lots of Copies Keeps Staff Save* (LOCKSS) foram citados. Segundo a pesquisa, os editores começam a perceber os benefícios econômicos que traz o “reuso” dos conteúdos das suas publicações (COX; COX, 2003).

Internacionalmente reconhecida, a base de dados JSTOR é o resultado de um consórcio entre instituições de ensino superior (IES) e institutos de pesquisa, promovido por uma organização independente, sem fins lucrativos, estabelecida nos Estados Unidos, desde 1995. A proposta do JSTOR é armazenar e tornar disponível o texto integral de coleções eletrônicas retrospectivas (*backfiles*) de periódicos, a partir do primeiro número. O serviço é pago e o período de abrangência de cada revista é variável.

2.3.2.2 LOCKSS

Um dos projetos direcionados a esse objetivo é o *Lots of Copies Keep Stuff Safe* da Stanford University (LOCKSS, 2003). O método de arquivamento desse projeto está direcionado para conservar a integridade das publicações

eletrônicas mantendo cópias em vários endereços eletrônicos (*sites*), conferindo periodicamente todas as cópias para verificar a congruência da informação armazenada. A solução LOCKSS procura minimizar o grau de impacto que um desastre possa causar em um único arquivo digital.

Depois de cinco anos de testes, em 2004 a Stanford University lançou o LOCKSS com o apoio financeiro da National Science Foundation (NSF), da Sun Microsystems e da Mellon Foundation. Desde então, uma corporação LOCKSS provê suporte tecnológico e para o gerenciamento das coleções digitais (REICH, 2003).

O funcionamento do *software* livre LOCKSS precisa de uma infra-estrutura de *hardware* de custo reduzido e uma administração técnica mínima. Considerando o alto custo do armazenamento em repositórios, uma metodologia como a do LOCKSS é muito importante para os países em desenvolvimento, onde a maioria das instituições carece de financiamento para as atividades relacionadas com a preservação digital de publicações científicas. A solução permite que os leitores autorizados tenham acesso às edições eletrônicas dos periódicos, respeita os direitos das editoras e não aumenta o orçamento das bibliotecas (REICH; ROSENTHAL, 2001). Ela compreende:

1) a preservação através da replicação distribuída;
2) a aplicação de uma tecnologia de apuração de dados pelos pares;
3) o uso do repositório como núcleo das atividades;
4) a inclusão de quatro das seis entidades funcionais do modelo OAIS: Recepção, Administração, Armazenamento e Acesso.[26]

O modelo LOCKSS foi desenhado para criar um sistema descentralizado de publicação e acesso permanente na *web*, apoiado na redundância de sistemas distribuídos para preservar os conteúdos, seu acesso e a habilidade de entender seus *bits*.

A preservação digital que o LOCKSS permite não se reduz à disponibilização perpétua dos materiais armazenados, mas se estende às funções que garantem a preservação da propriedade intelectual dos documentos em qualquer formato digital. Os responsáveis pelo seu desenvolvimento adquirem e disponibilizam cópias das publicações de uma forma análoga à da biblioteca, com

[26] Declaração de Conformidade com o a norma OAIS/ISO 14721:2003 Disponível em: http://www.lockss.org/lockss/OAIS Acesso em: 15/set/2008

uma diferença: os registros não são descartados e seu acesso é salvaguardado (LOCKSS, 2007).

Com o LOCKSS, as instituições que assinam periódicos científicos podem escolher a preservação dessas publicações, criando versões (*caches*) persistentes e autorizadas de baixo custo dos conteúdos que elas assinam. O servidor LOCKSS guarda permanentemente as versões que ele coleta das novas edições, comparando-as de forma lenta com os conteúdos de outras versões mantidas em diferentes bibliotecas, corrigindo e detectando possíveis erros nos arquivos de qualquer um dos repositórios que formem uma rede cooperativa de réplicas LOCKSS (Figura 12).

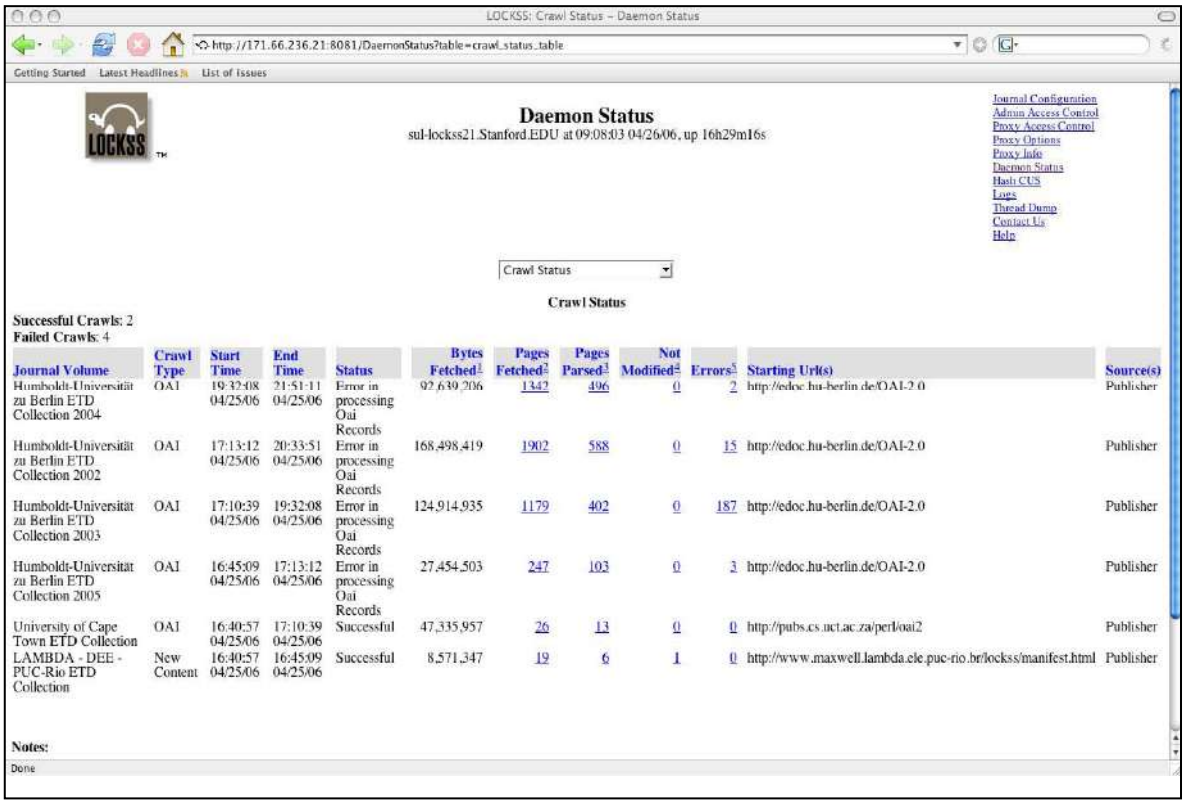


Figura 12 – Exemplo de monitoramento de caixas LOCKSS
Fonte: Projeto LOCKSS[27]

É recomendado que o uso dos repositórios ou "caixas" LOCKSS (LOCKSS, 2007) esteja sob a responsabilidade de bibliotecas que garantam o acesso a periódicos científicos na Internet. O sistema permite que as bibliotecas

[27] Disponível em: http://171.66.236.21:8081/DaemonStatus?table=crawl_status_table. Acesso em: 10/set/2007

voltem a tomar a custódia das revistas para fins de preservação, mesmo que os editores deixem de publicá-las tanto na versão impressa quanto na eletrônica.

Um compromisso se estabelece entre as editoras e as bibliotecas para que as instituições possam coletar, armazenar, arquivar e preservar conteúdo autorizado e salvaguardar o acesso das suas comunidades. Dessa maneira, o modelo LOCKSS reforça o controle de acesso por parte dos editores, sem mudar os seus modelos de negócios, sejam eles restritos ou abertos. Os editores permitem que as caixas do LOCKSS funcionem, publicando um termo que registra a sua aprovação para que o processo ocorra. Nos novos sistemas de gerenciamento eletrônico de editoração, como o *Open Journal Systems* (OJS/SEER) o termo já está incorporado, deixando ao editor a decisão de permitir que as caixas LOCKSS consigam identificar o endereço de IP e os conteúdos autorizados.

2.3.2.3 Gerenciamento editorial e preservação digital

A preservação digital começa a ser considerada uma extensão da missão das publicações científicas com vistas a preservar o conhecimento de uma área e justificar os recursos conseguidos para suas atividades. O trabalho de coletar e preservar materiais para o acesso futuro normalmente não é de responsabilidade direta do cientista ou do criador do material. Esse é um problema do editor e das bibliotecas que têm de preservar materiais por muitos anos. Raramente profissionais que não são bibliotecários ou arquivistas sabem dessa tarefa, até o momento que os documentos começam a sofrer danos ou são perdidos.

Na última década tem crescido o número de soluções para criação e gerenciamento de periódicos; com isso, os editores científicos não precisam se preocupar com a necessidade de aumento da visibilidade, uso e impacto dos seus conteúdos. A preocupação desses editores com a preservação recai agora nas potencialidades dessas soluções com relação ao arquivamento das versões originais.

Tanto na oferta de soluções comerciais como na de *softwar*e livre para a editoração eletrônica de revistas, a variedade oferecida aumenta a cada dia. O objetivo de todas elas está orientado a ajudar no cumprimento das tarefas editoriais

e na visibilidade das etapas da edição. As características que esses *softwares* possuem em comum são as seguintes:

1. cadastro de todos os usuários do sistema;
2. produção de relatórios e estatísticas de uso;
3. adaptabilidade ao processo editorial de cada publicação;
4. suporte técnico profissional;
5. produção e envio automático de mensagens;
6. submissão de trabalhos *on-line*;
7. variedade de formatos de publicação;
8. aplicativos opcionais.

Entre as ferramentas que surgiram na última década e que favorecem a edição completa de artigos científicos na *web*, estão as relacionadas no Quadro 4.

Nome	Ano Inicial	Software Livre	Software Proprietário	Instituição/Empresa
Open Journal Systems (OJS/SEER)	1998	X		British Columbia University, Fraser University
OpenACS	1998	X		ArsDigita
SciX Open Publishing Services (SOPS)	2001	X		University of Ljubljana (Slovenia)
Article System	2002	X		Sourceforge.net
TOPAZ	2003	X		Edgewall Software
Digital Publishing System (DPubS)	2004	X		Cornell University Library
GAPworks	2005	X		German Academic Publishers
Hyperjournal	2005	X		HyperJournal Association
EdiKit©	1999		X	Berkeley electronic Press
EJPress	1999		X	eJournalPress
Electronic Submission and Peer Review (ESPERE)	2000		X	Electronic Libraires Programme (eLib) of the Higher Education Founding Council for England (HFCE)
Rapid Review™	2000		X	Cadmus Journal Services
Bench>Press™	2001		X	HighWire Press®, Stanford University Libraries
Manuscript Central™	2001		X	ScholarOne da Thompson Business
Xpress Track™	2001		X	XpressTrack™
Editorial Manager®	2001		X	Aries Systems Corp.
Allen Track™	2002		X	eJournalPress.com

Quadro 4 – Ferramentas para editoração eletrônica de revistas

O *Open Journal Systems* (OJS), *software* livre, de fonte aberta, desenvolvido pelo *Public Knowledge Project* originário da British Columbia University (Canadá) foi personalizado e traduzido pelo Instituto Brasileiro de Informação em

Ciência e Tecnologia (Ibict), com o nome de SEER (Sistema Eletrônico de Editoração de Revistas). O instituto presta assistência aos editores em cada uma das etapas do processo de editoração e pode ser adaptado de acordo com as próprias políticas de publicação (MÁRDERO ARELLANO, 2004). É o primeiro *software* em português de editoração eletrônica que possui o protocolo OAI para intercâmbio de dados essenciais (metadados) e que contem mecanismos para a preservação de seu conteúdo em conjunto com o projeto LOCKSS (Figura 13).

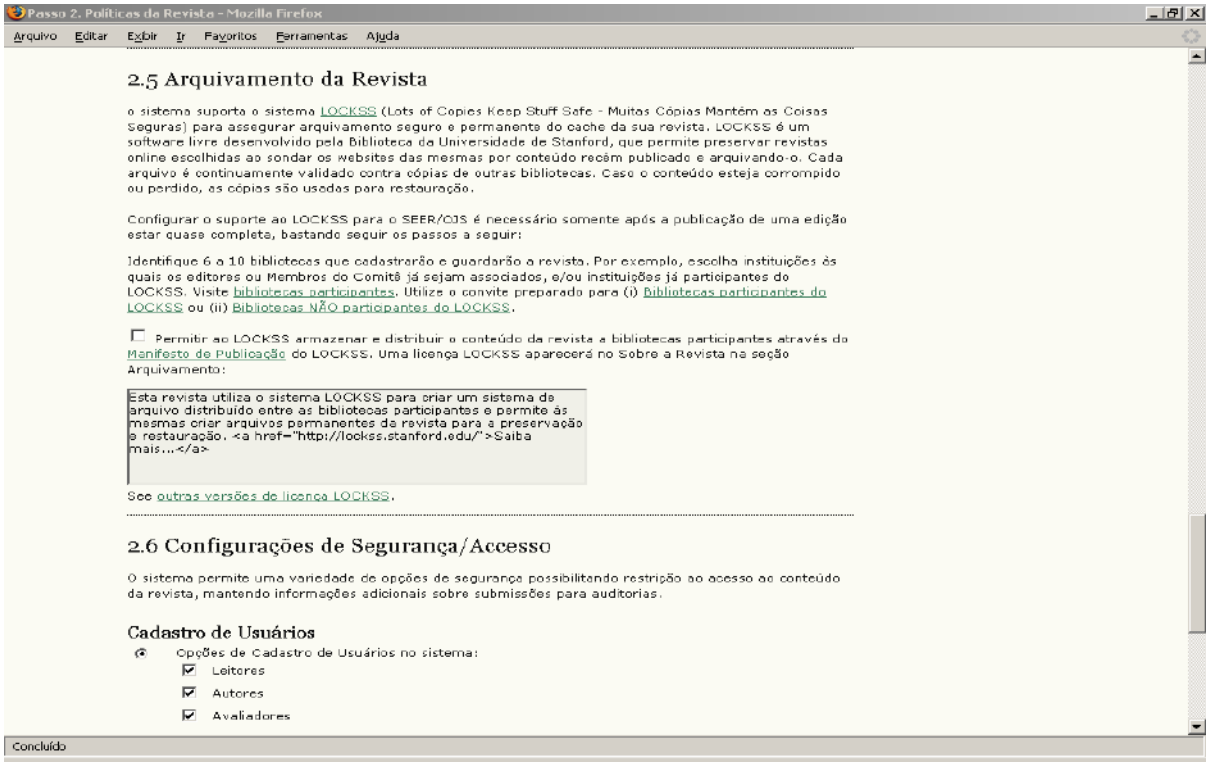


Figura 13 - Página do OJS/SEER opção Arquivamento da Revista no LOCKSS
Fonte: Portal do SEER[28]

Outro passo importante tem sido o que as associações de editores de periódicos científicos e de pesquisa, como a International Union of Pure and Applied Physics (IUPAP), têm recomendado aos pesquisadores, editores e bibliotecários da área que desenvolvam formas de registro dos documentos que incluam informações sobre *hardware* e *software* para que isso sirva como uma ajuda em casos de possíveis migrações ou risco de perda dos dados e o uso de XML e outros formatos padrão apropriados à área (SMITH, 2001).

[28] Disponível em: http://seer.ibict.br. Acesso em: 10/out/2008.

Os sistemas de gerenciamento editorial têm encorajado os editores a usar linguagens de marcação e esquemas XML para permitir que o material seja extraído, integrado a outros serviços, provendo aos aplicativos da *web* 2.0 os conteúdos que os usuários obtêm sob demanda (MORGAN, 2000). Os editores que optam por sistemas baseados em SGML/XML estão disponibilizando formatos próprios para a preservação como parte de seu processo normal de editoração.

2.3.3 Preservação digital em repositórios digitais

As bibliotecas, através do tempo, têm cumprido a missão de permitir a descoberta e a entrega de material acadêmico, mas elas não podem oferecer aquilo que não coletam ou mantêm. As bibliotecas na iternet são importantes para os pesquisadores porque viabilizam a descoberta do que foi catalogado e publicado *on-line*. Para organizar os documentos com informações científicas em uma biblioteca digital, os profissionais da informação preparam metadados (descritivos, estruturais e de preservação), arquivos para preservação, arquivos para acesso, diretórios, projetos de documentação e repositórios de objetos digitais, seguindo práticas adequadas de descrição.

As bibliotecas digitais, os repositórios digitais e os periódicos de acesso livre estão requerendo mudanças em seus usuários, no que se refere às práticas de trabalho e participação no esforço global, para definir e classificar seus documentos segundo novos modelos de comunicação científica.

O estudo da preservação digital nas subáreas da ciência da informação tem passado do reconhecimento de sua necessidade e da sua definição, para abordagens mais técnicas e propostas de ações mais claras. Para alguns especialistas, as bibliotecas digitais são consideradas o caminho mais adequado para a preservação dos recursos de informação (HILDRETH, 1995). Lesk (1997) resume essa afirmação destacando que as bibliotecas digitais são meios mais dinâmicos para as atividades de preservação digital do que as bibliotecas tradicionais, considerando a sua adaptação às freqüentes mudanças tecnológicas.

Em contrapartida, pesquisadores de vários projetos que envolvem *softwares* para repositórios digitais asseguram que os repositórios institucionais são

os lugares adequados para testar e formular as metodologias e políticas a serem adotadas pelos provedores de informação científica. O argumento no qual eles se baseiam é que os repositórios para preservação de objetos digitais devem estar localizados em instituições confiáveis e capazes de armazenar, migrar e dar acesso a coleções digitais (OCLC/RLG, 2002).

Os responsáveis pelos acervos digitais confiam nos repositórios institucionais poder preservar e dar acesso a material não apenas publicado em periódicos avaliados pelos pares, mas também à literatura cinzenta, como teses, relatórios, documentos governamentais e, ainda, materiais suplementares, conjuntos de dados, imagens, visualizações e simulações, comunicação informal como *e-mails*, *blogs*, *podcasts*, *websites*, *wikis* e apresentações.

Atualmente, os repositórios digitais podem integrar ferramentas de preservação digital, ou aquelas que sejam equivalentes nessa funcionalidade. Uma instituição pode também optar por definir um tipo de *workflow* que integre ferramentas em determinados momentos do processo. Entre as ferramentas atualmente disponíveis estão as seguintes:

a) as que geram e capturam metadados: NLNZ metadata extractor;[29]
b) as que identificam e avaliam os formatos de arquivo: JHOVE[30] e DROID[31];
c) as que padronizam arquivos para formatos preserváveis ou formatos abertos para a preservação: XENA[32] da National Archives of Australia, entre outros.

Esses são alguns exemplos de sistemas e soluções tecnológicas que estão sendo testados e que são considerados padrões de preservação digital pela comunidade científica, mas os pesquisadores estão muito distantes de uma automação completa em sistemas de preservação digital.

[29] Disponível em: http://www.natlib.govt.nz/about-us/current-initiatives/metadata-extraction-tool Acesso em: 13/out/2008
[30] Disponível em: http://hul.harvard.edu/jhove/ Acesso em: 13/out/2008
[31] Disponível em: http://droid.sourceforge.net/wiki/index.php/Introduction Acesso em: 13/out/2008
[32] Disponível em: http://xena.sourceforge.net/ Acesso em: 13/out/2008

2.3.3.1 Bibliotecas digitais

Segundo Webb (2000), as bibliotecas tradicionalmente são responsáveis por manter coleções para uso permanente, protegendo-as de ameaças, ou salvando-as e reparando-as para compensar seus impactos. No mundo das bibliotecas digitais, um desafio importante a ser enfrentado é a capacidade de poder acolher diferentes tipos de formatos inovadores, assim como o planejamento de suas adaptações no futuro.

Uma biblioteca digital pode incluir inúmeros programas e aplicativos usados para criar e estruturar uma coleção digital. São *softwares* e sistemas que servem para gerenciar e administrar esses recursos digitais, antes de serem depositados e publicados. Essas ferramentas influenciam a maneira como os documentos serão preservados e os respectivos custos envolvidos no processo.

Os sistemas informatizados para bibliotecas digitais atuais requerem acentuado esforço de gerenciamento de coleções digitais, que vai além das tarefas tradicionais das bibliotecas (ex.: aquisição, seleção, classificação, arquivamento), na construção da interoperabilidade de acervos digitais (arquiteturas, metadados, formatos padrão), a qual é possível por meio de sistemas relacionados e desenvolvidos para propósitos e comunidades específicas (ARMS, 2000).

Para Borbinha e Correia (2001), a biblioteca digital apresenta um dos primeiros agrupamentos dos requisitos associados com as atividades tradicionais de preservação como é possível acompanhar no Quadro 5.

Aquisição	Entrega pelo produtor	
	Captura pela biblioteca	
	Coletado pela biblioteca	
Verificação	Integridade física (meio)	
	Integridade do conteúdo	Integridade lógica
		Autenticação
Registro	Metadados	Descrição bibliográfica
		Instalação e manipulação
		Acesso
		Preservação
Preservação	Preservação física	Refrescamento do meio
		Migração de suporte
	Preservação lógica	Conversão de formatos
		Emulação
	Preservação intelectual	
Acesso	Condições de uso	Acesso local
		Acesso remoto

Quadro 5 - Estratégias de preservação no fluxo dos objetos digitais na biblioteca digital Fonte: Borbinha e Correia (2001)

Esses autores mencionaram que, entre as responsabilidades de preservação de uma biblioteca digital, estão as relacionadas com os seguintes aspectos:

a) conversão de formatos digitais;
b) tratamento dos metadados;
c) usabilidade dos dados disponibilizados;
d) infra-estrutura para manutenção dos sistemas;
e) desenvolvimento de *software*;
f) estruturação dos mecanismos de pesquisa;
g) consultoria sobre recursos digitais.

Entretanto, a preservação digital em mãos das bibliotecas nacionais evidencia outros subconjuntos de responsabilidades:

a) publicar: depósito externo, edição digital local, digitalização e metadados estruturais;
b) registrar e preservar: armazenamento, gestão de identificadores e catalogação;

c) acessar: local e remotamente;
d) disseminar e pesquisar: catálogos *on-line*, índices temáticos (*Web Services,* OAI-PMH, Z39.50).

Seja qual for o tamanho ou a importância da biblioteca, verifica-se uma diferença entre os repositórios e as bibliotecas digitais. Os repositórios de informação digital são responsáveis coletivamente por garantir, por meio de alguma estratégia de preservação, a integridade e o acesso de longo prazo da herança social, econômico-cultural e intelectual de um país na forma digital. Uma biblioteca digital é um repositório que coleta e provê acesso à informação digital, mas pode também não garantir o armazenamento de longo prazo e o acesso à informação.

Muitas bibliotecas digitais podem ou não ser repositórios digitais, mas os repositórios digitais realizam algumas funções de biblioteca digital pelo fato de elas também selecionarem, obterem, armazenarem e proverem acesso à informação digital. Especificamente na área de preservação digital, os repositórios digitais são considerados parte da infra-estrutura de serviços de uma biblioteca digital.

2.3.3.2 Repositórios Digitais

Um repositório digital é um serviço de armazenamento de objetos digitais que tem a capacidade de manter e gerenciar materiais por longos períodos de tempo e prover o seu acesso apropriado. Os repositórios digitais dividem-se em temáticos, institucionais e centrais. Os repositórios temáticos cobrem determinada área do conhecimento. Já os repositórios institucionais (RI) são sistemas de informação que armazenam, preservam, divulgam e dão acesso à produção intelectual de instituições e comunidades científicas, em formato digital, enquanto os repositórios centrais são provedores de serviços nacionais e internacionais que permitem a reunião de dados coletados tanto de bibliotecas digitais, quanto de repositórios temáticos e repositórios institucionais.

O propósito dos repositórios estaria dirigido a dois aspectos: o acesso e o armazenamento. As universidades de vários países estão adotando esta forma de publicação para apoiar a divulgação dos resultados de pesquisas científicas, criando

políticas institucionais para legitimar os repositórios de acesso livre como detentores de produção científica reconhecida por órgãos de fomento (LYNCH, 2003).

Os repositórios digitais compreendem um conjunto de ferramentas necessárias para os produtores, disseminadores e usuários de documentos digitais. Os projetos de preservação digital devem atender a pré-requisitos e serem parte de um processo transparente de disponibilização a longo prazo dos objetos digitais. Os desenvolvedores desses projetos são responsáveis pela aplicação dos padrões reconhecidos internacionalmente de preservação digital.

Já em 1996, a *Task Force on Archiving of Digital Information* da *Commission on Preservation & Access* (CPA) e a RLG analisou as possibilidades dos repositórios digitais, chegando às seguintes conclusões:

a) com a finalidade de preservar a longo prazo a informação digital, em uma escala adequada à demanda da pesquisa científica no futuro, as instituições requererão uma infra-estrutura apropriada capaz de suportar um sistema de distribuição de arquivos digitais;
b) um componente crítico da infra-estrutura de arquivamento digital é a existência de um número suficiente de organizações confiáveis capazes de armazenar, migrar e prover acesso às coleções digitais (WATERS & GARRETT, 1996 p. 46).

A gênese dos sistemas para repositórios digitais ocorreu no final do ano 2000, quando a University of Southampton na Inglaterra, lançou o pacote de *software* chamado EPrints. Desde então, o movimento para a construção de repositórios digitais tem crescido consideravelmente, motivado pela convergência entre a redução dos custos no armazenamento *on-line*, a proliferação da banda larga e das tecnologias de redes de *gigabits* e o desenvolvimento de padrões de metadados para descrever o conteúdo dos repositórios.

As características principais dos repositórios para objetos digitais foram mencionadas por Stewart Granger (2002) como aquilo que formaria a “infra-estrutura profunda”, relacionada com toda a parte organizacional, aspectos legais e culturais, assim como toda a parte tecnológica. Para ele, o problema principal está em que essa infra-estrutura realmente responda às necessidades da instituição e de seus usuários, e não apenas as das empresas que comercializam essas tecnologias. Para que isso seja possível, ele menciona a necessidade de que existam

repositórios certificados, assim como mecanismos de colaboração que facilitem o intercâmbio entre a comunidade envolvida no tratamento de materiais digitais.

Esses repositórios incentivam o autodepósito e publicação na rede gerenciada pelo pesquisador (*self-archiving*), utilizam novas tecnologias abertas (*open source*), tornando disponíveis as informações para serem acessadas permanentemente por diversos provedores de serviços em nível nacional e internacional.[33] Tal mudança na publicação científica iniciou-se com a constituição dos chamados repositórios digitais temáticos. A idéia evoluiu para o agrupamento desses repositórios sob a responsabilidade de instituições centradas na divulgação da produção científica local (CAFÉ *et al*., 2003).

O crescimento do número de repositórios digitais no mundo está sendo acompanhado pela Iniciativa dos Arquivos Abertos, como mostra a Figura 14.

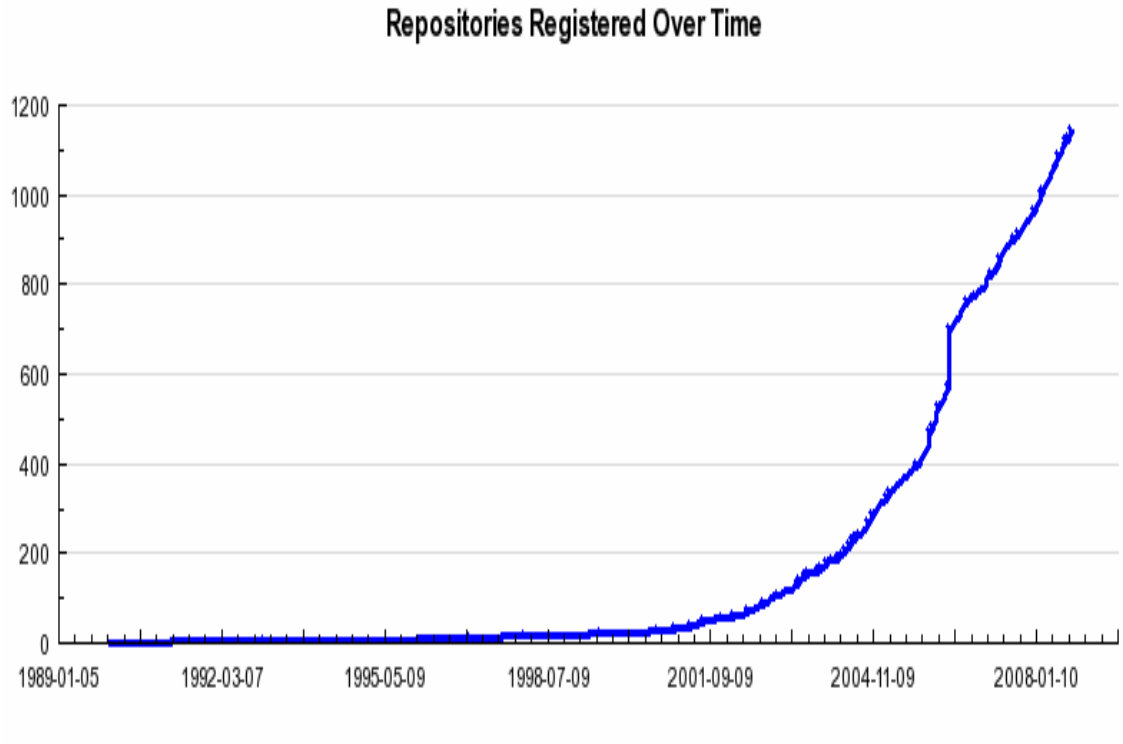


Figura 14 - Crescimento dos arquivos/repositórios[34]

[33] Ainda na sua fase inicial, poucos repositórios têm alcançado tamanho e importância consideráveis. Entre os principais estão Fedora (*Flexible Extensible Digital Object and Repository Architecture*), CDSWare, Eprints, Archimene, ARNO, iTor, MyCoRe. Alguns provedores de serviço baseados no Protocolo OAI, que permitem a busca em mais de 250 arquivos abertos, são Arc, Citabase, MyOAI, Perseus, Scirus, OAIIECSP, Callima, CYCLADES, OAIster, SAIL-eprints, TORII e PKP. Um dos primeiros projetos de repositórios temáticos foi o JSTOR (*The Scholarly Journal Archive* – URL: http://www.jstor.org/), criado para servir, apenas, como repositório de preservação digital de periódicos científicos de algumas áreas.

[34] Fonte: ROAR. Disponível em: http://roar.eprints.org/index.php?action=generate_chart Acesso em: 2/nov/2008.

Como demonstra a Figura 14, o aumento do número de repositórios digitais começou a ser expressivo a partir de 2002, o que se deve ao aparecimento de sistemas para repositórios institucionais e seu impacto na infra-estrutura das bibliotecas digitais existentes.

2.3.3.3 Repositórios Institucionais

Desde suas primeiras implementações, os repositórios institucionais têm sido voltados para a informação científica, contendo os trabalhos de pesquisa e ensino de comunidades acadêmicas, assim como a documentação de ampla variedade de atividades das próprias universidades. Os repositórios institucionais podem ser pensados como parte de uma ação dentro da gestão documental, que inclui sua criação, tratamento, transmissão e acesso. Crowe (2002, p. 4) define os repositórios institucionais como "[...] um arquivo digital de produtos intelectuais criados por uma comunidade de pesquisadores, estudantes e professores de uma instituição [...]". Para Lynch (2003), eles representam serviços e um compromisso:

> "Segundo meu ponto de vista, um repositório institucional acadêmico é um conjunto de serviços que a universidade oferece para os membros da sua comunidade, para o gerenciamento e disseminação do material digital criado pela instituição e pelos seus membros. É essencialmente o compromisso de uma instituição de cuidar do material digital, incluindo a preservação a longo prazo, quando apropriada, a organização, acesso e distribuição." (LYNCH, 2003 s/p.).

O primeiro impacto que os repositórios institucionais causaram foi a criação de bibliotecas digitais de teses e dissertações, que permitiu que outros documentos pouco divulgados por essas instituições fossem incluídos nos repositórios. Regularmente são materiais importantes para as equipes e empregados dessas organizações. A missão dos repositórios institucionais ficou centralizada na produção, manutenção, distribuição e preservação de grande variedade de recursos de informação em rede para pesquisadores, professores e estudantes de dentro e fora de uma instituição.

A primeira abordagem oficial a respeito dos repositórios institucionais foi da Association of Research Libraries, na qual são mencionados alguns exemplos

(CROWE, 2002). Nos Estados Unidos, o governo tem usado repositórios institucionais há muito tempo (Defense Technical Information Center, Department of Energy's Office of Scientific and Technical Information, NASA Center for AeroSpace Information, National Technical Information Service e *a* Government Printing Office). Na Europa, um dos primeiros países a adotar os repositórios institucionais foi a França, no seu Institut de l'Information Scientifique et Technique (INIST-CNRS).

Na década de 1990, diferentes tipos de organizações começaram a armazenar eletronicamente ampla variedade de materiais, resultando em grandes repositórios que precisam ser agora preservados. A responsabilidade de se ter optado por usar repositórios digitais vem crescendo, junto com a necessidade de preservar por longo prazo os materiais neles contidos, sua organização, acesso constante e distribuição (LYNCH, 2003). Dependendo das circunstâncias, a responsabilidade de alimentar os repositórios, de organizar o conteúdo e de preservá-lo terá prioridade na construção de repositórios digitais abertos (PINFIELD; JAMES, 2003).

A implementação de repositórios institucionais levanta uma série de questões complexas sobre os recursos e estratégias organizacionais, assim como questões sobre os papéis e responsabilidades. Muitos projetos de repositórios institucionais são motivados pelo desejo de mudar a comunicação científica. Essa mudança, se necessária, poderá questionar a responsabilidade das instituições de ensino de publicar sua produção, retirando esse papel das grandes editoras comerciais.

Na área de inovação tecnológica para o gerenciamento de conteúdos digitais, os repositórios institucionais (RI) estão sendo considerados uma forma eficaz de preservação da produção intelectual dos especialistas de uma ou várias instituições. Construídos no contexto das universidades e instituições governamentais, eles garantem a visibilidade das coleções digitais pela potencialização do acesso à informação e pela criação de indicadores da qualidade da produção científica nas organizações, mostrando a relevância de suas atividades e o valor da instituição.

Segundo Crowe (2002), os repositórios digitais são uma resposta a dois assuntos estratégicos das IES: a) prover um sistema que expanda o acesso a resultados de pesquisa, garantir o controle da produção acadêmica, aumentar a competição e reduzir o monopólio dos periódicos, diminuir os custos e destacar a

relevância da própria instituição e das bibliotecas; e b) servir como indicadores potenciais da qualidade da universidade e demonstrar a relevância científica, social e econômica das atividades de pesquisa, para aumentar sua visibilidade, *status* e valor público.

Porém, Pinfield (2002) afirma que os verdadeiros desafios na implementação de repositórios institucionais são as questões culturais e organizacionais. Uma das dificuldades existentes é conseguir levar os autores ao autodepósito sistemático de sua produção intelectual. Uma alternativa é que os responsáveis pelo repositório ou os bibliotecários depositem os itens em nome dos autores, pelo menos no início do desenvolvimento do serviço. Também pode haver a necessidade de formatação ou de conversão de arquivos de modo a poder transferi-los para o servidor do repositório. A mesma alternativa é sugerida como solução. Em ambos os casos, as bibliotecas podem facilitar o trabalho dos autores de contribuírem com suas pesquisas para os repositórios institucionais, porque estão mais bem preparadas para prover a normalização na preparação dos documentos digitais.

Para Pinfield, Gardner e Macoll (2002), a principal meta dos repositórios institucionais é envolver os pesquisadores e usuários. A participação destes agentes é fundamental para a manutenção do repositório. Eles precisam tanto contribuir com conteúdo para o RI, como também usá-lo para consultas e acesso à literatura técnico-científica de outras instituições. Para que um repositório seja valorizado pelos pesquisadores, é necessário que ele seja confiável, que os conteúdos não desapareçam aleatoriamente.

Um repositório institucional nunca é uma iniciativa isolada; ele permite que as bibliotecas possam trabalhar de um modo interdisciplinar, enquanto os canais formais e informais de comunicação continuam a evoluir. Assim, ele muda também o papel passivo das bibliotecas, de receptoras de material publicado para agentes ativos na determinação do gerenciamento do conteúdo relevante próprio (NIXON, 2002).

A capacidade de contar com um espaço onde possa ser depositada a produção científica de uma instituição propicia o ambiente necessário para a discussão entre os pares, o intercâmbio de idéias dentro de uma comunidade de pesquisadores e a revisão constante das versões de um documento. O debate crítico é alimentado de forma aberta pelos participantes dos repositórios de acesso

livre, e não se tem limite para o tamanho dos documentos ou extensão temporal da discussão de qualquer tema. São resolvidos os problemas de distribuição dos canais formais e o acesso pode ser gratuito.

Por serem um veículo de autopublicação por natureza, os repositórios abertos solicitam, apenas, uma promoção institucional para seu uso. As instituições reconhecem os materiais arquivados e os valorizam, estabelecendo as ligações entre a informação que precisam e aquela que desejam disponibilizar abertamente.

Um repositório digital pode ser operado em qualquer organização do setor público. Atualmente existem instituições encarregadas de armazenar e gerenciar material digital, mas são poucas as que asseguram ter a capacidade de armazenamento de longo prazo. Muitas empresas privadas afirmam que provêem acesso e armazenamento, mas o acesso público é restringido, e o proprietário original que gerencia o conteúdo em uma página comercial da internet passa o controle direto do material para terceiros. Nesses casos, segundo Simpson (2004), para manter os direitos e o controle sobre o material digital, deve-se consultar a legislação, assim como especificar se as pesquisas foram financiadas, estipulando que o material continuará sob o controle e propriedade da instituição que o gerou.

Um dos objetivos da publicação científica é salvaguardar a propriedade intelectual (OKUBO, 1997). Os repositórios institucionais lidam com essas tarefas e com o problema da cópia apropriada. Existem, também, alguns tipos de licença sobre a utilização de material intelectual (*Common License/Creative License*), que seguem as normas de direitos autorais. O único obstáculo atual para o acesso livre mediante o uso de repositórios institucionais é a ausência de políticas institucionais que requeiram o autodepósito da sua própria produção científica.

A opção pelo autodepósito temático e/ou institucional é certamente mais confiável. Além disso, à medida que o repositório institucional se consolida, a coleta automática aos repositórios temáticos e centrais poderá auxiliar na segurança dos dados, fornecendo *backups*, redundâncias e “espelho” dos metadados e documentos contidos nos repositórios temáticos.

As diferenças sutis entre esses objetivos são complicadas devido à variedade de materiais publicáveis. Por exemplo, nos repositórios, muitos dados são considerados literatura cinzenta: relatórios, folhetos, guias, informativos, dados orçamentários, memorandos e resultados de pesquisa que não foram publicados (muitos deles de organizações que não têm perfil de editoras). Acrescentem-se à

lista trabalhos artísticos, pedidos de orçamento, mapas, entrevistas de TV e rádio, filmes, registros musicais, fotografias, relatórios de consultoria, desenhos técnicos, pôsteres.

Por causa da preservação digital desse tipo de materiais, os repositórios institucionais se converteram em assunto muito relevante. Com eles, são colocados em funcionamento serviços para os membros das comunidades internas que visam ao gerenciamento e à disseminação de todo o material criado dentro de uma instituição e fora dela. A vantagem dos *softwares* de repositórios está em prover mecanismos para identificar consistentemente materiais que simplifiquem as atividades relacionadas com futuras migrações. O objetivo é possuir um *software* de repositório que possa garantir o gerenciamento das mudanças tecnológicas e a migração do conteúdo digital para novas versões de mídia eletrônica. Sem uma estrutura de repositório, a migração de materiais digitais pode chegar a ser um fator complicador.

2.3.3.4 Repositórios institucionais nas bibliotecas

Os repositórios institucionais começaram a ter importância para a comunidade científica a partir de 2002. A comunidade de pesquisadores da área de preservação digital percebeu os benefícios que esses sistemas poderiam trazer para as instituições acadêmicas. Para Lynch (2003), com os repositórios digitais as instituições acadêmicas podem realmente apoiar a pesquisa da qual elas se beneficiam, deixando seu papel passivo de editoras e participando do desenvolvimento da arquitetura de novas funções dos sistemas de informação. No início de 2003, o movimento dos repositórios institucionais já expressava clara missão: proporcionar às instituições acadêmicas a possibilidade de investir no desenvolvimento de sistemas de repositórios digitais.

Foi nesse momento que responsáveis pelas bibliotecas, pelos arquivos e pelos museus perceberam as possibilidades de integração de acervos criada pela homogeneidade dos documentos digitais com que lidam. Todos os objetos digitais podem ser abordados da mesma forma por qualquer dessas instituições, sem que uma perspectiva exclua as outras.

Na área da ciência da informação, os repositórios institucionais estão sendo considerados como uma das formas de mudar rapidamente a comunicação científica e também como um dos incentivos para que as bibliotecas entrem em um território inexplorado. A prioridade atual é o desenvolvimento de repositórios institucionais, *softwares* e serviços que gerenciam e disseminam materiais digitais de toda instituição. Os bibliotecários lideram esse trabalho de forma inovadora, como participantes de equipes colaborativas que podem incluir tecnologistas de informação, arquivistas, professores e administradores de universidades, assim como funcionários públicos e comunidades de usuários.

Como produtoras de pesquisas científicas, as instituições acadêmicas estão interessadas na captura, disseminação e preservação da produção intelectual de seus próprios membros. Tradicionalmente, as editoras e bibliotecas têm tido o papel complementar de facilitar a publicação e preservação da produção científica. Nas últimas décadas, as mudanças tecnológicas e de mercado têm-se acelerado, motivadas em parte pelo volume crescente da publicação de resultados de pesquisas.

Essa mudança de papéis está afetando a relação simbiótica entre editoras e bibliotecas. A combinação de uma rede quase ubíqua com o aumento crescente dos preços nos modelos tradicionais de publicação prepara o cenário para novas expectativas. Entre elas está o uso de repositórios institucionais para prover aos professores novas formas de criar e preservar objetos de aprendizagem, tais como ilustrações, visualizações, modelos e vídeos.

Para Peters (2002), o papel das instituições de ensino superior e de pesquisa envolvidas em projetos de preservação digital será de detentoras de repositórios digitais, que permitam controle da autoria e a cobertura de elevado número de áreas de conhecimento. Peters cita o exemplo do consórcio de universidades públicas de Ohio, nos Estados Unidos (OhioLINK), que está se desenvolvendo para ser um serviço de repositório digital por meio do “[...] auto depósito da informação de bases de dados de artigos de periódicos e bancos de teses e dissertações de bibliotecas participantes [...]” (PETERS, 2002, p. 415).

Outros autores, como Messerschmitt (2003) e Hitchcock (2003), também asseguram que os repositórios institucionais poderiam ser viabilizados mais facilmente através de sistemas distribuídos de preservação digital. Neles, os repositórios centralizados e mais bem organizados compartilhariam suas

metodologias com os outros repositórios temáticos e institucionais. Para Messerschmitt, o papel das bibliotecas como curadoras da informação digital é fundamental, uma função que elas têm realizado por séculos.

Como a maioria dos repositórios está em instituições acadêmicas, as bibliotecas universitárias começaram também a desenvolvê-los. As bibliotecas estão se unindo às agências de governo, sociedades históricas, museus e outras instituições culturais para estabelecer repositórios comunitários ou temáticos. Nesse intuito, vários repositórios estão sendo desenvolvidos em consórcio, uma vez que nem todas as instituições acadêmicas precisam ou desejam tê-los.

Os repositórios institucionais estendem significativamente o papel das bibliotecas, representando um compromisso sério e de longa duração, com muitos benefícios. Professores e cientistas que começam a usar repositórios institucionais para publicar e preservar seus trabalhos confiam na integridade, conhecimento e competência daqueles que gerenciam esses repositórios.

A experiência acumulada de adoção de padrões e protocolos de comunicação por parte dos profissionais da informação é fundamental na construção de repositórios institucionais confiáveis. Eles podem definir um conjunto mínimo de critérios para o arquivamento de informações científicas produzidas em uma instituição. Esse conjunto de critérios para o desenvolvimento de um repositório de preservação deve estar influenciado pelo modelo OAIS e pelo princípio arquivístico de custódia responsável.

2.3.3.5 *Softwares* para repositórios institucionais

Com o surgimento da Iniciativa dos Arquivos Abertos[35], inicialmente conhecida como UPS (*Universal Preprint Service*), na última década do século XX, diminuíram as dificuldades que os autores enfrentavam para disponibilizar o resultado das suas pesquisas internacionalmente, aumentando a sua visibilidade e a estimativa de impacto na comunidade, o qual é medido pelo grau de usabilidade alcançado (HARNAD, 1999).

[35] Disponível em: www.openarchives.org Acesso em: 23/out/2008

Através dos padrões de interoperabilidade do modelo OAI (metadados *Dublin Core*, protocolo de coleta de metadados OAI-PMH, dentre outros), os autores e os editores interessados em disponibilizar suas publicações abertamente podem ter seus trabalhos compartilhados com outras instituições. Esta é uma das formas utilizadas pelas universidades para apoiar a divulgação dos resultados de pesquisas científicas, criando mecanismos para legitimar e estimular a publicação dos trabalhos produzidos (LYNCH, 2003).

Para Harnad (2003), os repositórios institucionais têm cinco objetivos:

a) realizar o autodepósito da produção científica institucional em forma de *pré-prints*, *pós-prints*, teses, dissertações, coleções digitais, matériais digitais preserváveis, materiais para o ensino e as versões eletrônicas de periódicos e livros;
b) providenciar o gerenciamento da coleção digital institucional;
c) prover material para o ensino à distância;
d) viabilizar a publicação eletrônica;
e) preservar o conteúdo digital institucional.

Na perspectiva da Iniciativa dos Arquivos Abertos, nem todos os repositórios institucionais possuem as mesmas metas, mas eles compartilham objetivos similares. O contexto dos repositórios pode ser caracterizado pelos seguintes atributos:

a) repositórios que armazenam arquivos que nasceram digitais, sem análogos em papel;
b) repositórios que possuem atributos de confiabilidade e disponibilidade relevantes para a comunidade científica;
c) repositórios de acesso contínuo, atributo digital que garante a possibilidade de citar, descobrir, entregar e usar o recurso depois de sua criação e depósito no repositório.

Esses atributos devem ser garantidos permanentemente para prevenir as falhas dos formatos e controlar os efeitos das mudanças tecnológicas. Também, os processos de preservação digital devem estar apoiados em políticas bem definidas, na sua organização e nas estratégias adotadas.

Como mencionado anteriormente, o modelo OAIS fornece as especificações de um repositório e estabelece responsabilidades que uma organização deve distribuir para operar como um arquivo de acesso livre. O modelo

de referência define um sistema de informação para arquivamento aberto como aquele composto por uma organização de pessoas e sistemas que aceitam a responsabilidade de preservar informação e sua disponibilização para uma comunidade específica. A aplicação dos princípios do modelo OAIS, e particularmente, a implementação de um arquivo de acesso livre em concordância com os modelos de funcionalidade e estrutura da informação do OAIS é o pré-requisito chave para estabelecer repositórios confiáveis e garantir a preservação de longo prazo dos seus atributos digitais.

Para garantir a confiança nos repositórios digitais de acesso livre por parte dos autores e dos usuários, eles devem manter algumas propriedades de preservação digital mínimas:

a) autenticidade: a certeza de que um componente digital foi criado pela pessoa que afirma tê-lo feito; a autenticidade permite ter a certeza de que o criador do objeto digital não pode negar que foi ele quem o criou. As assinaturas digitais e as marcas d'água digitais são técnicas que garantem a autenticidade dos objetos digitais;

b) integridade: a habilidade de manter os dados completos e corretos, prevenindo mudanças acidentais ou maliciosas (corrupção dos dados). Entregando e salvando um *bit/byte checksum*, como o MD5 faz, constitui uma técnica básica para detectar se qualquer modificação produziu algum efeito nos objetos digitais depois de ter sido inserido no repositório;

c) confiança e disponibilidade: a confiabilidade está relacionada à habilidade dos componentes de *hardware* e *software* funcionarem de acordo com suas especificações sem erros ou defeitos. Disponibilidade é a porcentagem do tempo que o sistema está regularmente em funcionamento, em relação com o tempo total que ele deve operar. Algumas técnicas usadas para garantir altas porcentagens de confiabilidade e disponibilidade são *backups*, *softwares* de antivírus, *firewalls*, *operating system patches*, atualizações de aplicações de *software*, componentes de *hardware* de redundância e tolerância de falhas;

d) capacidade de reuso: habilidade de acessar um recurso digital pelo tempo que a instituição decida manter o repositório. Os objetos digitais

> científicos e acadêmicos que possuem valor por longo período de tempo devem ser recuperados apropriadamente e reusados ao menos por um período longo de tempo (uso de identificadores permanentes e mantendo formatos e mídias). A preservação digital deve incluir todos os componentes de dados da infra-estrutura dos repositórios digitais de acesso livre, não apenas os objetos digitais, mas também seus metadados e identificadores.

Com toda essa variedade de materiais e metas, é possível entender o porquê do aparecimento de sistemas de *software* tão diversificados. Mas não todas essas diferenças surgem por causa, apenas, do tipo de conteúdos que cada repositório coleta. Os sistemas existentes aplicam-se à ampla maioria dos contextos institucionais, permitindo uma variedade de funcionalidades e recursos para gestão de documentos eletrônicos.

A implementação dos RI pode ser realizada de forma simples: uma estrutura hierarquizada, acesso via Web e metadados coletados por meio do protocolo OAI-PMH, permitindo assim que os usuários, ao optar por qualquer mecanismo de busca da Iniciativa dos Arquivos Abertos, possam encontrar e recuperar o conteúdo do repositório.

A maior diferença entre os programas para repositórios institucionais está na origem do *software*, se ele é gratuito ou comercial. Os *softwares* comerciais não estão abertos para modificações locais na maioria dos casos. Nos *softwares* abertos, os códigos de programação estão disponíveis para que os usuários possam conhecê-los e modifica-los. Na maioria dos casos, os *softwares* não vêm com uma licença. O sistema operacional *Linux* é o exemplo mais conhecido.

Os espaços para o depósito e submissão de documentos digitais com informações científicas estão sendo desenvolvidos em alguns programas com características parecidas. Alguns pacotes de *software* livre para repositórios institucionais têm ganhado destaque, entre eles:

- EPrints, o mais antigo, possui a base de instalação mas variada. Disponibilizado pelo Joint Information Systems Committee na Inglaterra e pela National Science Foundation dos Estados Unidos, este sistema focaliza a produção científica tradicional em forma de pré-prints e pós-prints. A publicação no EPrints segue o modelo tradicional da publicação impressa. Mesmo sendo um *software* que provê

longevidade e acompanha o desenvolvimento na área da preservação digital, a primeira motivação do EPrints não é a preservação de longo prazo.

- Greenstone é um *software* que permite a construção e distribuição de coleções digitais de bibliotecas. Desenvolvido pelo New Zealand Digital Library Project, o Greenstone está sendo distribuído em colaboração com a Unesco e a Human Info NGO.
- i-Tor do Netherlands Institute for Scientific Information Services implementa um tipo de repositório de dados independente. O conteúdo e a interface de usuário funcionam como partes autônomas porque o i-Tor publica os dados de grande variedade de sistemas de bases de dados relacionais, tipos de arquivos de sistemas, e de *sites* na *web*.
- CDS Invenio (antigo *CDSWare*) cobre todos os aspectos de uma biblioteca digital e é usado entre outros repositórios no mundo pelo repositório da CERN, o maior entre todos os registrados no ROAR e o terceiro entre os repositórios institucionais (não centrais que coletam de outros repositórios institucionais). O repositório da CERN também, pelo seu elevado índice de textos completos é o que conta com mais aceitação institucional. Segundo Harnad (2003), o repositório da CERN tem a melhor política de repositórios institucionais no mundo, com o mandato de auto-arquivamento institucional mais compreensivo, sistemático e com sucesso[36]. Mas o CERN não promove a adoção de sua política institucional de auto-arquivamento, quando repassa o *software* CDS *Invenio*.
- Fedora™ (*Flexible Extensible Digital Object Repository Architecture*) Desenvolvido pela University of Virginia Library e o Cornell University' Digital Library Group, o Fedora é um *software* livre que focaliza a implementação e gerenciamento de bibliotecas digitais (STAPLES; WAYLAND; PAYETTE, 2003). Ele pode incluir serviços de preservação, pois segue o modelo OAIS nas suas principais funções de entrada e saída de dados. A empresa VTLS usa o *Fedora* como

[36] Disponível em: http://www.eprints.org/signup/fulllist.php Acesso em 12/0ut/2008

base em alguns dos seus produtos comerciais para sistemas de bibliotecas.

- DIAS (*Digital Information Archive System*) foi desenvolvido originalmente por uma empresa para a Biblioteca Nacional da Holanda, com o objetivo de lidar com o depósito de documentos digitais e informação multimídia. A Koninklijke Bibliotheek usou como base os resultados obtidos em vários projetos nos últimos anos, como no sistema *Deposit of Netherlands Electronic Publications* (DNEP) (STEENBAKKERS, 2000); o seu *design* observa os requisitos de um sistema de depósito, suporta a inserção automatizada e manual de materiais, que podem ser acessados via *web*, e é um sistema baseado no gerenciador de conteúdo da IBM (IBM, 2003).
- OCLC Digital Archive está baseado no modelo OAIS e proporciona acesso de longo prazo, armazenamento e preservação de objetos digitais; usa a estrutura de metadados do METS (*Metadata Encoding and Transmission Standard*) para permitir a interoperabilidade com outros arquivos e está conectado com o sistema de catalogação *OCLC's Conexion*.

A vantagem desses *softwares* de repositórios está em disponibilizar mecanismos para identificar consistentemente materiais que simplifiquem as atividades relacionadas com futuras migrações. O objetivo das instituições responsáveis pela guarda de informações digitais é possuir um *software* de repositório que realize as migrações de forma estável, apesar das mudanças tecnológicas.

2.3.3.6 Repositórios DSpace

Dentre os *softwares* citados, a ferramenta que tem sido implementada pelo maior número de instituições e que desde sua criação deu prioridade à preservação digital é o DSpace. Orientado para ser um programa usado em repositório que captura a produção intelectual das instituições de pesquisa, diferentemente dos

outros *softwares*, o DSpace estende-se para uma variedade maior de tipos de materiais e possui mais opções de preservação de objetos digitais.

O DSpace é um *software* livre desenvolvido pelo Massachusetts Institute of Technology (MIT) Libraries e pelos Laboratórios Hewlett-Packard para criação de repositórios institucionais e multidisciplinares para bibliotecas, arquivos e centros de pesquisa. Sua estrutura provê um modelo de informação organizacional baseado em "comunidades" e coleções, o qual pode ser configurado de modo a refletir todo o conjunto de unidades administrativas de uma instituição. Permite a configuração do processo editorial nos moldes dos periódicos tradicionais, incluindo a possibilidade de revisão pelos pares. Suporta os mais variados tipos de formatos de arquivos digitais, incluindo textos, som e imagem (Figura 15).

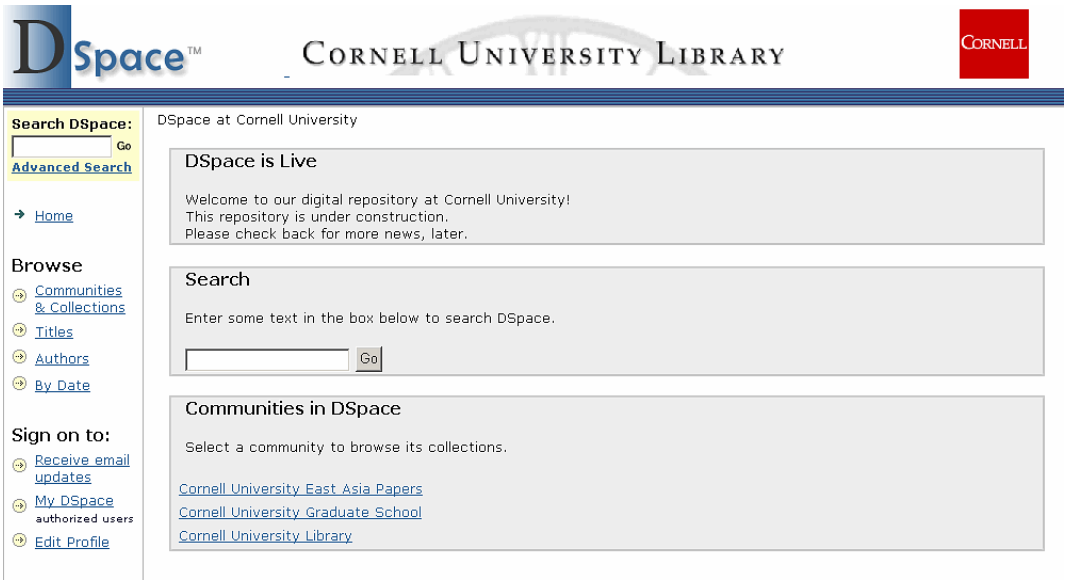


Figura 15 - Primeira interface do *DSpace* da *Cornell University*

O surgimento do DSpace em 2003 está relacionado à iniciativa da *Open Courseware* (OCW) da MIT, com o objetivo "prover acesso livre, recuperável e coerente aos materiais de ensino do MIT" (DSPACE, 2002). Eles são parte da infra-estrutura que o Council on Educational Technology (MITCET) formou em 2000. Também, o projeto estabelece conexão com outras iniciativas, como a que está sendo realizada na Harvard University, chamada *Digital Repository Service* (DRS). O DSpace representou a entrada do MIT na comunidade de desenvolvimento de *software* livre, com um sistema completo e disponível gratuitamente. O interesse foi

expressivo desde as etapas pré-operacionais do projeto, sendo considerado pelos pesquisadores como um modelo na criação da estrutura de outros sistemas.

O DSpace possui uma natureza operacional específica de preservar os objetos digitais que são de interesse da comunidade científica. Criado inicialmente para a implementação de arquivos institucionais, tem sido usado para o compartilhamento da produção científica nas universidades. O *software* é resultado da implementação de padrões já aceitos internacionalmente, como o *Open Archive Initiative*, o *Open Archival Information System Reference Model* e o Dublin Core. Atualmente está sendo utilizado para promover uma aliança entre repositórios federativos e instituições.

A missão do DSpace é oferecer um armazenamento estável a longo prazo para todos os materiais produzidos pelos professores e pesquisadores do MIT. A preservação de longo prazo seria em ampla variedade de formatos, incluindo texto, áudio, vídeo, imagens, bases de dados, entre outros; assim como possibilitar o acesso remoto a todos os materiais através de uma interface coerente. As funcionalidades de preservação estiveram presentes desde as primeiras etapas do *design* da arquitetura do repositório, envolvendo aspectos sociais (CHUDNOV, 2001).

Para os desenvolvedores do DSpace, além do seu uso como repositório, o *software* foi implementado para ajudar na pesquisa dos sistemas de gerenciamento de conteúdo que contemplam a preservação, o arquivamento e a distribuição de materiais digitais. A plataforma utiliza o modelo OAIS, seu vocabulário para descrever sistemas de arquivamento de mídias e a verificação dos planos operacionais e funcionais para determinado repositório. A influência do OAIS no *design* da arquitetura do DSpace faz dele um exemplo de repositório com funcionalidades de arquivamento digital.

Os repositórios digitais podem abordar uma preservação a curto ou longo prazo (CROW, 2002), promovendo o ambiente necessário para a disseminação e processos de reprodução, assim como criar o compromisso institucional com a preservação. No DSpace, a preservação ocorre no nível dos *bytes* onde a ordem de cada item é armazenada, assim como sua habilidade de preservação em alto grau de representação, que pode facilitar na identificação e servir como evidência das partes de um registro. Muitos repositórios enfrentam dificuldades para chegar a esse

grau de representação. Ainda são poucos os repositórios adequadamente equipados para lidar com metadados de preservação.

Os níveis de preservação no DSpace são fundamentais devido, principalmente, ao uso de metadados na sua concepção. No DSpace encontrou-se três tipos básicos de metadados – descritivos, administrativos e estruturais. A estrutura dos metadados segue o modelo de metadados METS. Os metadados descritivos estão baseados no *Library Application Profile for Elements and Qualifiers*, mas qualquer instituição que faça opção pelo DSpace pode mudar o conjunto de metadados usado.

No nível das comunidades e das coleções, usam-se metadados descritivos simples derivados de um subconjunto do Dublin Core. Os metadados estruturais contêm informações sobre como apresentar um item ou *bitstream*, assim como informações sobre as relações entre os itens. Nos metadados administrativos estão incluídos os metadados de preservação, de origem e a informação sobre a política de autorizações. Muitos deles são organizados segundo o esquema de relacionamento DBMS. As informações sobre Origem (*Provenance*) são mantidas em Dublin Core, mas em uma descrição simplificada. Nesta área de metadados administrativos espera-se que o dicionário de metadados da OCLC/RLG PREMIS possa ser utilizado na identificação dos elementos para metadados de preservação.

O DSpace faz uso da terminologia de comunidades para definir os espaços dos conteúdos onde os usuários podem localizar a informação. Em cada comunidade os usuários encontram coleções, e nelas estão os itens ou "[...] agrupamentos de conteúdos e metadados que permitem seu arquivamento em pequenas unidades [...]" (DSpace, 2002).

Os itens no DSpace são os registros ou arquivos, organizados em *bundles* ou agrupamento de *bitstreams*, como aqueles contidos num arquivo *html* e suas imagens associadas, tendo no seu nível mais elementar a organização dos itens por *bitstream*, identificados pelo seu formato de *bitstream*.

Os itens constituem conjuntos de informações que descrevem o conhecimento sobre o formato e a codificação do *bitstream*, incluindo uma versão MIME e o nome do tipo de formato (ex. *Adobe PDF*). As informações também contêm especificações do formato e código fonte para manipulação do formato. A divisão hierárquica dos *bitstreams* e seus agrupamentos na forma de metadados adicionados correspondem aproximadamente ao modelo de informação do OAIS.

A preservação desse tipo de conteúdo é gerenciado na arquitetura do DSpace pelo *Bitstream Storage Manager*. O administrador do armazenamento dos *bitstream* é uma das mais importantes funcionalidades de preservação na arquitetura do DSpace, pois nela são armazenadas as informações sobre os arquivos submetidos ao nível de *bitstream*, permitindo uma capacidade limitada de transação para os *bitstreams*.

A complexidade inicia no momento em que o gerenciador do armazenamento e seus administradores começam a identificar o nível do serviço de preservação para cada um dos itens e, dessa forma, determinar o grau de suporte para cada tipo de *bitstream*.

Os administradores do DSpace dividem os formatos de *bitstream* em duas categorias: conhecido e desconhecido. Os tipos de *bitstream* conhecidos são divididos também em formatos suportados e os que o sistema não suporta (TANSLEY *et al.*, 2003). Os conhecidos são aqueles aos quais os administradores do DSpace nomearam o formato do arquivo, sua versão, *mimetype* e outras informações relevantes (ex.: sua relação com outros formatos) dentro do registro dos formatos do *bitstream* do sistema. Além disso, um processo foi ativado para identificar automaticamente as futuras submissões, usando o formato já nomeado.

Ser suportado significa que os administradores procuraram e armazenaram especificações documentais do formato suficientes no DSpace ou em formato análogo numa biblioteca confiável.

O conhecimento dos administradores dos diferentes formatos de *bitstream* e seu suporte possibilita novo nível de serviços de preservação. O mais alto nível é denominado Suportado Nível 2. Esse serviço é estabelecido quando o *bitsream* é mantido e recuperado sob pedido no futuro. Acrescente-se a isso que a instituição confia na sua habilidade de usar e entender que o material depositado poderá ser razoavelmente preservado e existe um comprometimento com essa finalidade, da forma mais apropriada, segundo o contexto que envolve cada material. As técnicas de preservação a serem aplicadas incluem estratégias como emulação, migração, serviços de transformação de formatos.

Todas as funcionalidades do Gerenciador de armazenamento de *bitstream* (*Bitstream Storage Manager*) podem ser contextualizadas dentro da concepção de planejamento da preservação (*Preservation Planning*) no modelo OAIS. Da mesma forma, alguns componentes também podem participar da Administração e

Armazenamento de Arquivamento (*Administration and Archival Storage*). Isso mostra como o modelo OAIS pode ser aplicado em todas as funções de um repositório. Os desenvolvedores do DSpace tomaram como exemplo o desenho do repositório *Harvard DRS* na arquitetura do DSpace, especialmente os componentes relacionados com a identificação e gerenciamento dos níveis de preservação da informação digital que ele suporta.

Uma peculiaridade que faz do DSpace um repositório orientado para a preservação digital é o fato de que, além de utilizar tecnologias como as mencionadas, ele opera em um ambiente computacional de ferramentas de código aberto (*software* livre). Isso permite assegurar a viabilidade da estrutura de longo prazo, apoiada na comunidade de usuários e desenvolvedores que mantêm as informações sobre o DSpace sob uma licença BSD.

O DSpace é o exemplo de um repositório confiável que garante a manutenção de formatos de arquivos gerenciando-os, identificando-os e aplicando processos de preservação que o sistema é capaz de suportar. As instituições que adotam o DSpace conseguem, além de identificar quais são os arquivos ou coleções que precisam de maior proteção, realizar um planejamento sobre a manutenção de cópias de *software* e *hardware* e sua documentação necessária para identificar os tipos de arquivo indispensáveis para o uso dos objetos digitais no futuro. Essas ações servem para que os usuários do sistema possam ter maior grau de confiança de que seus documentos receberão os cuidados necessários à sua preservação.

2.3.3.7 Repositórios e identificadores persistentes

Outra área relacionada com a preservação digital é a técnica de nomeação persistente dos objetos digitais ou incorporação na arquitetura de um sistema de URN (*Uniform Resource Name*). Identificar e localizar consistentemente conteúdo são cruciais para os *softwares* de repositórios. Existem vários URN, tais como o Handle System®, o DOI® (*Digital Object Identifier*) da CrossRef e o PURL (*Persistent Uniform Resource Locators*) da OCLC. Para facilitar o acesso

permanente, cada objeto deve ter um identificador que transcende o próprio *software*.[37]

Os identificadores permanentes estão relacionados diretamente com os padrões de preservação digital, sendo um componente importante da infra-estrutura dos repositórios digitais, para ajudar a garantir que os objetos digitais possam ser gerenciados e localizados no futuro. Os identificadores permanentes devem permanecer válidos, mesmo se o conteúdo for migrado para um novo sistema ou se a administração responsável pelo repositório institucional mudar.

O PURL foi desenvolvido pela OCLC baseado no padrão URL (*Uniform Resource Locator*) do protocolo HTTP. O funcionamento do PURL está estruturado na conexão de URL que apontam para um serviço de resolução (*service resolver)* mantido para redirecionar um endereço ou *link* de páginas *web* que não esteja funcionando para o endereço ativo. Três exemplos de instituições que usam PURL para a preservação do acesso às informações nas suas páginas *web* são: o Digital Archive da OCLC, a US. Government Printing Office e o US Department of Energy's Office of Scientific and Technical Information.

O Handle System®, criado pela Corporation for National Research Initiatives (CNRI), como um *software* de URN permite que qualquer item possa ser encontrado no decorrer do tempo, apesar de haver mudado de local físico. Os protocolos do sistema possibilitam que de forma distribuída um servidor remoto guarde os "*handles*" ou identificadores permanentes dos recursos de informação e reconheça a informação necessária para localizar e acessar tais recursos. Essa informação pode mudar quando necessário, para ser capaz de descrever o estado atual do objeto digital, sem mudar o prefixo ou "*handle*". O sistema *Handle* também usa um serviço de resolução de identificação, mas no *Handle*, o serviço permite que um único prefixo resolva uma ou mais URL com versões, formatos e locais diferentes para o mesmo documento.

No caso do DSpace, cada repositório obtém um *handle* ou prefixo do do CNRI que permite a identificação persistente do Handle System® para todos os documentos depositados, como um componente do repositório institucional e

[37] PURL: Online Computer Library Center Inc. http://purl.oclc.org
URN: Internet Engineering Task Force. http://www.ietf.org/html.charters/urn-charte.html
DOI: International DOI Foundation. http://www.doi.org
Handle System: Corporation for National Research Initiatives. http://www.handle.net/index.html

convertendo-o globalmente em único. Um *handle* pode ser designado a comunidades, coleções ou itens, menos aos *bitstreams*, porque ele pode mudar para dar suporte a novas atividades de preservação. Instituições como a Koninklijke Bibliotheek na Holanda e a Defense Technical Information Center nos Estados Unidos têm incluído os identificadores persistentes do Handle System® (STEENBAKKERS, 2002).

Algumas organizações internacionais já trabalham no sentido de criar seus próprios sistemas de identificadores permanentes. Um dos mais reconhecidos é o DOI® da International DOI Foundation, que em 1998 implementou o Handle System® especificamente para a gestão da propriedade intelectual, gerenciando identificadores permanentes de *links* referenciados entre publicações, com seu próprio registro de nomes de autoridades. Em 2003, a DOI Foundation começou a designar seus identificadores persistentes a documentos com dados científicos, e tem sido usado para referenciar e citar documentos originais, aumentando com isso a possibilidade de que os documentos sejam encontrados e reusados.

Entre os Handles e PURLS alternativos que vêm surgindo, podem ser mencionados os seguintes:

- *Archival Resource Key* (ARK) criado pela Califórnia Digital Library é usado em todos seus objetos digitais e serviços (KUNZE, 2003); permite o acesso a longo prazo a informações contidas em qualquer tipo de suporte, pois contém partes que são flexíveis para suportar redirecionamento de URL, associando permanentemente um “nome” a um objeto específico.[38]
- URN baseados no ISSN e no ISBN, como no projeto DiVA da Uppsala University na Suécia (MUELLER *et al.*, 2003) e nos projetos EPICUR e CARMEN da Alemanha.
- *Local Identifier Scheme* da National Library of Australia para todas as publicações no projeto *PANDORA Archive*, permitindo a criação automática de identificadores permanentes incorporada no software PANDAS, fazendo com que o esquema local possa ser convertido em um sistema global.

[38] Disponível em: http://www.cdlib.org/inside/diglib/ark/ Acesso em: 16/out/2008

- N2T (*Name-to-Thing*) é um serviço de resolução de identificadores persistentes e um consórcio de organizações que lidam com a memória cultural (http://n2t.info/).

2.3.3.8 Repositório uniforme do INPE

No Brasil, a biblioteca do Instituto de Pesquisas Espaciais (INPE) começou a desenvolver em 1995 o *Uniform Repository for a Library* (URLib)[39] para resolver o problema da integração de documentos produzidos por pessoas utilizando sistemas de arquivos independentes (BANON; BANON, 2005). Para gerenciar os repositórios uniformes que alimentam o acervo da biblioteca, foi desenvolvido o *URLibService*. A biblioteca contém documentos eletrônicos formados por coleções de arquivos e diretórios depositados em vários repositórios uniformes ou acervos locais utilizando o Apache como servidor *web*.

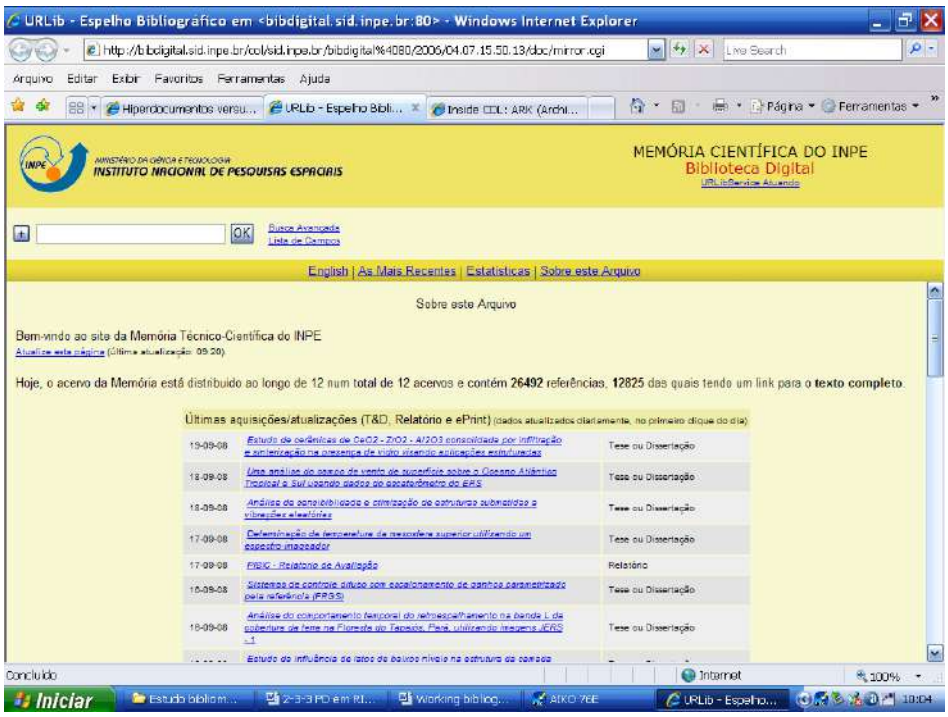


Figura 16 – Interface da Biblioteca Digital Memória do INPE[40]

[39] Repositório Uniforme para uma Biblioteca
[40] Fonte: INPE. Disponível em: http://bibdigital.sid.inpe.br Acesso em: 15/dez/2008

Assim como o ARK, no sistema de Identificação UR*Lib*Service a geração dos identificadores pode ser feita pelos próprios provedores de dados, se precisar de cadastramento junto ao resolvedor de identificação, como no caso do Handle System®, estabelecendo-se um vínculo automático entre documentos depositados em dois provedores distintos. Também a vantagem do *URLibService* é de que a diferencia do DOI®, pois consegue resolver "[...] a distinção entre um documento e suas cópias, oferecendo para o usuário final a garantia que o documento acessado é sempre o mesmo, seja ele uma cópia ou não [...]." (BANON, 2006).

Com o *URLibService*, o Inpe desenvolveu o primeiro serviço de identificação global de nome no Brasil, sendo a primeira solução nacional para os problemas de *links* e citações permanentes. Assim como, também, é o primeiro sistema de autodepósito nacional que respeita os direitos autorais e que certifica a originalidade dos trabalhos. O programa é livre para instituições públicas e associações sem fins lucrativos (BANON, 2006).

Em 2007, as instituições brasileiras como o CNPq, a Bireme e o Ibict começaram a usar em alguns de seus sistema de informação o redirecionamento que o identificador DOI® estabelece, para citar artigos ou partes de artigos dos pesquisadores (Sistema LATTES) e no endereço dos artigos nas revistas na base de dados SciELO e no sistema OJS/SEER versão 2.1.

2.4 GERENCIAMENTO DA PRESERVAÇÃO DIGITAL

Como parte da construção cultural da realidade, a informação reflete as visões do mundo de uma sociedade em determinado tempo. No século XX, a informação técnico-científica tornou-se relevante para o progresso econômico das nações. O desenvolvimento científico e tecnológico trouxe consigo maior oferta de recursos de informação para a população mundial. As novas tecnologias de informação colocaram em evidência o papel dos meios e técnicas digitais de gerenciamento da informação digital. O material digital, seja aquele criado em computador ou digitalizado, está presente na maioria dos serviços de informação (WEBB, 2000).

Em vários países, há urgência na definição das políticas, obrigações e metodologias mais apropriadas para a administração dos recursos digitais. Internacionalmente, iniciativas dos setores público e privado se multiplicam, tentando estabelecer metodologias e estratégias que incluem técnicas de preservação que possam responder às necessidades de garantir a longevidade dessas informações (Apêndice C).

A compreensão da preservação digital por parte das instituições responsáveis por acervos culturais e históricos tem focalizado os obstáculos técnicos e a habilidade de provar a validade das soluções tecnológicas adotadas. Os modelos de sistemas da preservação digital dentro do contexto das instituições produtoras de conhecimento em C&T e das bibliotecas universitárias têm sido analisados por vários autores. Alguns temas conexos, como a produção científica e aspectos específicos sobre preservação, como *hardware e software* envolvidos, a seleção do que será preservado, o custo da preservação têm sido tratados, assim como existe uma ênfase na necessidade de planejamento do que será selecionado para se preservar.

Desde a primeira proposta de um esquema conceitual que abordou o gerenciamento da preservação digital, feita por Daniel Greenstein em 1998, detectou-se que os requisitos da preservação digital influenciam a maneira como os recursos digitais estão sendo estruturados, documentados, armazenados, validados e as condições como eles podem ser acessados. Greenstein (1998) descreveu o gerenciamento do material digital em sete módulos interligados, nos quais as boas práticas dependem das decisões dos desenvolvedores das coleções sobre aspectos como tipo de conteúdo e formato (Quadro 6).

Módulo	Submódulo	Sub-submódulo
Dados de Seleção e Avaliação		
Dados de Criação		
Dados de Gerenciamento	Dados Estruturais	
	Dados de Documentação	
	Dados de Armazenamento	
	Dados de Validação	Dados de Valorização
		Dados de Cópia
		Restauração da Mídia
Dados de Divulgação		
Dados de Uso		
Dados de Preservação		
Gerenciamento de Direitos		

Quadro 6 – Gerenciamento dos dados preservação digital
Fonte: versão traduzida e adaptada de Greenstein, (1998)

O modelo de Greenstein trata dos tipos de decisões que devem ser tomadas no gerenciamento de recursos digitais. Algumas delas independem do controle dos responsáveis pelas coleções. São informações gerenciais sobre a estrutura do dado (como e quantas vezes ele foi formatado e codificado), informações sobre a sua origem, suporte e locais de armazenamento, procedimentos adotados para garantir a integridade do dado (completude, funcionalidade e consistência), permissões de uso e direitos autorais. Divididas em módulos desde o momento da sua criação do recurso digital, passam pela avaliação da usabilidade e relevância do conteúdo.

O modelo de Greenstein apontou a necessidade de implementar uma estrutura para o gerenciamento de objetos digitais que possa ser usada para descrever as necessidades dos responsáveis por esses registros. Sua proposta foi uma das primeiras a ressaltar a área da preservação digital como aquela que requer maior atenção, quando se levar em consideração os custos com a manutenção dos recursos e a adoção prática por parte dos administradores (HENDLEY, 1998).

Tanto para Greenstein como para Hendley, um sistema de preservação digital podia ser visto como uma combinação de:

a) capacidade de armazenamento;
b) funções de inserção;

c) coleções de metadados e gerenciamento;
d) estratégias de preservação e disseminação de conteúdos arquivados em sistemas administrados de forma centralizada;
e) outro tipo de estrutura organizacional na qual os componentes do processo de preservação estão separados em serviços distribuídos entre as várias instituições participantes, cada uma delas especializada em um segmento do processo.

Em 2004, Lavoie e Dempsey resumiram a quatro níveis a concepção de um sistema de preservação digital. No primeiro, estariam o *hardware*, *software* e a infra-estrutura que dá apoio ao armazenamento e distribuição do conteúdo digital. No segundo nível, estariam incluídos os serviços mais especializados para o gerenciamento do conteúdo arquivado no sistema, incluindo a criação de metadados, validação da autenticação e integridade dos materiais.

A criação de indicadores, por sua vez, é implementada no terceiro nível, incluindo o monitoramento do ambiente da preservação, para acompanhar o impacto das mudanças na habilidade de acessar e usar esses conteúdos, assim como facilitar os processos de migração e emulação. O último nível abrangeria os serviços relacionados com a busca, navegação, solicitações de aceso, permissões e disseminação seletiva.

Essa concepção já estava influenciada pela percepção do funcionamento de um sistema de preservação baseado em repositórios digitais segundo o modelo OAIS. Com ele ficaram evidentes os aspectos mais importantes do *workflow* da informação a ser preservada: a entrada, o acesso, a administração, gerenciamento de dados, o planejamento da preservação e o armazenamento com fins arquivísticos, incluindo o armazenamento da mídia e o gerenciamento de *software*.

Os estudos sobre os sistemas de preservação digital foram surgindo na última década focando no desenvolvimento e propriedade dos *softwares* para repositórios digitais e no fluxo interno da gestão da informação dos objetos digitais.

Um estudo que pode servir como exemplo dos trabalhos sobre a identificação das etapas de gerenciamento da preservação digital foi o realizado por Anderson e Pinfield (2004), no qual foram identificadas sete etapas no gerenciamento do ciclo de vida dos objetos digitais (Figura 17).

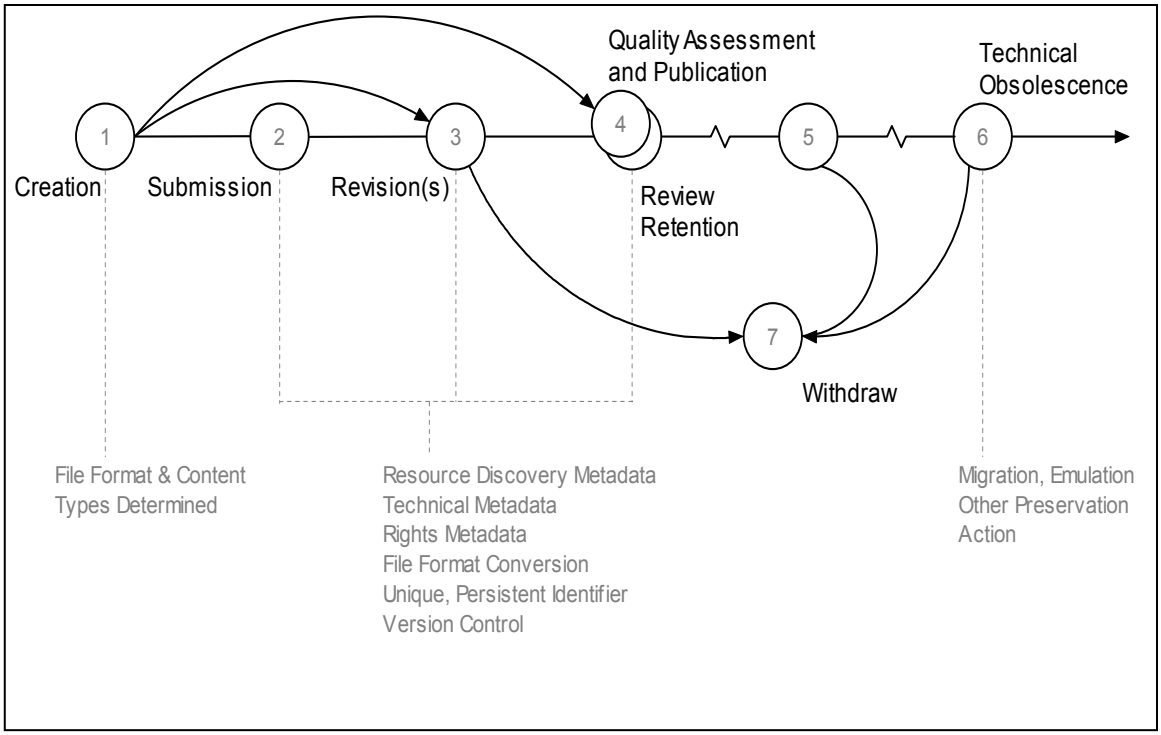


Figura 17 – Ciclo de vida dos objetos digitais
Fonte: Anderson e Pinfield (2004)

A primeira é a da criação do documento, quando se escolhe o formato do arquivo e se determina o tipo de conteúdo; na segunda etapa são realizados os processos de submissão; na terceira são feitas as revisões e na quarta a retenção para revisão. Nessas três etapas (submissão, revisão e retenção) são inseridos os metadados de descoberta, técnicos e de direitos autorais, assim como também o tipo de formato para conversão, o identificador único persistente e a versão para controle. No passo quatro podem ser recebidos diretamente os trabalhos criados já no passo um que marca a atribuição de qualidade para sua publicação. O passo cinco pode significar a decisão para a retirada do documento, nas mesmas circunstâncias dos passos três e sete. Já no passo seis, decide-se a ação a ser realizada com o documento que enfrenta um processo de obsolescência técnica (migração, emulação e outras estratégias).

Mas existe ainda na comunidade da preservação digital a necessidade de ferramentas acessíveis, tecnologia e treinamento no uso dos sistemas que gerenciem os objetos digitais. (LUPOVICI; MASANÈS, 2000; McGOVERN, 2007; ROTHENBERG, 2000). É também aceito o pressuposto de que não existe uma

única mídia de armazenamento permanente e que é necessário o desenho de sistemas que gerenciem as mudanças inevitáveis de um sistema para outro.

Para a Unesco (2007), é mais conveniente que se invista na construção de sistemas sustentáveis do que em suportes supostamente permanentes. No seu relatório de 2007 sobre *softwares* livres para sistemas de arquivamento, a organização conclui que é necessário um sistema simples e sustentável que proporcione estratégias para gerenciar e identificar as funções de preservação.

Especificamente sobre a integração de ferramentas ao *workflow* das instituições, estão surgindo estudos sobre como o *workflow* da preservação digital, que é modular, é capaz de integrar de forma fácil e personalizada as novas ferramentas de preservação digital. Há avanços na área de extração automática de metadados no momento da submissão, mas existe a necessidade de investigar as ferramentas e os tipos de *workflow* que podem ser compartilhados entre organizações.

Uma das mais recentes iniciativas que tentam acabar com a barreira entre as infra-estruturas organizacional e tecnológica é o projeto europeu PLANETS[41], coordenado pela British Library, que promete um tipo de planejamento tecnológico e ferramentas que refletem as políticas organizacionais. O foco do projeto é o processo chamado de planejamento da preservação, definido no modelo OAIS, no qual a preservação digital está centrada nas decisões certas no tempo certo para garantir que a informação de valor não seja perdida. Com a mesma finalidade surgiu o projeto PLEDGE, coordenado pelo Massachusetts Institute of Technology, que promete um protótipo de mecanismo automatizado de políticas de preservação digital[42].

Aguarda-se que, com mais instituições desenvolvendo repositórios digitais confiáveis, seja possível o compartilhamento dos *workflows* e a diminuição dos custos. Existe um esforço em desenvolvimento do padrão ISO, que vai integrar várias iniciativas e requisitos relevantes das áreas de tecnologias da informação e segurança.

---

[41] Disponível em: http://www.planets-project.eu/ Acessado em: 13/out/2008
[42] Disponível em: http://pledge.mit.edu/index.php/Main_Page Acessado em: 13/out/2008

2.4.1 Infra-estrutura tecnológica

A infra-estrutura tecnológica necessária refere-se à combinação de estratégias (procedimentos, protocolos, documentação, redes, medidas de segurança, *workflow*), o pessoal (habilidades técnicas e arquivísticas), os equipamentos (*hardware* e *software*, mídia de armazenamento) e outros meios para alcançar o objetivo de preservar os objetos digitais.

Para Kenneth Thibodeu (2002), diretor da *Electronic Records Archives Program* da National Archives and Records Administration as abordagens tecnológicas da preservação de objetos digitais estão direcionadas ao acesso permanente. Ele afirma que, em contraste com tecnologias como a migração ou "arqueologia digital," surgirão cada vez mais tecnologias que focalizam a documentação necessária dos arquivos digitais para possibilitar sua reprodução no futuro.

Foi na década de 1990 que começaram os estudos sobre a permanência dos objetos digitais, *RLG DigiNews* foi o primeiro periódico científico a publicar artigos dedicados ao tema preservação digital, chegando a ser um dos veículos de comunicação mais consultados pela comunidade de pesquisadores da área. Para a comunidade preocupada com a preservação digital, é considerado um momento significativo a publicação em 1996 do relatório da RLG e a CPA (McGOVERN, 2007).

Nos anos que se seguiram à publicação do *Preserving Digital Information* feita pelo CLIR e o *Research Libraries Group* em 1996, o crescimento considerável dos recursos de informação digital em conteúdo *web*, livros eletrônicos, periódicos *on-line*, mídia digital e material didático tem dificultado a antecipação dos desafios inerentes à sua preservação (SMITH, 2005).

Os progressos alcançados desde o relatório publicado pelo *Task Force on Digital Archiving* multiplicaram-se:

a) consolidou-se o modelo de referência para sistemas de preservação digital (OAIS) também chamado em português de Sistema Aberto de Arquivamento de Informação (SAAI),

b) estabeleceu-se o conjunto de atributos para repositórios digitais confiáveis (TDR), que delimita o contexto organizacional da preservação digital,
c) publicaram-se dicionários de dados para metadados de preservação (PREMIS) e,
d) surgiram os repositórios institucionais facilitados pelo aparecimento de novos *softwares*, como o *Dspace*.

Eles têm favorecido o desenvolvimento de estratégias de preservação mais específicas, como o PRONOM para tipos de conteúdo digital. Juntou-se a eles a norma ISO 19005-1, publicada no dia 13 de setembro de 2005, que especifica o formato *Portable Document Format* - PDF16 ou PDF/A como padrão universal para arquivamento de documentos eletrônicos em longo prazo.

Com relação aos *softwares* para repositórios digitais de *software* livre, alguns já têm se consolidado na comunidade científica e suas diferenças são claras. Os mais usados (EPrints e DSpace) explicam sua orientação da seguinte maneira:

a) EPrints: enfatiza particularmente o conteúdo de acesso livre a artigos de periódicos e teses; o *workflow* do EPrints está baseado numa interface *web* que lida com o processo de submissão; os documentos são transferidos para um armazenamento temporário que será moderado pela equipe do repositório. As etapas subseqüentes serão realizadas no documento e no registro dos metadados, o que inclui migração do formato do arquivo, ajustes nos metadados ou eliminação do documento (http://www.eprints.org).
b) DSpace: concentra-se no aspecto de curadoria digital em geral; o *workflow* é organizado para simplificar o processo de submissão e reunir informação relevante no repositório; ele inclui ferramentas para importar e exportar documentos e uma estrutura de diretório simples, na qual os metadados Dublin Core são armazenados em arquivos XML (http://www.dspace.org).

As pesquisas na área continuam a surgir e os objetivos são cada vez mais variados. No *RLG DigiNews* volume 9 número 4 de 2005, foram mencionados 10 programas promissores de pesquisa em preservação digital. O que caracterizam esses projetos é a necessidade de que a tecnologia possibilite mais procedimentos

automatizados de preservação digital e que existam mecanismos que integrem requisitos de auditoria no gerenciamento da preservação de objetos digitais.

Também existem recomendações publicadas pela National Science Foundation (NSF) nos Estados Unidos e pela Community Research & Development Information Service (CORDIS) na Inglaterra, sobre tipos de infra-estrutura que sejam proporcionais ao crescimento do tamanho dos repositórios de conteúdo digital.

Os padrões de conversão digital e as melhores práticas têm evoluído, assim como as iniciativas mais importantes de metadados e protocolos de intercâmbio, além das tecnologias para mover informação digital de um *hardware* e software para outro. Contudo, o desenvolvimento das infra-estruturas da preservação da informação digital tem sido lento.[43] Mas um sinal de maturidade da área é, segundo Lavoie e Dempsey (2004), o entendimento de que a preservação digital não é apenas uma parte técnica que acontece como um "processo isolado" mas, sim como parte de um ambiente maior que compreende toda a informação digital.

2.4.2 Infra-estrutura organizacional

Existem algumas áreas da infra-estrutura organizacional que precisam de definições mais claras, tais como a necessidade de integrar as políticas organizacionais para a preservação digital com as implementações tecnológicas e a necessidade de desenvolver e envolver pessoas com conhecimentos sobre preservação digital.

Segundo Nancy Y. McGovern (2007), o "pilar" organizacional da preservação digital determina o tipo de procedimentos que estão sendo empregados, se ela é resultado de um mandato ou uma norma institucional, seu escopo, seus objetivos e o pessoal comprometido. Na década de 1990, a infra-estrutura organizacional era o item menos observado pelas iniciativas de preservação digital, mostrando a ausência de comprometimento na própria missão

[43] Na página *Digital Curation Tools* do Digital Curation Centre encontra-se uma lista completa de ferramentas e tecnologias em software livre para os processos de preservação digital: http://www.dcc.ac.uk/tools/digital-curation-tools/ Acesso em: 12/out/2008.

das instituições de alguma declaração relacionada com a permanência das informações e a falta de programas de preservação digital sustentáveis. Naquele período, esperava-se por uma solução única que fosse o suficientemente forte e que servisse para inibir o desenvolvimento da responsabilidade organizacional pela preservação digital.

Já em 2002, com a publicação do relatório da RLG/OCLC sobre os atributos e responsabilidades dos repositórios digitais confiáveis e o desenvolvimento dos *softwares* abertos para repositórios e o Modelo de Referência OAIS (SAAI), a infra-estrutura organizacional da preservação digital se fortaleceu, facilitando o desenvolvimento de políticas de preservação digital nas organizações e o reconhecimento do papel central que tem a certificação.

Como mencionado no subseção 2.2.5, o modelo de referência OAIS/SAAI representa a expressão mais formal e compreensiva do processo de arquivamento disponível à comunidade que trabalha com preservação de objetos digitais; identifica documentos que são necessários, tais como um acordo de submissão, formatos padrão, controle de acesso físico, administração de bases de dados, gerenciamento de armazenamento, recuperação de desastres, evolução do sistema, padrões de migração e outros procedimentos.

O modelo proporciona uma linguagem comum e um conjunto de funções para serem discutidas e para poder mapear o desenvolvimento organizacional na questão da preservação digital. Por ser considerado um padrão para arquitetura dos repositórios e arquivos digitais, o OAIS/SAAI proporciona a estrutura onde os metadados de preservação e outros padrões podem ser desenvolvidos. Ele também possibilita que um *software* de repositório possa exportar de forma padronizada seus metadados arquivados em formato XML para facilitar a migração para outro sistema.

Considerado, desde 2003, por vários pesquisadores como a base conceitual dos repositórios de informação científica, o modelo OAIS/SAAI é um padrão ISO e a maioria dos sistemas de preservação digital está sendo reestruturada para ficar em conformidade com ele. Essas iniciativas estão tentando inserir Pacotes de Submissão de Informação (PSI) e produzindo Pacotes de Disseminação de Informação (PDI), com a finalidade de se adequar ao modelo.

Como uma entidade que opera na administração do OAIS, o planejamento da preservação tem como objetivo monitorar as mudanças tecnológicas que podem afetar tanto os consumidores como os produtores de um repositório digital. Nesta

parte do processo são realizadas análises periódicas das ferramentas que vêm surgindo, para desenvolver padrões abertos de formatos e sistemas.

Uma das funções de preservação do OAIS é suportar a migração e transformação dos dados com algum grau de automação (WALKER; THOMA, 2004). O desenvolvimento de futuros conversores de formatos poderá ser facilitado, se mais repositórios digitais programarem seus metadados seguindo o modelo OAIS, instalando aplicativos já em funcionamento no seu repositório (ex.: a ferramenta de conversão de documentos *MyMorph* da National Library of Medicine e a ferramenta de validação e extração de formato de objetos digitais *JHOVE* da Harvard Unversity Library), ou pelo o uso de serviços de migração de formato como o *TOM* (*Typed Object Model*) da Carnegie-Mellon University.

Apesar da importância das tecnologias a serem usadas para combater a obsolescência dos suportes e mídias, as mudanças organizacionais continuam a ser uma das ameaças para a estabilidade da informação de longo prazo (LAVOIE; DEMPSEY, 2004). O planejamento da preservação deverá considerar o uso de *software* livre, estruturado em linguagem de banco de dados SQL, para prover uma fácil interoperabilidade entre diversos sistemas.

2.4.3 Consolidação das iniciativas de preservação digital

Os sistemas de preservação digital se estendem para fora das instituições, usando ferramentas agregadoras, organizando comunidades de usuários de recursos digitais, incentivando o aumento do material disponibilizado nas bibliotecas digitais. Ainda assim, a preservação digital continua sendo um território desconhecido para muitas instituições públicas e privadas. Organismos de pesquisa de vários países têm manifestado suas reservas em relação às implicações econômicas da preservação a longo prazo, assunto que não tem sido tratado diretamente com as agências financiadoras.

Para Thibodeau (2002), entre os mais importantes desenvolvimentos na preservação digital desde 1996, está o alerta que a National Archives e a Library of Congress dos Estados Unidos (LC), têm levantado sobre a possível perda catastrófica da informação atual. Conseqüentemente, as duas agências têm

conseguido obter mandatos e financiamento do Congresso daquele país para abordar o problema da preservação digital.

As instituições tradicionalmente comprometidas com a preservação da herança cultural e da história têm empreendido alguns esforços com financiamento de governos locais. Os arquivos e bibliotecas nacionais têm manifestado o interesse oficial dos estados de documentar aspectos culturais relevantes de determinado período ou evento. Nos Estados Unidos, a Library of Congress têm realizado alguns dos maiores esforços (*American Memory*) que atraem o financiamento da iniciativa privada, mas o interesse da LC pelo crescente número de objetos nascidos digitais só começou em 2000, com a iniciativa chamada *National Digital Information Infrastructure Program* (NDIIP) (FRIEDLANDER, 2002).

Partindo da idéia de que exista a necessidade latente de garantir o acesso à informação de longo prazo, algumas ações têm sido propostas aos gerentes das coleções digitais tanto em instituições detentoras de acervos culturais, quanto nas dos acervos científicos. São elas: colaboração e intercâmbio de experiências, adoção estratégia estruturada de preservação digital e criação de uma infra-estrutura de repositórios de preservação digital.

Duas instituições que têm acompanhado o desenvolvimento das práticas da preservação digital são o International Council for Scientific and Technical Information (ICSTI) e o Grupo e Trabalho de Preservação Digital CENDI (U.S. Federal Information Managers Group). Os relatórios produzidos por essas duas instituições têm servido para acompanhar os projetos implementados por instituições das áreas de C&T (HODGE; FRANGAKIS, 2004). Os dados coletados são aqueles relacionados com práticas - divididas por tipo de documento, formato, *workflow* estabelecido pelo sistema operacional escolhido, atividades padrão - e a disponibilização dos "pacotes de *software*". No levantamento de 2004, foram coletados dados de mais de 50 sistemas ou projetos de preservação digital de vários países e instituições. As principais conclusões alcançadas foram:

a) um panorama do estado atual das práticas nas áreas de C&T;
b) a identificação dos mais importantes temas abordados;
c) um conjunto de lições aprendidas e futuras ações.

No Quadro 7, podem ser visualizadas as estratégias de preservação adotadas em vários projetos ou sistemas operacionais selecionados na pesquisa do CENDI/ICSTI (HODGE; FRANGAKIS, 2004).

		ESTRATÉGIAS					
		Política de Arquivamento	**Migração**	**Padrões Abertos**	**Metadados de Preservação**	**Colaboração**	**Repositório Institucional**
SISTEMAS	AIA		X		X	X	X
	AIP	X	X		X		
	ALEXA	X			X	X	
	DIAS		X	X		X	
	DiVA	X	X		X		X
	Dspace	X	X	X	X	X	X
	ESD					X	
	EROS	X	X		X		
	FEDORA™		X	X	X	X	X
	GPO		X		X	X	
	IUCR	X	X		X		
	JSTOR	X	X		X	X	
	LSDA		X		X		
	LOCKSS			X		X	
	NASA		X		X		X
	NLM	X	X				X
	NMM	X	X				
	OCLC						
	PANDORA	X	X	X	X	X	
	PUbMed	X	X		X		
	VERS	X	X		X		X
TOTAL		12	17	5	15	10	7

Quadro 7 – Estratégias de preservação adotadas pelos 21 sistemas operacionais
Fonte: CENDI/ICSTI – Hodge e Frangakis (2004)

Os 21 projetos selecionados por Hodge e Frangakis gerenciam ampla gama de recursos de informação científica. No Quadro 7, são mostradas as estratégias detectadas nos diferentes projetos de sistemas de preservação digital de informação científica. A maioria dos projetos possuía políticas de arquivamento e normas relacionadas para a criação de metadados. A estratégia de migração foi a mais usada nos sistemas levantados. A estratégia de emulação não foi mencionada como uma das mais importantes para essas organizações. A colaboração foi assinalada como significativa junto à comunidade de desenvolvedores de sistemas de preservação digital.

A União Européia criou o projeto Erpanet, estabelecido em novembro de 2001, tendo como finalidade unir as diversas iniciativas na área de preservação digital. O projeto europeu publicou o *Erpanet Digital Preservation Charter*, uma declaração dos países europeus que serve como fundo para a cooperação entre as diversas iniciativas, envolvendo cada um dos participantes do ciclo de vida dos

objetos digitais. O Erpanet apontou os motivos da importância de participar de projetos cooperativos de preservação digital:

1) alcançar melhor administração dos riscos;
2) participar de um marco comum de trabalho para *benchmarking* do acesso digital e das melhores práticas de preservação;
3) construir parcerias;
4) chamar a atenção os principais atores envolvidos com objetos digitais;
5) evitar redundância;
6) maximizar os esforços de preservação.

Nos países do Primeiro Mundo, estão sendo criadas redes de comunicação entre especialistas de uma ou mais nações, com o objetivo de discutir temas específicos da preservação de objetos digitais de longo prazo com características locais, como o idioma. Blake (2004) narra que na Alemanha foi criada uma rede bilíngüe para intercâmbio de experiências fora e dentro daquele país. A experiência tem favorecido a discussão sobre estratégias e políticas de arquivamento digital.

O tema da colaboração tem sido discutido também no contexto da interoperabilidade. É nessa estrutura que estão apoiados todos os aspectos dos repositórios digitais de acesso livre. Estrategicamente, essa abertura e cooperação geram uma multiplicação de benefícios para todos os participantes, incluindo novas habilidades e experiências, modelos de negócios sustentáveis, estratégias de preservação mais estruturadas etc. O uso de padrões abertos e de protocolos para a interoperabilidade dos arquivos permite a incorporação de colaboradores internacionais, formalizando redes de informação que minimizam a redundância dos recursos e custos na sua manutenção.

As idéias básicas do modelo OAIS podem ser detectadas na maioria dos projetos de pesquisa de preservação digital (LeFURGY, 2002). São casos como o da Library of Congress, baseado na digitalização usando a padronização dos formatos e o uso de metadados de preservação; outro exemplo já mencionado é o VERS da Australia National Library, que usa metodologias detalhadas para a criação de documentos digitais no gerenciamento arquivístico; cita-se também o *CURL Exemplars for Digital Archives* (CEDAR) na Inglaterra, que desenvolve metodologias baseadas no modelo de representação das propriedades dos objetos digitais; e o

projeto do Electronic Record Archives (ERA) dos Estados Unidos, que está utilizando o modelo OAIS nas suas propostas de gerenciamento de materiais digitais persistentes, definindo as estruturas e metadados necessários para seu acesso e recuperação.

As iniciativas nacionais de preservação digital têm surgido a partir de grandes instituições depositárias e de centros de pesquisa que vêm optando pela proposta dos Arquivos Abertos e o Acesso Livre. Algumas delas já foram destacadas por Muir e por Studwell e Ghetu, em 2003. Também uma lista mais abrangente encontra-se no Apêndice D. Entre as mais citadas na literatura estão as seguintes:

1. DIAS, projeto para *software* de arquivamento;
2. *Digital Archive*, o arquivo de documentos governamentais da *National Archives*;
3. HD-ROM e HD- Rosetta produto de preservação;
4. *Internet Archive*, projeto para *software* de arquivamento;
5. InterPares, projeto internacional;
6. JISC e SHERPA, plano e projeto de instituições da Inglaterra;
7. LOCKSS, projeto da Stanford University;
8. NDIIPP, plano de preservação digital da Library of Congress;
9. PADI e PANDORA, plano e arquivo *web* da National Library of Australia;
10. *Preservation Metadata Extraction Tool*, componente do programa de preservação digital da National Library of New Zealand;
11. UVC estratégia de cópia centralizada.

Os projetos mencionados têm como origem os grupos de pesquisa que, mesmo separadamente, possuem o mesmo objetivo. Alguns têm como centro de suas atividades a produção acadêmica e a construção de consórcios. Com o crescimento de repositórios institucionais, espera-se que essas instituições adotem mais os sistemas distribuídos e surjam mais iniciativas colaborativas.

Alguns esforços estão apontando nessa direção, como é o caso do projeto SHERPA (*Securing a Hybrid Environment for Research Preservation and Access*) na Inglaterra, que estuda implementação prática das recomendações para uma infra-estrutura técnica de preservação desagregada. As funções de preservação desagregada, segundo Beagrie (2002), são serviços de preservação institucionais

que podem dar suporte e beneficiar o planejamento de mecanismo de armazenamento à distância. As razões para o surgimento dessa proposta estão na natureza dos repositórios institucionais que, segundo James *et al.* (2003), apresentam as seguintes características:

a) a maioria não tem como inerente os processos de preservação digital;
b) não contam com pessoal suficiente com conhecimentos sobre preservação digital e,
c) replicam serviços e custos.

Os trabalhos nas iniciativas e projetos de preservação digital continuam reproduzindo-se de forma rápida, com desenvolvimentos significativos de sistemas, pacotes de *softwares*, legislações de depósito, parcerias e propostas de novos padrões. Mesmo com isso, ainda existem algumas áreas a serem abordadas. Na seção a seguir, trataremos dos fatores que levam à sustentabilidade dos sistemas de preservação digital emergentes.

2.5 SUSTENTABILIDADE DA PRESERVAÇÃO DIGITAL

As pesquisas internacionais com os responsáveis por coleções digitais mostram que as organizações ainda não podem garantir recursos para a preservação digital. Essa deficiência de recursos é uma das grandes ameaças à preservação digital identificadas por especialistas da área (BEAGRIE, 2003; BRADLEY, 2004; WHEATLEY, 2003), além de projetos e políticas insuficientes, suporte inadequado por parte das equipes, a obsolescência tecnológica e a falta de experiência dos responsáveis.

Diferentemente da infra-estrutura organizacional que fica incorporada no documento do TDR, e da infra-estrutura tecnológica definida dentro do modelo de referência OAIS, a parte dos “recursos necessários” para a preservação digital não tem nenhum documento produzido pela comunidade que expresse qual é seu escopo e seus requisitos. O que a literatura reporta é que poucas instituições têm alocado recursos dos seus orçamentos para os cuidados com os materiais digitais de longo prazo. A necessidade de realizar ações imediatas para preservar coleções

digitais fica atrás do desejo de criar e disseminar novas formas de conteúdo digital (LAVOIE; DEMPSEY, 2004).

A sustentabilidade econômica é a habilidade de conduzir recursos suficientes e de forma constante para alcançar os objetivos da preservação digital. O fato de haver incluído o atributo “sustentabilidade financeira” no documento TDR (RLG/OCLC, 2002) simboliza a primeira vez que os custos da preservação digital foram expressos explicitamente. Um avanço nesse aspecto foi o fato de que, pelo menos no exterior, algumas instituições do ensino superior destinam fundos para a preservação digital (ex.: o MIT para o DSpace, University of Cornell para projetos de coleções digitais). Também é considerado um avanço o fato de existirem programas de preservação digital que têm um tempo de vida maior do que os projetos de preservação digital (ex.: JISC, NEH, NSF, NHPRC). Esses programas desenvolveram políticas e receberam fundos destinados à pesquisa.

Além desses avanços, a comunidade da área tem produzido ampla literatura sobre os custos que a preservação digital envolve. Alguns desses trabalhos tratam sobre modelos econômicos (LAVOIE, 2003; SANNETT, 2002; 2003) e ferramentas para comparar custos entre a migração e a emulação ao longo do tempo (OLTMANS; KOL, 2005) entre outros. Esses documentos têm ajudado a entender os custos, mas não têm sido suficientes para a criação de um documento produzido pela comunidade capaz de sustentar o impacto que os custos representados pelo desenvolvimento e as implementações das estratégias de preservação podem causar em cada caso, particularmente.

Para os teóricos da preservação digital, a distribuição dos custos está determinada pelo contexto organizacional e está amarrada à implementação tecnológica no programa de preservação digital de uma organização. Em resumo, seria o “o que” e o “como” da preservação digital determinando o “quanto”. A designação de fundos por parte das organizações para a preservação digital e um documento por parte da comunidade que expresse os custos gerenciais de estratégias de preservação digital compõem os dois aspectos que ainda se encontram em construção (LAVOIE, 2008).

Cada estratégia técnica de preservação e de acesso implica diferentes custos e cronogramas. A preservação digital requer recursos disponíveis permanentemente que começam a se delinear no momento da criação do recurso.

Muitos materiais são produzidos fora dos centros de informação e das bibliotecas. Nesses casos, são os editores e outros criadores de conteúdos que devem adotar os padrões apropriados e as tecnologias que darão subsídios à preservação. Segundo Feeney (1999), “[...] diferentemente da situação que se aplica aos livros, o arquivamento digital requer investimentos relativamente freqüentes para superar a obsolescência rápida produzida pelas mudanças tecnológicas”.

Os provedores de material digital, no correr do tempo, precisam investir para criar documentação e metadados, gerando novas formas de material para manter o acesso. Esse investimento deve ser levado em consideração no momento de discutir os direitos de uso e reuso dos objetos digitais. Do mesmo modo, entre os aspectos que devem ser identificados em qualquer estratégia de preservação, está a necessidade de contratação e de capacitação de pessoal. Se não houver nenhum plano administrativo que cubra esses itens, nenhuma manutenção de acervos de longo prazo será bem-sucedida.

Quanto à preservação digital, Chilvers (2000) deixa claro que sua viabilização está relacionada aos aspectos financeiros, gerenciais e técnicos da instituição onde ocorrerá, e tem como objetivo prolongar a vida útil do dado. Existe uma lacuna entre estabelecer uma estratégia de preservação e a sua implementação. Uma vez que uma das grandes vantagens do documento digital é a sua interoperabilidade com vários outros mecanismos ao mesmo tempo, se a questão da informática (*hardware*, *software* e obsolescência tecnológica, entre outros) não estiver resolvida, sofrendo constantes análises e avanços, o documento perderá uma de suas características vitais e que o distingue de outras formas de documentos.

O problema maior aparece no quesito financiamento, já que a maioria dos projetos consegue recursos apenas por curto período de tempo, quando qualquer projeto de preservação digital requer um investimento a longo prazo. A natureza dos registros proporciona a dimensão do valor que eles podem chegar a ter no decorrer do tempo (SANT’ANNA, 2001).

O *Digital Archiving Working Group,* da *JISC/NPO*, analisando o gerenciamento dos custos da preservação digital, sugere que os materiais digitais sejam arquivados em formatos padrão, em mídia padrão, e gerenciados dentro de alguns dos poucos modelos de sistema de preservação operacionais existentes.

Para Hendley (1998), existem algumas dificuldades no desenvolvimento de modelos para custos com a preservação digital. Segundo ele, as organizações precisam levar em consideração a quantidade e o nível de acesso que pretendem oferecer, pois isso influencia os custos. Também, o treinamento de especialistas e técnicos deve ser contemplado nos custos das atividades da estratégia de centros de preservação digital (CHAPMAN, 2003).

O custo de estruturar o que envolve esta tecnologia, para garantir o trabalho cultural e intelectual da era digital, será ainda mais notável se desenvolvido ainda no começo do planejamento dos custos, para que estes sejam minimizados e efetivamente aplicados. Pontos-chave como este devem ser pensados para identificar e selecionar práticas apropriadas de custo-benefício para cada estágio no ciclo de vida dos recursos digitais.

A preservação digital pode ser considerada uma atividade economicamente sustentável, apesar de serem insuficientes os orçamentos atribuídos a várias iniciativas até hoje. A decisão de como devem ser atribuídos os recursos financeiros continua influenciada pela vontade dos administradores de criar e disseminar novas formas de conteúdo digital e pela atribuição de fundos para projetos de curta duração. Uma perspectiva colaborativa propiciaria a distribuição dos custos entre várias instituições, mitigando gastos com infra-estrutura e pessoal especializado.

Em geral, os custos com a preservação digital, se não forem compartilhados entre instituições parceiras, podem chegar a ser significativos e até ultrapassar os requeridos para sua contrapartida em papel. Economicamente, a sustentabilidade da preservação digital depende dos seguintes fatores:

a) reconhecimento dos benefícios de preservar materiais digitais seguindo critérios específicos de seleção;
b) incentivos apropriados para provocar às autoridades para atuar a favor do interesse público;
c) mecanismos que garantam a destinação de recursos para as atividades de preservação digital nas organizações, local e coletivamente;
d) uso eficiente dos escassos recursos financeiros, e
e) planejamento organizado e com autoridade das atividades definidas nas políticas de preservação digital.

No exterior, o financiamento estatal tem sido imprescindível para o desenvolvimento de projetos de preservação digital. Um exemplo é o da Library of Congress, que em 2000, desenvolveu o *National Digital Information Infrastructure and Preservation Program* (NDIIP), como uma arquitetura nacional para a preservação de material digital, por meio de uma rede de parceiros formada por bibliotecas, editoras, agências federais e do setor privado. O plano estratégico inicial aprovado com orçamento de US$ 25 milhões contou com parcerias entre empresas e comunidades nacionais e internacionais nas áreas de editoração, tecnologias e direitos autorais. Para 2007, o financiamento federal e as contribuições particulares chegaram a US$ 75 milhões.

Agências de fomento estão começando a participar; fundações como a Andrew W. Mellon Foundation têm financiado projetos de arquivamento de publicações eletrônicas, um deles sobre os arquivos digitais disponibilizados na rede e outro que focaliza aqueles que são armazenados pelas editoras. Com as pesquisas desenvolvidas nesses projetos pode-se extrair informações para os bibliotecários e editores sobre a necessidade de confiar em seus parceiros no campo digital. A confiança é um dos elementos fundamentais da infra-estrutura da preservação digital a ser criada em qualquer país.

2.5.1 Sustentabilidade comercial

É importante reconhecer que cada estratégia técnica de preservação e de acesso implica diferentes custos e cronogramas. Os grandes projetos de repositórios digitais precisam de uma sustentabilidade financeira comprovada para serem considerados “confiáveis”. O documento da RLG/OCLC (2002) recomenda que os repositórios digitais devam estar aderidos a corretas práticas de negócios e possuírem planos economicamente sustentáveis.

O grau de sustentabilidade financeira pode estar influenciado pelo tipo de organização que financia e pelas características dos serviços de informação. As iniciativas mantidas pelos governos e com obrigatoriedade no depósito legal e guarda de informações oficiais podem ser mais facilmente beneficiados, enquanto aquelas instituições que tenham como propósito o acesso à informação (ex.: as IES)

podem se respaldar nas necessidades dos usuários e no desenvolvimento dos seus acervos.

As empresas privadas têm estado envolvidas na preservação digital, desde que foram introduzidas as primeiras versões dos *hardware* e *software*. O interesse na preservação de dados armazenados nas primeiras versões se faz presente na oferta das novas soluções. Seus efeitos nos projetos de preservação digital podem ser percebidos na estimativa dos custos na implementação desses projetos. Eis alguns dos serviços comerciais de preservação digital usados no exterior são:

a) *Archivas, Permabits, Bamboo*
b) *Chronopolis grid storage* (SDSC, USA)
c) *IBM Digital Library system*
d) *Iron Mountain, Akamai*
e) *JSTOR*
f) *OCLC digital archive*
g) *Portico* (*Ithaka Harbors, Inc.*).

Em vários países as companhias que oferecem seus serviços e aplicações de preservação digital têm trabalhado colaborativamente com instituições de pesquisa e agências governamentais. Por outro lado, o setor privado tem criado soluções para a preservação da informação digital de arquivos de *e-mail*, texto completo e bases de dados, aplicáveis em diferentes momentos do ciclo de vida do documento digital. Entre as dificuldades que apresenta o trabalho com as companhias privadas no campo da preservação. Dois aspectos se ressaltam:

1) ausência do registro da guarda dos dados e dos dados estatísticos das falhas; esses dados são considerados como “rotinas” ou proprietários;
2) os dados técnicos dos registros são guardados como sendo departamentais ou pessoais e não corporativos, ficando distribuídos em ferramentas e versões operadas sem controle das atualizações.

Para Gladney (2001) o problema da preservação digital na área comercial está centrado no fato de que os beneficiados não vislumbram a utilidade dessas atividades senão até depois de muitos anos de o material haver sido salvo. Hodge e Caroll (1999) afirmam que os atores principais não percebem o problema da mesma forma: para os criadores dos documentos digitais, o envolvimento com a permanência dos dados está reduzido apenas ao processo de arquivamento

(começando a perceber a importância dos metadados) e para os editores o mínimo de ação (procedimentos de *backup*) já é suficiente. Alguns editores comerciais responsáveis pela guarda permanente de dados individuais apóiam a cobrança do acesso e armazenamento dos registros digitais.

Uma tendência das organizações ao tratar de obter recursos para a construção de seus acervos digitais está sendo a oferta de serviços com valor agregado relacionados com a preservação dos materiais armazenados. Um exemplo é o OCLC, que oferece um serviço para o gerenciamento de coleções de bibliotecas digitais chamado *ContentDM*, disponível para ser usado de forma gratuita no servidor do OCLC apenas para os associados ao consórcio *Digital Co-op*, que ganham descontos em outros serviços, como cursos e suporte técnico.

O OCLC disponibiliza serviços terceirizados de arquivamento digital no seu *OCLC Digital Archive*, consistindo numa inscrição anual e cobrando pelos custos de armazenamento mensalmente. A cobrança está determinada pelo número de *gigabytes* do conteúdo armazenado. Os participantes do *Digital & Preservation Coperative* também recebem descontos.

Um caso de modelo econômico para financiamento de projetos de preservação baseados em plataforma de acesso livre é o do MIT, que tem como plano (ex. *DSpace*) oferecer serviços pagos de suporte para a criação de metadados específicos para serviços de apoio a instituições que adotam o *DSpace*.

A oferta de serviços padronizados de preservação digital dependerá do crescimento da demanda das organizações por formas economicamente sustentáveis de atividades eficientes de preservação, dentro de sistemas distribuídos. Segundo Lavoie (2004), a chave para um mercado sustentável de serviços de preservação é cultivar a confiança entre aqueles que oferecem os serviços e os que são potencialmente responsáveis pelo uso desses serviços.

Finalmente, a sustentabilidade econômica da preservação digital precisa de mais dados sobre os custos reais das infra-estruturas tecnológicas e organizacionais, e fica constatada a impossibilidade de manter um projeto de preservação digital desenvolvido localmente se não houver parcerias e distribuição dos custos com maior número de instituições. Esse estágio será alcançado quando os projetos de preservação digital passarem da fase de inicial de avaliação da reformatação dos suportes físicos, do estabelecimento de repositórios digitais e da inserção padronizada de registros digitais.

2.6 POLÍTICAS DE PRESERVAÇÃO DIGITAL

Muitos criadores de material digital com informação científica e tecnológica precisam ter o controle do acesso a sua produção. Por não possuírem conhecimentos de estratégias de preservação, estão preocupados com o fato de os materiais eletrônicos serem fáceis de copiar e redistribuir. Em contrapartida, os provedores do material no correr do tempo precisam investir para gerar documentação revisada, metadados e criar novas formas de material para manter o acesso (LAVOIE; GARTNER, 2005). A solução ideal é a aliança entre a organização habilitada no gerenciamento dos dados digitais e os criadores, de tal forma que possam criar conhecimento do material mantendo o controle das decisões sobre o conteúdo a ser preservado e os intervalos para essa atividade.

Os sistemas de gerenciamento de conteúdos digitais, tais como repositórios institucionais e bibliotecas digitais de teses e dissertações, propõem-se a preservar os conteúdos digitais dos quais as instituições possuem os direitos autorais. Algumas delas assumem esse direito e outras obtêm a confirmação dessa permissão com a assinatura dos autores desses trabalhos (ex.: a Biblioteca Digital de Teses e Dissertações no Brasil).

Em países como os Estados Unidos (*Digital Millennium Copyright Act*, 1998), as legislações nacionais requerem que seja solicitada permissão dos detentores dos direitos autorais para preservar cópias de conteúdos protegidos com alguma forma de direito autoral. Tanto os materiais de acesso livre como os de acesso restrito podem exigir algum tipo de permissão.

No que diz respeito aos repositórios de acesso livre, todos os trabalhos depositados continuam sendo propriedade intelectual dos seus respectivos autores. Cabe a estes a responsabilidade de decidir o que deve ou não ser depositado e que documentos podem ou não ser consultados na íntegra.

Nas instituições em geral, a maior parte dos documentos e publicações técnico-científicas é resultado das atividades de seus recursos humanos e, por isto, pode ser autodepositada e disponibilizada no repositório. O acesso livre sem restrições de *copyright* é viável pelos seguintes motivos: a) os autores são os que detêm os direitos intelectuais das suas teses e dissertações, da maioria dos trabalhos

apresentados em eventos (a menos que tenham transferido todos os direitos explicitamente), dos relatórios técnicos, *working papers* etc.; b) a maioria significativa dos periódicos científicos (aproximadamente 92%) permite algum tipo de autodepósito dos pré-*prints* (versões de trabalhos anteriores à avaliação pelos pares e aceite para publicação) e/ou de pós-*prints* (versão final, depois de ter sido avaliada pelos pares aprovada para publicação) dos artigos que eles publicam (HARNAD, 2007); e c) mesmo quando autores assinaram alguma forma de transferência de *copyrights* restrito, a consulta àqueles que detêm os direitos pode resultar na autorização (ou no mínimo na não proibição) do autodepósito no RI.

A disponibilização de documentos em acesso livre pode constituir uma infração apenas nos casos de licença concedida a terceiros (editores, por exemplo) pelo autor/instituição, ou por conter assuntos confidenciais ou que sejam destinados à comercialização (livros). Neste caso, apenas a sua referência (e seus respectivos metadados) pode ser disponibilizada no repositório.

Um exemplo de sistema que especifica que unicamente serão depositados materiais com permissão de acesso é o *LOCKSS*. Ele replica conteúdos de materiais em bibliotecas e estabelece que seja acrescentada uma declaração por parte da organização detentora dos direitos autorais das publicações preservadas, na qual é garantida a permissão do acesso aos materiais para os leitores da instituição que hospeda o *LOCKSS*. Também, com a adoção de licenças *Creative Commons* os sistemas estão providenciando as permissões necessárias para a preservação digital.

Para as instituições de depósito legal, como as bibliotecas nacionais, a legislação para a preservação dos documentos já está estabelecida e isso incentiva os detentores de direitos autorais a cooperar e colocar seus materiais à disposição. Em janeiro de 2000 a Inglaterra foi o primeiro país a introduzir a recomendação para o depósito legal de publicações eletrônicas, sendo plausíveis de negociação caso por caso.

O projeto *Internet Archive* toma uma abordagem alternativa ao coletar e preservar conteúdos protegidos com direitos autorais sem ter solicitado permissão. A estratégia não favorece uma política de preservação digital porque ela força os gerentes dos sistemas a retirar imediatamente qualquer material que seja identificado como protegido por leis de direitos autorais. A forma a ser adotada por gestores de

sistemas de informação seria obter e guardar a declaração expressa da permissão para preservar cada documento.

2.6.1 Permissões para preservar

A elaboração de políticas e o estabelecimento de iniciativas pioneiras de preservação digital estão colaborando para que os padrões e normas que estão surgindo possam afirmar-se como estratégias confiáveis e soluções a serem seguidas pelas instituições responsáveis pela informação científica em formato digital.

Um dos princípios estabelecidos em convênios internacionais e em várias legislações nacionais é a obrigatoriedade do depósito em arquivos ou bibliotecas de uma ou mais cópias dos trabalhos produzidos por instituições públicas e privadas. Mas na maioria dos países o depósito legal ainda não é obrigatório para os documentos digitais, apesar de, como mencionado anteriormente, esse tipo de material requerer um gerenciamento capaz de garantir que o material produzido em um país esteja disponível permanentemente.

A meta do depósito legal é garantir o acesso. Ao se estabelecer uma legislação para o depósito legal de objetos digitais, levantam-se questões ligadas ao acesso, tais como os custos e a proteção dos direitos do autor/editor (PADI, 2003). Países como a Nova Zelândia já estenderam suas políticas aplicadas às versões impressas para os materiais eletrônicos e versões *on-line*. Na Noruega, a mesma legislação está voltada em primeira instância para a preservação cultural e a proteção dos direitos autorais de qualquer “[...] trabalho que possa ser lido, ouvido, transmitido e que possa ser escrito e reproduzido em formatos eletrônicos no futuro [...]” (VAN NUYS, 2003, p. 16).

As bibliotecas nacionais de vários países estão começando a aplicar algumas normas e leis que combinam diferentes formas de desenvolvimento de coleções e proteção dos direitos autorais. Elas mencionam a importância de desenvolver rotinas para a seleção, descrição, identificação e armazenamento de todo tipo de documento digital. Uma negociação entre as bibliotecas nacionais e as

editoras começa a criar esquemas que permitam o depósito voluntário, seguindo acordos bilaterais (HODGE; FRANGAKIS, 2004).

Em vários países, com o crescimento no número de projetos de preservação digital, começam a aparecer documentos e decisões que representam o fortalecimento do aspecto organizacional da preservação digital. Os responsáveis por esses projetos detectaram os elementos essenciais para programas de preservação digital fortes, tais como: seguir e adotar os projetos já testados, implementar as "melhores práticas", trabalhar colaborativamente e documentar as políticas, procedimentos e padrões adotados (PADI, 2003).

Desde 1996 vem surgindo políticas de preservação digital em diferentes tipos de organizações, universidades, arquivos nacionais, bibliotecas nacionais, etc. Algumas das mais conhecidas citadas por McGovern (2007) são as seguintes:

- Arts and Humanities Data Service (2004)
- British Library (2002)
- Columbia University Library (2000)
- Cornell University Library (2004)
- Inter-University Consortium for Political and Social Research (2007)
- Library and Archives Canada (2002)
- National Archives UK (2000)
- National Library Australia (2001)
- National Library of New Zealand (2003)
- National Library of Wales (2003)
- Nestor Policy Project da Alemanha (2004)
- North Caroline Department of Cultural Resources (2000)
- UK data archives (2005)

Os argumentos relacionados com uma política de preservação digital observados nos exemplos citados e publicados no exterior podem ser exemplificados por duas das primeiras instituições a divulgar os pontos focais das legislações propostas.

O primeiro exemplo é o modelo de política de preservação digital da NLA, promulgado em 24 de fevereiro de 2002, que se destaca pela sua natureza política e enfoque institucional, declarando a intenção de "[...] preservar e manter documentos (digitais ou não) [...]", contando com a colaboração de instituições parceiras.

Depois de mencionar o escopo da política, adota-se uma definição de preservação digital e reafirma-se que as ações serão aplicadas a qualquer tipo de material digital segundo a natureza dos documentos. O modelo a ser seguido é o OAIS e ele fica explícito na declaração dos seus princípios operacionais:

1) necessidade de desenvolver práticas adequadas e coerentes para o tratamento mais apropriado dos materiais digitais;
2) uso do modelo de referência OAIS como um exemplo para a construção e gerenciamento do seu arquivo;
3) comprometimento no uso de normas e padrões internacionais no desenvolvimento de dos sistemas e da infra-estrutura para a preservação digital;
4) reconhecimento da importância da experimentação e desenvolvimento de processos que compreendam modelos de negócios e cumprimento da norma OAIS;
5) definição do Pacote de Arquivamento de Informação a ser usada seguindo o OAIS.

Além de se comprometer a incentivar a pesquisa na área e a dar apoio ao desenvolvimento de padrões e normas, a política da NLA expressa o comprometimento da instituição com relação às coleções digitais nesse país. Ela abre espaço para a especificação do uso de estratégias de preservação digital e do trabalho com órgãos oficiais para fundamentar a sustentabilidade econômica.

O segundo exemplo de política de preservação digital é da *British Library*, formalizado no dia 14 de novembro de 2002. Centrada na sua missão de preservar materiais digitais, a British Library adota uma abordagem estratégica na implementação das ações de preservação de objetos digitais, entre as quais estariam as de colaboração, estudo e pesquisa das características, desafios e especificações desses materiais.

O comprometimento institucional da British Library vai até o estabelecimento de requisitos com relação a estratégias a serem implementadas, controle de formatos, projetos de digitalização e disposição para a implementação de soluções novas.

As duas políticas citadas são exemplos da formalização das intenções de programas e projetos de instituições amparadas nas leis de depósito legal que estão começando a formular critérios para o estabelecimento do elemento preservação

nas soluções a serem adotadas para o armazenamento de objetos digitais (FOOT, 2001).

2.6.2 Gerenciamento dos direitos de propriedade

Dentro do contexto estratégico de criação e preservação de recursos digitais, o relatório da Arts and Humanities Data Service (AHDS) da Inglaterra ressalta o contexto das políticas aplicáveis para os três estágios dentro do ciclo de vida de um recurso digital: criação, administração/preservação e uso (BEAGRIE; GREENSTEIN, 1998). Uma das principais conclusões foi a de que as decisões que podem afetar a disponibilização dos materiais digitais precisam ser tomadas no início do ciclo de vida do objeto digital.

Para os materiais digitais, a ligação entre sua criação e a preservação é muito mais importante devido a que decisões sobre a forma como os objetos digitais são criados influenciam a maneira como eles podem ser preservados. Do mesmo modo, as decisões tomadas no momento da preservação podem influenciar como esse material será acessado no futuro.

O gerenciamento dos direitos de propriedade intelectual de materiais digitais tem um significado mais argüível do que na mídia tradicional, pois deve considerar não apenas o conteúdo, mas também qualquer assunto relacionado ao *software*. Qualquer ação (cópias, encapsular conteúdo, emulação de *software*, migração de conteúdo) envolve atividades que podem infringir permissões específicas dos que mantêm os direitos.

Esse tipo de gerenciamento cobre todos os processos que envolvem a definição e a propriedade dos direitos dos depositantes e os direitos dos usuários dos centros de informação. Greenstein (1997) define direitos autorais como aqueles que incluem direitos de propriedade intelectuais e assuntos legais relacionados com a proteção dos dados e confidecialidade. Ele assegura que os direitos inseridos em um recurso podem não apenas determinar como ele pode ser acessado e usado, mas também determinar o como e se eles podem legalmente ser preservados por terceiros.

No que se refere ao direito autoral, dois grandes acordos regulam a questão: a Convenção de Roma, que é o principal tratado internacional em direito conexo, e a Convenção de Berna, que é o tratado autoral do direito internacional principal. Esta última não faz uma isenção específica para copiar todos os tipos de trabalho para finalidades da preservação (BERNE, 1986). Fornece, sim, uma regra geral sobre o direito de reprodução:

1. as isenções relativas ao direito de reprodução não devem ser demasiadamente amplas, devem ser incorporadas na lei para um alvo específico e devem servir ao interesse público. Se uma isenção estatutária fosse adotada especificamente para finalidades da preservação, tais circunstâncias seriam satisfeitas;
2. uma isenção não deve se opor à exploração normal do trabalho (WIPO, 1996).

No Brasil, a lei que trata desta questão é a de nº 9.610, sancionada em 19 de fevereiro de 1998 pelo então presidente Fernando Henrique Cardoso. Para as unidades de informação, a definição de publicação, constante do art. 5º, é primordial: "I - publicação - o oferecimento de obra literária, artística ou científica ao conhecimento do público, com o consentimento do autor, ou de qualquer outro titular de direito de autor, por qualquer forma ou processo".

Preservar significa copiar de alguma maneira, e o artigo 5º desta lei define reprodução como:

> "VI - a cópia de um ou vários exemplares de uma obra literária, artística ou científica ou de um fonograma, de qualquer forma tangível, incluindo qualquer armazenamento permanente ou temporário por meios eletrônicos ou qualquer outro meio de fixação que venha a ser desenvolvido".

Já o art. 29 afirma que depende de autorização prévia e expressa do autor a utilização da obra, por quaisquer modalidades, tais como "I - a reprodução parcial ou integral [...]", e o art. 33 completa mostrando que ninguém pode reproduzir obra que não pertença ao domínio público, a pretexto de anotá-la, comentá-la ou melhorá-la, sem permissão do autor.

No caso do armazenamento de documentos digitais em bases de dados, a lei 9.610 confirma a determinação existente de que depende da autorização prévia do autor, pelo fato de, ao serem incluídos, transformar-se em uma forma de disseminação e reprografia (GANDELMAN, 1997).

Na Koninklijke Bibliotheek (KB) em Haia, a comissão do Institute for Information Law da Universidade de Amsterdã examina os problemas relacionados ao direito autoral que poderiam impedir esforços para preservar e manter a acessibilidade de dados eletrônicos, uma vez que a preservação implica a reprodução em grande parte dos casos.

De acordo com as diretrizes da KB, há dois métodos para garantir a preservação de longo prazo de publicações eletrônicas:

1) assegurar sua acessibilidade: a publicação pode ser retida em seu formulário original junto com o *hardware* e a tecnologia necessários para acessar, isto é construindo uma "passagem" entre a velha e a nova tecnologia;
2) transferir a publicação às gerações mais novas de configurações do *hardware* e *software* do computador, assim que a mais velha ameaçar se tornar obsoleta. Isso envolve um número das técnicas e procedimentos (incluindo conversão e transferência de dados) geralmente conhecidos pelo termo geral "migração".

Se a primeira opção for escolhida, a publicação eletrônica ainda assim precisará, eventualmente, ser copiada para ser preservada, já que, como todas as mídias existentes, as publicações eletrônicas estão suscetíveis à deterioração. A conclusão final do *Preservation of New Technology*, um estudo compilado a pedido da American Commission on Preservation and Access of the Council on Library and Information Resources, é que preservação significa copiar (BOERES; MÁRDERO, 2005).

Em 2003 a *NARA* (U.S. National Archives and Records Administration) adotou a norma ISO 15489 para o gerenciamento dos registros eletrônicos em todo o ciclo de vida dos documentos digitais de caráter oficial do governo americano. Essa norma surgiu na Austrália (AS 4390) e está sendo usada como base para políticas de preservação de documentos digitais de várias instituições daquele país. A norma considera a gestão dos “registros” como uma seqüência integrada de processos, que vai desde quem usa para criação até seu acesso contínuo, ampliando as responsabilidades dos criadores e gerentes desses registros para todos os empregados da instituição mantenedora.

O pré-requisito de preservabilidade pode ser alcançado unicamente mediante uma estratégia organizacional. Para a maioria das instituições de ensino

superior, a carência de tempo, recursos e conhecimentos necessários para garantir a preservação de sua produção intelectual de longo prazo, bem como a adoção de uma política de preservação poderiam facilitar o entendimento dos gestores sobre as principais conclusões a que chegam os estudos atuais sobre o tema (BOÊRES, 2004; ERPANET, 2003; FEATHER, 2004; FOOT, 2001):

1. a preservação de informação digital requer colaboração entre organizações e pessoas envolvidas na criação e no gerenciamento;
2. a preservação começa no *design* de sistemas que darão suporte à criação de objetos digitais;
3. deve ter-se claro o papel de cada um dos envolvidos no processo para identificar e apontar responsabilidades;
4. a preservação deve ser parte integrante de qualquer estratégia que inclua o uso de tecnologia de informação;
5. as estratégias de preservação e seus métodos deverão estar integrados em todas as atividades ou sistemas que criam ou usam informação digital.

Uma política aceitável de preservação digital implica observar e aplicar procedimentos que podem ser inclusive aceitos como estratégias de preservação. Entre eles estão os relativos à tecnologia da informação, mais especificamente no tocante à compatibilidade de *hardware, software* e migração dos dados (conversão para outro formato físico ou digital, emulação tecnológica e “espelhamento” dos dados), à observação da integridade do conteúdo intelectual a ser preservado; à análise dos custos envolvidos no processo; ao desenvolvimento de uma criteriosa política de seleção do que será preservado e, intimamente atrelada a isso, a observação das questões mencionadas sobre direito autoral.

2.7 MAPEAMENTO DOS CRITÉRIOS DE PRESERVAÇÃO DIGITAL

Entre as possibilidades de abordagem ao tema da preservação digital, está tentar identificar os critérios que sirvam como padrão para comparar e julgar as melhores práticas nas organizações. Numa área relativamente nova e com uma variedade ainda não completamente definida de aplicações, torna-se necessária a

formulação de indicadores que possam auxiliar na fundamentação de uma estratégia a ser escolhida.

Devido ao escopo do tema desta tese, o estudo dos critérios de preservação digital focalizou-se nos projetos e iniciativas relacionadas com a informação técnico-científica; são exemplos predominantemente resultantes de instituições de ensino e pesquisa que estão sendo consideradas como boas práticas e experiências de aprendizagem no campo dos repositórios digitais.

A avaliação de qualquer prática relacionada com os recursos digitais envolve estimá-los segundo critérios práticos e técnicos. Greenstein foi o primeiro a enfocar os critérios práticos de preservação relacionados a um conjunto de atividades que determinam a manutenção de coleções digitais de informações científicas e culturais. Essas condições básicas (BULLOCK, 1999) foram identificadas e aplicadas por vários especialistas (bibliotecários, arquivistas, gerentes de bibliotecas digitais) em diferentes iniciativas das quais partiram as primeiras políticas de preservação digital.

O desenvolvimento dos programas e projetos de preservação tem apontado a necessidade de chegar a um acordo, no que seria o conjunto mínimo de critérios para o funcionamento de um repositório de preservação da produção científica em formato digital. A maioria das propostas de critérios que vem surgindo menciona a influência do modelo de referência OAIS/SAAI, assim como, também, elas apresentam um conjunto de critérios para o planejamento de repositórios digitais.

Alguns esboços de critérios encontravam-se presentes em propostas de preservação digital que focalizavam:

- o controle físico e intelectual dos materiais para garantir sua preservação a longo prazo, com credenciamento de formatos de arquivos requeridos;
- a padronização de estratégias de migração, validação de dados e especificações normativas para mídia física;
- a padronização no uso de metadados de preservação.

No centro dessas propostas, a maioria proveniente da área arquivística, já estava contido o conceito de arquivos ou repositórios digitais confiáveis. A confiabilidade vem diretamente do principio arquivístico de custódia responsável.

Como observa Thomaz (2007), a confiança nos repositórios digitais se desenvolve em no mínimo três níveis:

1. a confiança de que os produtores estão enviando as informações corretas;
2. a confiança de que os consumidores estão recebendo as informações corretas; e
3. a confiança de que os fornecedores estão prestando serviços adequados.

A primeira lista de atributos e responsabilidades de repositórios digitais confiáveis foi estabelecida pela Research Libraries Group (RLG) e o Online Computer Library Center (OCLC) no seu relatório publicado em 2002. Esse conjunto de atributos está influenciado pelo modelo OAIS/SAAI e estabelece que essa obediência ao modelo deva ser considerada o primeiro critério a ser observado por um repositório confiável.

Para esse grupo, os repositórios confiáveis devem incluir também, atributos que possam dar suporte a sistemas de segurança e aos procedimentos adequados e significativos. Eles apontam que todo repositório confiável deve incluir atributos que sustentem os seguintes aspectos:

a) responsabilidade administrativa;
b) viabilidade organizacional;
c) sustentabilidade financeira;
d) adequabilidade tecnológica e procedimental;
e) sistema de segurança;
f) responsabilidade de procedimentos (certificação).

O relatório da RLG e da OCLC codifica esses atributos no conjunto de responsabilidades baseado no modelo OAIS/SAAI, manifestado suas responsabilidades de custódia e definindo, assim, o que pode ser considerado como uma lista de requisitos funcionais (Figura 18).

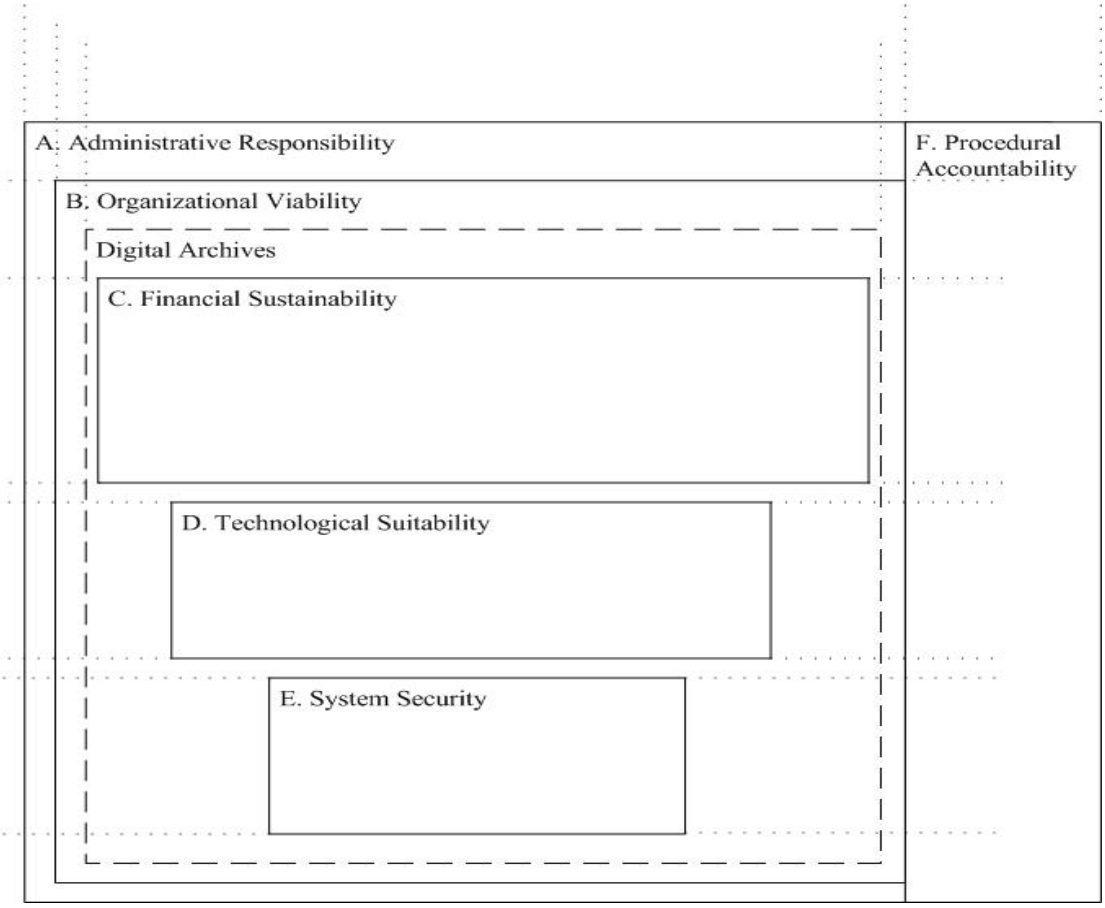


Figura 18 - Modelo de *Trusted Digital Repository* (TDR) da RLG e OCLC

O modelo *Trusted Digital Repository* representa a primeira forma de expressar a infra-estrutura organizacional da preservação digital. Os atributos do TDR converteram-se em padrões para a comunidade da preservação digital, pois antes deles não existia uma expressão formal do contexto organizacional da preservação digital.

Uma nova versão do modelo TDR surgiu do projeto de preservação digital da Cornell University, sugerindo que o modelo a ser seguido deve sair do entendimento de dois documentos-chave: o que propõe os atributos de um repositório digital confiável (TDR) da RLG/OCLC na implementação tecnológica, e da proposta do modelo de referência OAIS/SAAI para a estrutura do contexto organizacional. A representação dessa integração de propostas está no diagrama elaborado por Nancy Y. McGovern (2007) como mostra a Figura 19.

Usando o diagrama desenvolvido pela University of Cornell, McGovern aborda a questão da fronteira chamada de *Digital Archives Border* entre os primeiros dois atributos e o restante, para esclarecer sua importância naqueles casos em que

uma instituição mantém mais de um repositório. Os primeiros dois atributos se aplicariam a todos os repositórios na instituição, já que mais de uma organização pode chegar a gerenciar apenas um repositório (como no caso de um consórcio).

Figura 19 - Diagrama da integração dos dois pilares da preservação digital baseado no modelo TDR da RLG/OCLC
Fonte: McGovern (2007)

Desde 2003 a Cornell University vem criando parcerias e repassando para a comunidade sua experiência com a aplicação desses princípios, colocando tutoriais *on-line* e cursos para gestores de projetos de preservação digital. A Cornell University foi uma das primeiras instituições de ensino que utilizaram as duas grandes propostas para repositórios confiáveis; com elas foi estabelecida sua política de preservação digital e iniciados alguns dos seus projetos (CORNELL

UNIVERSITY LIBRARY, 2004). Também nessa universidade o trabalho seguiu alguns princípios norteadores:

a) focalizar as ações na redução de riscos;
b) proteger as versões arquiváveis;
c) entender os componentes dos objetos digitais;
d) reconhecer algum tipo de perda como inevitável;
e) estabelecer um ciclo de planejamento e manutenção.

Após a definição desses princípios, foi detectada a necessidade do trabalho colaborativo, ou seja, uma única instituição não conseguiria atender todas suas necessidade de preservação. Todos os projetos e iniciativas mapeados buscam a realização da sua intenção original que é comum para todos eles: preservar a produção intelectual e cultural das instituições e garantir o acesso permanente à informação (Apêndice C).

A possibilidade de se criar um modelo genérico de serviços de preservação digital para repositórios institucionais usando o modelo de referência OAIS é uma das necessidades nas iniciativas públicas de projetos e programas colaborativos. Alguns dos projetos começam a elaborar guias e manuais para auxiliar no processo de inserção de dados (*ingest process*), a estimular o depósito de arquivos em formatos-padrão para diminuir custos operacionais de longo prazo e a recomendar melhores práticas a serem implementadas.

No estágio inicial do projeto *Securing a Hybrid Environment for Research Preservation and Access* (SHERPA) foi realizada uma revisão crítica do modelo de referência OAIS. Usando o modelo para identificar os direitos e as responsabilidades executadas pelo repositório institucional, foram delineados os requisitos mandatórios e as entidades funcionais dentro do projeto de serviço de preservação distribuída do SHERPA (Figura 20).

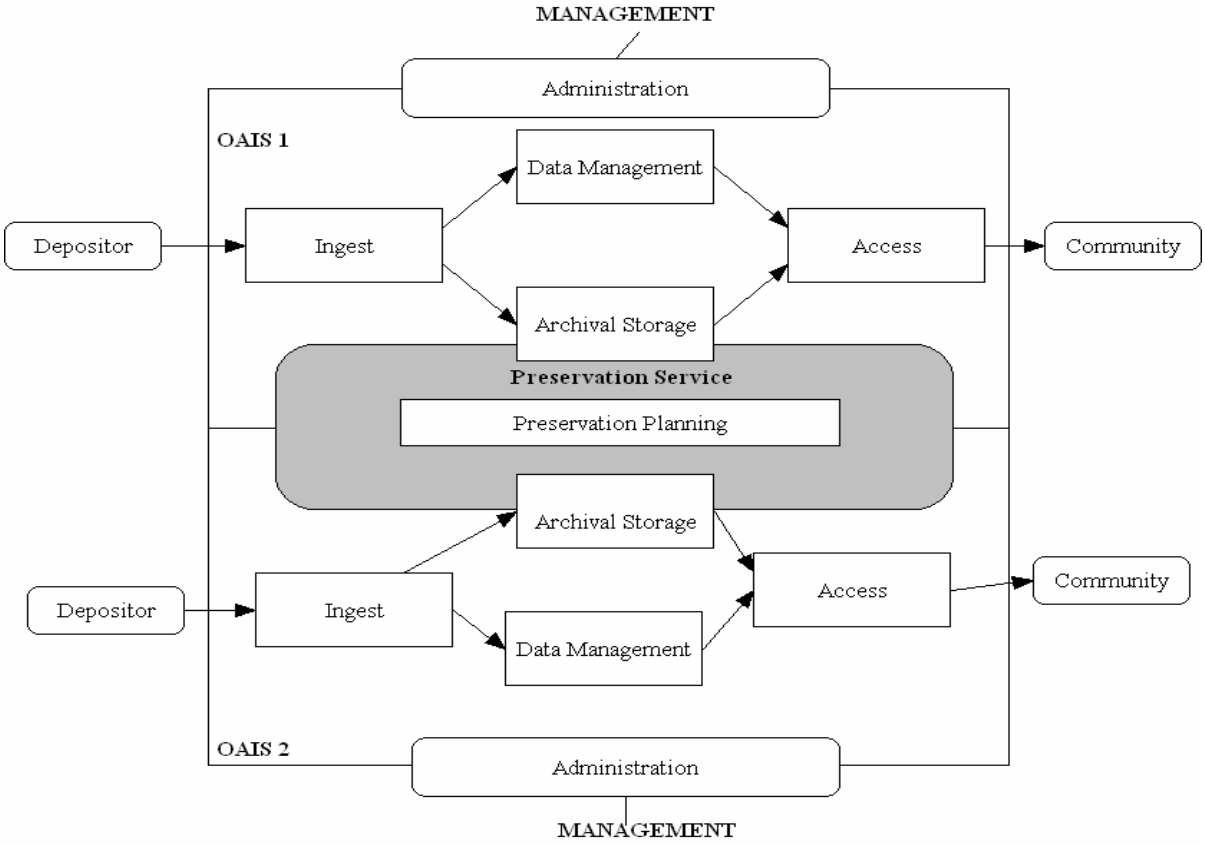


Figura 20 – Modelo funcional OAIS no projeto SHERPA

Projetando já um possível processo de auditoria para atribuir/certificar a concordância com o modelo OAIS, as primeiras ações desenvolvidas pelos responsáveis pelo projeto foram:

a) mapear as seis entidades de um repositório que esteja em concordância com o OAIS (inserção, armazenamento arquivístico, administração, gerenciamento de dados e acesso) dentro de uma estrutura existente;
b) garantir que a terminologia de um domínio específico pode ser mapeada por um equivalente OAIS.

O projeto SHERPA mostrou que a auditoria pode começar no processo de inserção dos dados, já que os sistemas de repositórios digitais podem ser divididos em dois grupos, aqueles em que o autor ou editor deposita diretamente o conteúdo (ex.: DSpace), e os que usam algum tipo de coleta desde as páginas a repositórios na *web* (ex. LOCKSS). Os dois processos não são imunes a algum tipo de ameaça e a auditoria pode confirmar a autenticidade do conteúdo inserido (ROSENTHAL *et al*., 2005).

2.7.1 Atributos de repositórios digitais

Os atributos e responsabilidade dos repositórios digitais em uso atualmente no exterior estão sendo definidos dentro do contexto dos desafios relacionados com a habilidade das instituições de integrar o gerenciamento de materiais digitais na sua estrutura organizacional. Outra característica desses centros é ter entre seus objetivos a pesquisa e o desenvolvimento das melhores práticas para maximizar os benefícios das novas tecnologias.

O reconhecimento da importância dos atributos de preservação digital e as políticas e procedimentos para sua aplicação constituem a primeira linha de ação no estabelecimento de sistemas de gerenciamento de informação digital.

Apesar do pouco tempo de implementados, os atributos de preservação digital podem ser acompanhados já no processo de escolha das ferramentas e formação dos repositórios digitais (CatalysIT, 2006; WHEATLEY, 2003) Com base na aplicação dos atributos TDR e do modelo OAIS/SAAI, os gestores desses sistemas de gerenciamento de informação digital podem propor algumas ações técnicas e organizacionais;

1) avaliar o funcionamento das primeiras migrações para novos formatos;
2) analisar os indicadores de desempenho dos repositórios e ferramentas adotadas;
3) definir políticas de seleção e depósito e retenção de documentos;
4) definir os perfis dos responsáveis e tipo de aperfeiçoamento para cumprir eficientemente as tarefas designadas.

Para efetivar as duas primeiras ações, alguns atributos técnicos de preservação digital têm sido usados, tais como:

1) segurança;
2) interoperabilidade;
3) qualidade e capacidade de configuração das ferramentas no *workflow;*
4) internacionalização – interfaces multilíngües;

5) licença para *software* livre.

Para as duas últimas ações, duas responsabilidades orientadas para os aspectos organizacionais são as mais indicadas: facilitar a configuração segundo o tipo de usuário e ter o suporte da comunidade.

2.7.2 Certificação de repositórios digitais

Em países como os Estados Unidos, Inglaterra e Alemanha já estão em funcionamento iniciativas de auditoria de arquivos digitais. Instituições nos Estados Unidos como o Research Library Group (RLG) e o National Archives and Records Administration (NARA) que criaram um grupo de trabalho sobre certificação de repositórios digitais e na Inglaterra como o Center for Research Libraries (CRL) com seu projeto de auditoria e certificação de arquivos digitais, estão interessadas em produzir um conjunto de princípios de auditoria em comum. São iniciativas que estão focalizando os benefícios e as ferramentas necessárias para a autovalorização e auditorias terceirizadas.

Sobre a capacidade de auditoria e certificação, existem iniciativas em funcionamento que vêm proporcionando ferramentas para auto-avaliação, mas as organizações precisam de meios para participar de auditorias externas. Processos de certificação podem criar maior padronização e credibilidade dos arquivos digitais que vão ao encontro das necessidades das bibliotecas e seus usuários. Os exemplos locais e as lições aprendidas estão contribuindo nesse sentido.

Com o reconhecimento da importância das políticas de preservação, ficou implícito o papel que elas têm dentro dos requisitos de evidência necessários para criar mecanismos de auditoria e certificação dos documentos e dos repositórios digitais. Em 2006 a RLG/NARA publicou o documento *Audit Checklist for Certifying Digital Repositories,* e no ano seguinte o Center for Research Libraries CRL lançou o *Trustworthy Repositories Audit & Certification: Criteria and Checklist (*TRAC), demostrando que as auditorias identificam realmente os pontos fortes e fracos nos programas de preservação digital, e como elas podem ajudar a definir planos de desenvolvimento que gradativamente cumpram com o conjunto de critérios definidos para os repositórios digitais confiáveis (McGOVERN, 2007).

A *checklist* da RLG/NARA define o conjunto de políticas gerenciais que são organizadas como critérios de preservação digital pela instituição, as funções do repositório, os processos e procedimentos, a comunidade alvo a usabilidade da informação e as tecnologias e a infra-estrutura técnica (RLG/ NARA, 2006).

A adoção de ferramentas de preservação digital que podem ser personalizadas, intercambiáveis e adaptadas aos *workflows* nos repositórios digitais está levando as instituições a abrir espaços para integrar requisitos de auditoria e de medição, dos pontos fracos e fortes dos programas de preservação digital.

Entre as conclusões a que chegaram os pesquisadores do RLG-NARA, no seu documento *Criteria for Measuring Audit Checklist for Certifying Digital Repositories,* está a de que as ferramentas usadas para fazer auditoria dos repositórios precisam ser desenvolvidas pelos próprios executores dos projetos de preservação digital.

O estabelecimento de programas de certificação e de critérios para serem usados em auditorias é uma necessidade identificada na comunidade da ciência da informação que lida com repositórios institucionais e que precisa de modelos de certificação dessas ferramentas (LYNCH 2003).

Algumas iniciativas que usam ferramentas como JHOVE, DROID e XENA estão apresentando, além da funcionalidade de preservar os objetos digitais na sua integração dessas ferramentas aos *softwares* de repositórios digitais, elementos que propiciam a comparação dos limites e capacidades das organizações no cumprimento dos requisitos de preservação apontados pelos modelos TDR e OAIS.

Projetos como o *Network of Expertise in Long-Term Storage of Digital Resources* (Nestor) da University Library of Humboldt-University Berlin, que desenvolveu seu “Catálogo de Critérios para Repositórios Digitais” são exemplos do crescente movimento para o desenvolvimento de parâmetros para medir a qualidade o confiabilidade de um repositório. O grupo de trabalho do Nestor tem entre seus objetivos “[...] formular critérios para repositórios digitais confiáveis e recomendações para procedimentos de certificação de repositórios digitais [...]” (DOBRATZ; SCHOGER; STRATHMANN, 2007 s/p.).

A complexidade dos sistemas de preservação digital em repositórios confiáveis foi apontada pelo grupo de trabalho do projeto Nestor em 2004; ele destacou a importância de abordar todo o processo gerencial de informação digital das organizações onde se deseja que a confiabilidade seja confirmada (NESTOR,

2006). A partir daí, todo o ambiente da preservação digital começou a ser tomado em consideração, assim como a recomendação de que os resultados da auditoria fossem comunicados de forma transparente ao público para gerar mais confiança.

Propostas como a do catálogo do projeto Nestor estão sendo usadas como instrumentos para auto-avaliação das etapas de desenvolvimento de projetos de repositórios digitais confiáveis. Sua aplicabilidade internacional tem suas limitações por razões geopolíticas, mas é levada em consideração internacionalmente para o estabelecimento de um processo formal de certificação ISO.

Em 2007 com o lançamento do "*checklist" da CRL Criteria for Measuring Trustworthiness of Digital Repositories and Archives: an Audit & Certification Checklist*, representantes de projetos de vários países estão construindo uma colaboração formal para auditoria e certificação de repositórios de preservação digital (CRL, 2007).

Tanto o Grupo de Trabalho do projeto NESTOR na Alemanha, quanto o DCC na Inglaterra e o CRL nos Estados Unidos coincidem nos princípios básicos para a aplicação de critérios institucionais de preservação digital, como segue:

1) Documentação (evidência): as metas, os conceitos, especificações e implementações de preservação digital de um repositório deverão ser documentadas adequadamente. A avaliação inicial do repositório como um todo baseada na documentação pode prevenir erros e implementações inapropriadas.
2) Transparência: a publicação da documentação leva à transparência. Segundo a CRL, "[...] apenas o repositório que expõe seu *design*, especificações, práticas, políticas e procedimentos para análise de riscos pode ser considerado confiável [...]" (CENTER OF RESEARCH LIBRARY, 2007, s/p).
3) Proporcionalidade: nenhum padrão ou norma deve ser tomado como único e absoluto para avaliação de todos os aspectos de um repositório digital, mas ela deve ser realizada baseada nos objetivos e tarefas aplicadas à preservação digital.
4) Mensurabilidade: apenas para alguns casos, devido aos aspectos de temporalidade dos processos de preservação, algumas formas de controle não são viáveis, mas a instituição deverá informar os

indicadores para medir o grau de confiabilidade, segundo seu nível de transparência.

As conclusões a que esses projetos chegaram são de que a certificação dos repositórios envolve mais do que a aplicação de critérios resultantes desses princípios básicos. Ela deve prover ferramentas para o planejamento da auto-avaliação e formas de auditorias internas e externas. O contexto de ambas é geopolítico, mas o processo não varia significativamente.

A comunidade de desenvolvedores desses repositórios vem confirmando a importância do processo de auditoria com mais freqüência do que o de certificação, pois a auditoria permite que os desenvolvedores dos repositórios analisem e respondam de forma sistemática às carências e acertos na criação dos sistemas.

No final de 2008, discute-se a aplicação de procedimentos de auditoria automatizados de contextos específicos da preservação digital, validando políticas locais definidas na aplicação de *softwares* para repositórios (MOORE; SMITH, 2007). Assim, espera-se também, a participação mais efetiva dos repositórios com informações oficiais dos governos, a partir do momento em que o processo de padronização da auditoria e certificação dos repositórios digitais fique formalizado numa norma ISO.

Para os especialistas em preservação digital, um apoio veio ao serem formalizados os primeiros centros de curadoria digital em 2005. Duas instituições que trabalham nesse sentido são a *Digital Curation Centre* (DCC) na Inglaterra e a UNC Chapel Hill nos Estados Unidos. O DCC é um centro que oferece serviços e produtos para a comunidade que trabalha com a curadoria de materiais digitais, oferecendo apoio à comunidade para trabalhar em rede e facilitar formas de acrescentar valor aos conteúdos digitais. Outras instituições estão discutindo a criação de programas de pós-graduação dentro das comunidades de ciência da informação, biblioteconomia, arquivologia e museologia, para capacitar seus estudantes nos trabalhos na área da curadoria da informação digital.

2.8 CONCLUSÕES DA REVISÃO DE LITERATURA

A metodologia de organização e descrição dos objetos digitais deve observar a natureza da informação a ser preservada. Isso é premissa básica para que os critérios de preservação digital sejam propostos. Porém, o conceito geral e a urgência estão presentes na ciência da informação, em suas disciplinas, demandando a criação de sistemas nos quais os recursos humanos, tecnológicos e procedimentais estejam em alinhamento com o estado-da-arte acerca das políticas, atributos e critérios de preservação digital.

Foi apresentado o estado atual das atividades de preservação digital da informação científica em países como os Estados Unidos, Austrália, Nova Zelândia, Holanda, Inglaterra, Alemanha e Canadá. Trata-se de estratégias que pretendem garantir a inalterabilidade dos registros digitais. Todos os métodos mencionados procuram evitar o risco de criar barreiras para um uso pleno dos recursos no futuro. Para resolver esse problema, foram elaborados e testados nos últimos anos esquemas conceituais que permitem identificar as partes integrantes do processo de preservação digital, assim como a sua padronização em modelos de referência para iniciativas de repositórios digitais (CCSDS, 2002).

Uma percepção duradoura da preservação digital precisa abarcar várias gerações de sistemas e tecnologias e unir as mudanças organizacionais com as necessidades de atualização dos responsáveis pelas coleções digitais. As perspectivas aqui citadas têm como respaldo especialistas e organizações, além de tecnologias assimiladas por comunidades de usuários. No exterior, os gerentes dos sistemas de preservação digital estão se adequando aos delineamentos dados pelos modelos de repositórios confiáveis e de referência existentes.

A questão da preservação digital possui várias frentes de desenvolvimento. Das políticas às técnicas, um grande conjunto de pontos críticos é encontrado e, por isso, os profissionais interessados na manutenção das informações sobre sua responsabilidade devem entender e se envolver com a elaboração das respostas. O interesse comum dos gestores dos mais diversos tipos de acervos em conhecer a preservação digital é o motivo que leva à pesquisa e o conseqüente desenvolvimento de soluções de aplicação na realidade.

Algumas leis e padrões relacionados com o ciclo de vida da informação em formato digital começam a surgir, permitindo a formulação de políticas de preservação digital estruturadas a partir da natureza da informação a ser guardada de longo prazo. Modelos conceituais estão sendo utilizados de forma a servir de base para políticas adaptadas a diferentes acervos por profissionais de formação e atividades diversificadas (arquivistas, bibliotecários, editores, museólogos e curadores). Os estudos sobre a preservação digital têm estabelecido que a imediata implementação de políticas de preservação digital é a forma mais efetiva de garantir o armazenamento e uso de recursos de informação por longos períodos de tempo.

A proposta de criação de sistemas de preservação digital distribuídos, que sigam modelos econômicos sustentáveis está partindo de projetos de preservação de publicações eletrônicas, que estão extraindo informações para bibliotecários e editores sobre a necessidade de confiar nos parceiros no campo digital, aspecto já mencionado, desde 1996, pelos primeiros especialistas a tratar do tema.

No caso específico da informação científica e tecnológica, os repositórios digitais são semelhantes em algumas características básicas no que se relaciona à preservação digital. Um exemplo é o *DSpace Institutional Digital Repository System* (projeto colaborativo da MIT Libraries e a Hewlett-Packard) orientado à preservação digital (JAMES *et al*., 2003). O DSpace como programa livre, ao ser adotado pelas organizações, transfere a estas a responsabilidade e os custos com as atividades de preservação (NIXON, 2002). Isto promove uma nova postura na comunidade, que é garantir efetivamente o acesso permanente à informação.

Várias pesquisas vêm centrando-se no modelo de preservação digital dos repositórios digitais, enfocando a necessidade de atuais e futuros usuários de contar com materiais autênticos e certificados por instituições reconhecidas. A descrição em metadados, de todos os detalhes que expressem a história de criação de um objeto digital está sendo considerada, tanto na área da arquivologia como na de ciência da informação, uma metodologia que pode garantir a “originalidade” de um registro eletrônico.

Os repositórios digitais estão começando a definir os níveis de serviços de preservação que podem oferecer, dependendo do conteúdo e do tipo de objeto. Para alguns conteúdos o repositório pode comprometer-se a chegar no nível da preservação dos *bits*, garantindo retornar ao autor do depósito a seqüência de *bits* sob demanda. Para outros registros, alguns sistemas de repositórios podem aplicar

estratégias de preservação como a migração ou até trabalhar em parceria com outros repositórios usuários das mesmas ferramentas.

Mas a falta de políticas de preservação na maioria dos projetos de repositórios digitais sugere a carência de conhecimentos técnicos sobre a importância das estratégias de preservação de materiais digitais existentes. A literatura da área mostra que a maioria dos projetos está nos seus primeiros estágios e apenas metade deles relata seguir alguma forma de modelo de negócios e processos de auditoria planejados (SMITH, 2008).

Essa lacuna informacional por parte dos responsáveis pelas políticas de implementação de informação digital precisa ser destacada. A segunda lacuna que deve ser mencionada diz respeito à importância da existência de bibliotecas digitais e de centros especializados no estabelecimento e monitoração de estratégias de preservação digital bem-sucedidas.

A amplitude do problema tratado conduziu a pesquisa bibliográfica ao momento do questionamento sobre a transformação de atributos e responsabilidade de preservação digital para critérios básicos formulados nas maiores iniciativas de pesquisa da área.

Embora o contexto atual dos trabalhos sobre auditoria e certificação de repositórios de preservação não proporcionem ainda elementos suficientes para perceber os benefícios obtidos na aplicação dos critérios de preservação digital adotados, espera-se que a definição de normas de auditoria e certificação seja reconhecida internacionalmente e que sua assimilação geopolítica mantenha as características interoperáveis dos sistemas.

As propostas de critérios de gestão de preservação digital da informação técnico-científica partiram de um entendimento sobre a realidade, gerado nas observações e análises dos aspectos fundamentais das práticas realizadas com esse objetivo. Espera-se que a ampla gama de condições sob as quais esses critérios sejam apropriados possua alta probabilidade de aplicação.

Com base na literatura analisada, pode-se afirmar que os critérios a serem propostos neste trabalho expressarão e justificarão um método de abordagem para uma realidade concreta, neste caso, aquela das instituições de pesquisa no Brasil.

Padrões, normas e práticas de preservação são adotados, quando existe consciência e interesse por parte dos responsáveis pelos acervos em desenvolver recursos que sejam permanentemente acessados. Em nível nacional, a definição

das políticas, obrigações e metodologias mais apropriadas para a preservação dos documentos eletrônicos deve levar em consideração a implementação de pacotes de *software* livres a fim de verificar se atendem às necessidades das instituições de ensino e pesquisa e se estão em concordância com os já testados padrões internacionais, que promovem o arquivamento digital da produção científica de longo prazo (UNESCO, 2007).

O trabalho exposto motivou o estudo bibliométrico sobre a produção científica da área (ver detalhes no item 3.3.1) e preparou a indagação sobre as expectativas e necessidade dos responsáveis pelos acervos de informação científica no Brasil. A realidade a ser detectada levou a uma proposta dos critérios que devem conduzir as implementações de sistemas de preservação digital no país.

3 METODOLOGIA

O método científico pode ser definido como o estudo de um fenômeno observável para fins de conhecimento ou de predição, descrevendo sua estrutura e funcionamento. As pesquisas que enfatizam os processos e os significados não medidos rigorosamente são denominadas qualitativas; elas abordam a relação entre o pesquisador e seu objeto de estudo e a natureza do contexto social.

> "A pesquisa qualitativa é um processo de investigação que obtém dados do contexto onde os eventos ocorrem, em uma tentativa para descobri-los, como uma forma para determinar os processos nos quais os eventos estão localizados e as perspectivas dos indivíduos que deles participam, utilizando a indução para derivar as possíveis explicações baseadas nos fenômenos observados." (GORMAN; CLAYTON, 1997 p. 23).

Para Silverman (2000), as metodologias na pesquisa social precisam de um planejamento cuidadoso e podem ser definidas de forma muito geral (qualitativa ou quantitativa) ou muito particular (por ex. análise de discurso, *grounded theory*).

Para Carlin (2003), desde seu início, o campo da biblioteconomia apropriou-se de instrumentos metodológicos de investigação e teóricos de outras áreas das ciências sociais, com exceção da bibliometria ou análise de citações.

Na ciência da informação, as ações relacionadas com a preservação digital estão associadas ao contexto dos problemas de informação que uma sociedade enfrenta no momento em que decide armazenar e dar acesso ao acervo considerado importante para ser preservado. Preservar, no ambiente digital, significa permitir a existência de longo prazo de objetos que contêm informação relevante.

Os levantamentos bibliográficos e buscas em recursos de informação *on-line* detectaram a ausência de argumentos teóricos ou abordagens metodológicas no estudo da preservação digital. Podemos afirmar que não existe uma interpretação única ou teorias que expliquem o tipo de fenômeno que está sendo produzido pelas ações de preservação digital no mundo.

Considerando que a temática da preservação digital encontra-se na fase da construção de hipóteses mais precisas e que os estudos atuais começam a elucidar questões complexas sobre sua natureza, esta pesquisa pretendeu prover subsídios e elucidar conceitos associados à preservação digital.

Na literatura da área, os critérios de preservação digital estão sendo apontados como elementos integrantes dos sistemas de gestão de preservação digital, divididos em três aspectos: o organizacional, o tecnológico e o dos recursos necessários. O aspecto organizacional está representado pela observância do modelo de referência OAIS na construção de sistema de gestão da preservação digital. Para tratar do aspecto tecnológico, está sendo seguido o modelo dos TDR na estruturação de repositórios de preservação confiáveis. E, para o terceiro aspecto, estão os esforços relacionados à identificação de alguns critérios que contemplem os custos e a formação de recursos humanos (GRANGER, 2002; LAVOIE; DEMPSEY, 2004; LYNCH, 2003; RLG, 2002).

A principal preocupação mencionada nesses estudos é que os profissionais da informação responsáveis pelos documentos digitais estabeleçam seus próprios critérios na criação de sistemas de gestão de preservação digital em suas instituições. Esta necessidade pode ser detectada nas pesquisas que têm sido realizadas em instituições internacionais (FRIEDLANDER, 2002; HODGE; FRANGAKIS, 2004; McGOVERN, 2007).

Na revisão da literatura foram citados os critérios apresentados nas principais iniciativas de preservação digital na última década. Ao considerar os tópicos tratados como reveladores dos critérios básicos de um sistema de gestão de preservação digital da informação técnico-científica e de sua possível adequação ou não à realidade brasileira, nossa pesquisa assumiu os seguintes pressupostos:

1) compreensão da gestão dos documentos digitais como um contínuo integrado de processos que envolvem todo seu ciclo de vida documental;
2) existência de graus de completeza/completude dos sistemas organizacionais de gestão da preservação digital;
3) possibilidades de adequação dos critérios para a preservação digital;
4) percepção da necessidade de preservação digital como determinante nos critérios que definem o sistema de gestão de preservação digital a ser adotado.

No caso específico dos problemas relacionados com a contextualização da realidade da preservação digital no Brasil e a formulação de critérios ainda não explorados, a pesquisa aponta para a unificação de abordagens qualitativas e quantitativas. Seguindo a proposta de Silverman (2000) sobre a caracterização dos

métodos de pesquisa, as técnicas utilizadas nos projetos com metodologias quantitativas podem ter apoio dos métodos qualitativos.

Neste trabalho foram utilizadas as duas abordagens, na tentativa de explorar dados bibliográficos e das percepções dos atores envolvidos com a preservação digital de informação científica no Brasil. Esta decisão foi adotada por se tratar de uma área que não tem constituído objeto de estudo no país e por não haver sido produzido conhecimento anterior. Como parte de primeiro reconhecimento do objeto de estudo, partimos para o encontro das percepções dos membros da comunidade investigada que detinham algum conhecimento sobre o tema, para desse modo, instrumentalizar uma posterior abordagem alternativa.

A pesquisa elaborada adotou como tema norteador a identificação dos critérios categorizados como indicadores de preservação digital da informação científica e tecnológica, representados na literatura científica especializada, e como esses critérios são contemplados nas unidades de pesquisa do Ministério da Ciência e Tecnologia do Brasil.

Seguindo a tipologia da pesquisa científica de Abramo (1979), este trabalho sustenta-se na caracterização de uma pesquisa interdisciplinar, porque envolve vários setores do conhecimento, que usa como técnicas a entrevista e a observação direta (LAKATOS & MARCONI, 1991) e no nível da interpretação dos dados, como uma pesquisa exploratória e descritiva. A estratégia metodológica adotada compreendeu quatro etapas distintas:

1) levantamento e análise bibliométrica da produção científica sobre preservação digital de informação científica e tecnológica em artigos, teses, dissertações e livros no período de 1996 a 2007;
2) contextualização da realidade da preservação digital nas unidades de pesquisa do Ministério da Ciência e Tecnologia, resultado de levantamentos, observação de práticas e análises de documentos dessas instituições;
3) realização de uma análise etnográfica de conteúdo publicado sobre a preservação digital no Brasil. Coleta de documentos, palestras, cursos, artigos e publicações, observando os padrões emergentes, os temas, ênfases e processos mais freqüentemente citados;
4) estruturação de um conjunto de critérios de preservação digital da informação técnico-científica.

Foram consideradas como material relevante as informações sobre padrões e práticas de preservação digital de projetos que não estavam relacionados com a preservação de informação científica e tecnológica, no intuito de identificar o estado-da-arte da pesquisa sobre o tema. A investigação pretendeu levantar as questões mais significativas para os envolvidos no processo de preservação digital estudando-as como experiências reais e com significados construídos socialmente. Os documentos técnicos selecionados nos contextos pesquisados foram entendidos como representações das percepções sobre a preservação digital dentro do processo analítico de conteúdo.

3.1 PROBLEMA INVESTIGADO

A pesquisa pretendeu responder à seguinte indagação: quais critérios de preservação digital, usados na construção sistemas de gestão da informação científica e tecnológica reconhecidos internacionalmente, podem ser adotados pelas instituições de pesquisa do Ministério da Ciência e Tecnologia do Brasil?

As respostas a essa questão oferecerão um conjunto de critérios de preservação digital apropriados para subsidiar futuras iniciativas relacionadas à manutenção e controle de acervos digitais relevantes para o desenvolvimento científico e tecnológico do país, assim como conhecimento sobre a necessidade de critérios de preservação digital no contexto do gerenciamento da informação científica e tecnológica.

3.2 Universo de pesquisa e amostra

No Brasil, a pesquisa científica apoiada pelo governo nasceu na década de 1930, como produto de uma reforma administrativa realizada durante a ditadura Vargas. Já nos anos 1950 foi criado o Conselho Nacional de Pesquisa (CNPq atualmente Conselho Nacional de Desenvolvimento Científico e Tecnológico) como

resposta às ameaças do uso da energia nuclear e à necessidade da segurança nacional do Estado. As demandas das universidades e centros de pesquisa por mais recursos na década de 1960 promoveram a criação de instituições de fomento como a Finep, para diminuir a burocracia no repasse dos recursos públicos para o desenvolvimento científico e tecnológico (DIAS, 2002).

A partir dos anos 1970 se consolida a formação da infra-estrutura governamental de pesquisa, com a criação de diversos centros de excelência na pesquisa científica, a atividade científica toma uma posição hegemônica na vida intelectual do país, e a criação do MCT conferia, sem dúvida, um *status* político à atividade (BURGOS, 1999).

Até o ano de 2000, as unidades de pesquisa do MCT encontravam-se distribuídas em três grandes grupos: quatro subordinadas diretamente ao Ministério; dez, ao CNPq; e cinco à Comissão Nacional de Energia Nuclear – CNEN, do extinto Ministério Extraordinário de Projetos Especiais.

Em julho de 2008, a estrutura organizacional do Ministério da Ciência e Tecnologia estava dividida em quatro categorias de instituições, unidades de pesquisa, organizações sociais, entidades vinculadas e empresas estatais, cada uma com suas características organizacionais:

a) unidades de pesquisa do MCT: executam programas e projetos visando ao fortalecimento da pesquisa científica e tecnológica brasileira. Essas atividades são acompanhadas pela Subsecretaria de Coordenação das Unidades de Pesquisa;
b) organizações sociais: representam instituições que abrangem laboratórios ou institutos especiais, com os quais o MCT mantém Contratos de Gestão. As atividades das organizações sociais são promovidas, supervisionadas e avaliadas pela Subsecretaria de Coordenação das Unidades de Pesquisa, para cumprir a lei que outorga essa função ao MCT (Lei nº 9.637 de 15 de maio de 1998);
c) entidades vinculadas: estão subdivididas em autarquias, fundações e empresas públicas;
d) empresas estatais.

Cronologicamente, as unidades de pesquisa têm origem variada, mas a integração da maioria delas ao Ministério aconteceu na última década do século XX. Essa integração foi realizada por decreto oficial, algumas como autarquias federais,

mas a caracterização de todas elas dentro do Ministério mudou em diferentes épocas, até chegar ao estado atual (Quadro 8).

Categoria	Instituição	Localização	Fundação
UNIDADES DE PESQUISA	Instituto Nacional de Pesquisas Espaciais (Inpe)	São José dos Campos/SP	1962
	Laboratório Nacional de Astrofísica (LNA)	Itajubá/MG	1985
	Observatório Nacional (ON)	Rio de Janeiro/RJ	1827
	Centro Tecnológico Mineral (Cetem)	Rio de Janeiro/RJ	1978
	Instituto Nacional de Tecnologia (INT)	Rio de Janeiro/RJ	1921
	Centro Brasileiro de Pesquisas Físicas (CBPF)	Rio de Janeiro/RJ	1949
	Museu de Astronomia e Ciências Afins (MAST)	Rio de Janeiro/RJ	1985
	Laboratório Nacional de Computação Científica (LNCC)	Petrópolis/RJ	1980
	Centro de Pesquisas Renato Archer (CenPRA)	Campinas/SP	1982
	Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)	Brasília/DF	1954
	Instituto Nacional do Semi-Árido (Insa)	João Pessoa/PB	2004
	Instituto Nacional de Pesquisas da Amazônia (Inpa)	Manaus/AM	1954
	Museu Paraense Emílio Goeldi (MPEG)	Belém/PA	1866
ORGANIZAÇÕES SOCIAIS (Lei nº 9.637 de 15 de maio de 1998)	Associação Brasileira de Tecnologia Luz Síncroton (LNLS-ABTuS)	Campinas/SP	1987
	Instituto de Desenvolvimento Sustentável Mamirauá (IDSM)	Tefé/AM	1999
	Associação Instituto Nacional de Matemática Pura e Aplicada (Impa)	Rio de Janeiro/RJ	1953
	Associação Rede Nacional de Ensino e Pesquisa (RNP)	Brasília/DF Rio de Janeiro/RJ	1989
	Centro de Gestão e Estudos Estratégicos (Cgee)	Brasília/DF	2001
ENTIDADES VINCULADAS	**AUTARQUIAS**		
	Agência Espacial Brasileira (AEB)	Brasília/DF	1994
	Comissão Nacional de Energia Nuclear (Cnen)	Rio de Janeiro/RJ	1957
	FUNDAÇÃO		
	Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)	Brasília/DF	1951
	EMPRESA PÚBLICA		
	Financiadora de Estudos e Projetos (Finep)	Rio de Janeiro/RJ	1967
EMPRESAS ESTATAIS	Indústrias Nucleares do Brasil (INB)	Rio de Janeiro/RJ	1988
	Nuclebrás Equipamentos Pesados (Nuclep)	Itaguaí/RJ	1975

Quadro 8 – Classificação das unidades de pesquisa do Ministério da Ciência e Tecnologia

Cada uma das 24 instituições com as suas peculiaridades em relação ao tipo de produção científica, características das comunidades científicas, contexto histórico e alocação de recursos materiais e humanos é apresentada no Apêndice E.

Para fins desta pesquisa, a população investigada resumiu-se às unidades de pesquisa do Ministério da Ciência e Tecnologia. A amostra foi composta pelos centros de informação e bibliotecas existentes nessas organizações. Os respondentes foram selecionados com base no seu grau de envolvimento com o gerenciamento dos documentos digitais produzidos nessas instituições, e sua seleção não foi determinada pelas suas categorias profissionais.

Os setores de informação estudados realizam atividades relacionadas à informação científica e tecnológica em suportes digitais, especificamente, com a produção científica do ministério que foi transformada para formato digital e daquela que já nasceu digital.

Segundo os dados coletados no primeiro levantamento sobre os centros de informação e bibliotecas do MCT, mais tarde confirmados na pesquisa de campo, há 19 unidades de pesquisa que contam com esse tipo de serviço para seus funcionários e público externo. O Quadro 9 mostra a lista de unidades, o setor de informação e o total de responsáveis por esses serviços.

Unidade	Setor de Informação	Responsável	Total
CBPF	Biblioteca	Bibliotecária	01
CenPRA	Biblioteca/DCOG	Bibliotecária	01
Cetem	Biblioteca	Bibliotecária e Cientista da Informação	02
Cgee	Núcleo Centro de informações	Cientista da Informação	01
Cnen	Centro de Informações Nucleares	Bibliotecária	01
CNPq	Biblioteca “Lygia Portocarrero Velloso”	Bibliotecária	01
Finep	Serviço de Informação	Bibliotecárias	02
Ibict	Biblioteca	Bibliotecária	01
Impa	Biblioteca	Bibliotecárias	02
Inpa	Biblioteca	Bibliotecária	01
Inpe	Biblioteca	Bibliotecárias	02
INT	Biblioteca	Bibliotecárias	02
LNA	Biblioteca	Bibliotecária	01
LNCC	Biblioteca	Bibliotecária	01
LNLS-ABTLuS	Biblioteca – Departamento Científico	Bibliotecária	01
Mast	Biblioteca	Bibliotecárias	02
MPEG	Biblioteca “Domingos Soares Ferreira Penna”	Bibliotecária	01
ON	Biblioteca/DID	Bibliotecária	01
RNP	Biblioteca/DGI	Cientista da Informação	01

Quadro 9 – Local de preservação e acesso da informação científica em formato digital produzida nas unidades de pesquisa do MCT

Entre 2004 e 2007, a pesquisa identificou 25 responsáveis por acervos digitais lotados nos centros de informação e bibliotecas. A localização foi possível por intermédio de consultas, realizadas via internet, incluindo-se as pessoas que tinham endereços eletrônicos disponibilizados nas páginas do MCT (amostragem não probabilística). Também a indicação das pessoas que teriam algum grau de comprometimento com a preservação dos acervos em formato digital foi feita pelos funcionários, pesquisadores, bibliotecários, arquivistas e profissionais de informática entrevistados durante a visita a esses centros.

3.3 INSTRUMENTOS DE COLETA DE DADOS

Vários autores mencionam que a experiência e comportamentos humanos não são “explicáveis” por fatores externos, mas podem ser “compreendidos” por meio da empatia comunicativa com a visão subjetiva e as razões internas do outro (MAYRING, 2002). Esse entendimento precisa do emprego de um conjunto de instrumentos que determinam a forma de abordar os fatos ao longo de um percurso para alcançar um objetivo preestabelecido.

Para compreender qualitativamente as práticas reais e os métodos mais utilizados para preservar a informação técnico-científica em documentos digitais, aplicaram-se as seguintes técnicas de pesquisa: estudo bibliométrico, questionário, entrevista, observação de curta duração e análise etnográfica de conteúdo.

O esquema da pesquisa incluiu no primeiro estágio um levantamento bibliográfico que se prolongou durante todo o período da investigação. Em seguida, utilizou-se um questionário para mapeamento da realidade das unidades de pesquisa do MCT.

Determinada qual seria a amostra do estudo e o local das visitas, partiu-se para o levantamento de dados coletados diretamente com os responsáveis pelos acervos digitais dessas unidades. Na pesquisa de campo, além da realização de entrevistas e da observação de curta duração, reuniram-se dados documentais encontrados em arquivos, bancos de dados, relatórios e publicações oferecidas pelos próprios entrevistados. As experiências relatadas durante as entrevistas funcionaram como parâmetro para a informação, obtida pelo método da observação. As informações foram organizadas cronologicamente e incluídas na análise etnográfica de conteúdo.

3.3.1 Estudo bibliométrico

Após tratamento bibliométrico realizado na bibliografia consultada, identificaram-se os autores e os títulos de periódicos mais citados, os temas e as datas de publicação. Esse levantamento incluiu a organização dos dados em dez variáveis, a saber:

- ano;
- área;
- aspecto discutido;
- autor;
- país.
- periódico;
- proposta;
- sub-tema;
- tipo de autoria;
- tipo de publicação;

Essas variáveis foram submetidas a um tratamento estatístico descritivo (análise de freqüência). Pretendeu-se quantificar a produção individual dos principais autores em diferentes períodos e seu impacto no desenvolvimento da área.

A continuidade desse estudo fundamentalmente bibliométrico tornou-se possível avaliando a produção científica sobre o tema da preservação digital no Brasil e no exterior, considerando apenas os trabalhos citados revisão da literatura e nas listas de documentos de projetos e iniciativas da área publicadas até 2007 (ver Apêndice C)

3.3.2 Questionários

A técnica de coleta de dados por meio de questionário é utilizada quando se pretende atingir um número considerável de pessoas. O conjunto de questões formuladas deve refletir os objetivos do estudo. Para caracterizar as atividades de preservação digital identificadas nas Unidades de Pesquisa do MCT, durante o período de 2004 a 2007, tornou-se necessário localizar as informações oficiais sobre os acervos digitais nessas instituições. O levantamento incluiu as informações oficiais nos *sites* do Ministério e o envio por *e-mail* de um questionário para os responsáveis pelas informações disponibilizadas na internet.

Na primeira aproximação, remeteu-se via correio eletrônico, a todos os responsáveis pelas bibliotecas e centros de informação do Ministério de Ciência e Tecnologia, questionário elaborado em colaboração com a Biblioteca de Política em

Ciência e Tecnologia (BPCT) da Coordenação-Geral de Gestão e Inovação (CGGI) do próprio Ministério. O propósito foi conseguir levantar, com questões abertas sobre fatos (ações) e de opinião (atitudes), dados sobre as atividades relacionadas com a preservação dos seus acervos em formato eletrônico bem como identificar os responsáveis e as principais atividades relacionadas com acervos digitais. (ver Apêndice A).

3.3.3 Pesquisa de campo

O estabelecimento de relações constantes entre determinadas condições pode ser obtido com a pesquisa de campo (RUIZ, 1979). Para Minayo (1992), o trabalho de campo nas abordagens qualitativas compreende duas técnicas fundamentais: a entrevista e a observação. Elas se complementam, aumentando o controle dos dados nos estudos exploratórios (SELLTIZ *et al.*, 1975).

Na pesquisa de campo foram analisadas as ações do Ministério da Ciência e Tecnologia para implementar (desenvolver, efetivar, planejar e utilizar) sistemas de preservação digital. Com isso, a investigação esteve centrada em descobrir:

a) os objetivos e características das ações propostas no contexto do MCT com fins de preservação digital;
b) as infra-estruturas e recursos adotados no MCT para a construção de sistemas de preservação digital;
c) como essas infra-estruturas padronizam, armazenam e ajudam a gerenciar a informação científica e tecnológica a ser preservada digitalmente.

Na fase após o primeiro levantamento, procedeu-se a visitar as unidades de pesquisa e a entrevistar os responsáveis previamente identificados. Os registros obtidos ajudaram a formar o conjunto de dados oficiais sobre as coleções eletrônicas institucionais.

Conforme a viabilidade de deslocamento do pesquisador para algumas localidades (Belém, Petrópolis, Rio de Janeiro, Campinas, São José dos Campos e Itajubá), observaram-se as experiências de preservação digital em andamento,

registrando a natureza das necessidades de preservação digital expressas nas restrições situacionais de cada uma das unidades examinadas.

A realização da pesquisa de campo compreendeu a observação direta de curta duração (um a três dias) para coleta de dados complementares acerca da realidade das práticas e necessidades de preservação digital nas unidades de pesquisa. Foram coletados dados das opiniões dos atores identificados, examinando-se as percepções dos envolvidos sobre as necessidades e as práticas de preservação digital, descobrindo-se temas decorrentes, padrões emergentes e processos enfatizados e adotados.

3.3.4 Entrevistas

A apropriação de um método qualitativo para a compreensão das percepções dos envolvidos nas atividades de preservação digital contemplou a necessidade de entrevistas, como verificação e registro das narrativas históricas das memórias das fontes interessadas.

Descombe observa que a entrevista representa uma espécie de conversação elaborada de maneira sistêmica:

> As entrevistas são uma proposta atrativa para os pesquisadores que realizam algum projeto. Aparentemente, não envolvem muita parafernália técnica na coleta da informação – talvez um caderno e um gravador portátil – e a técnica básica depende da habilidade que os pesquisadores já têm para conduzir uma conversa. Não é necessário contar com ferramentas complexas ou passar tempo aprendendo novas habilidades: esta é uma receita particularmente tentadora. (DESCOMBE, 2003, p. 163).

Na fase exploratória, estabeleceram-se conversas orientadas com alguns dos responsáveis identificados pelas pessoas que responderam aos questionários enviados. O propósito era recolher informações mais precisas sobre as pessoas que realmente tinham algum grau de comprometimento com a preservação dos acervos em formato digital dessas instituições.

Durante a 56ª. Reunião Anual da SBPC, celebrada na cidade de Cuiabá, Mato Grosso, em julho de 2004, realizaram-se as primeiras entrevistas semi-estruturadas (Apêndice B) com alguns dos responsáveis pelos centros de informação, bibliotecas e arquivos das unidades de pesquisa do MCT. Tal atividade

dependeu da disponibilidade dos respondentes durante o evento. Os dados obtidos permitiram identificar os responsáveis pelas práticas de preservação digital existentes, os materiais e formatos mais freqüentemente usados, além de captar opiniões sobre as necessidades de preservação digital nas unidades de pesquisa.

Em seguida ao primeiro contato, fizeram-se visitas às unidades de pesquisa, aplicando-se entrevistas presenciais semi-estruturadas, registradas em fita magnética como recurso especial para coleta de dados, posteriormente transcritas de forma precisa para entender como a interação foi estabelecida (SILVERMAN, 2000).

Durante a fase de pesquisa de campo, utilizou-se o mesmo roteiro de entrevista em profundidade (FINK; KOSECOFF, 1985) usado com o primeiro grupo de respondentes, orientado para obter informações relacionadas com as atividades de preservação dos acervos digitais.

As perguntas foram divididas em três grupos. O primeiro relacionava a formação profissional e as características das tarefas realizadas pelos entrevistados, a saber:

a) profissão;
b) titulação máxima;
c) tempo de serviço;
d) cargo ocupado atualmente;
e) tempo no cargo atual.

No segundo grupo, as questões dirigiam-se para a descrição das atividades desenvolvidas pelos respondentes que envolvem algum tipo de tratamento técnico nos arquivos digitais da instituição:

a) Quais atividades científicas são realizadas pela sua instituição?
b) Que tipo de produção científica já está em formato digital?
c) Quem são os usuários dessa produção científica?
d) Quais atividades de preservação de documentos digitais estão sendo realizadas pela sua instituição?

O terceiro grupo de perguntas permitiu a coleta de opiniões sobre o material a ser preservado e seus responsáveis:

e) Qual produção científica deve ser preservada pela sua instituição?
f) Quem são os responsáveis por essas atividades de preservação?

Com elas, pretendeu-se detectar variáveis que não haviam sido contempladas no quadro conceitual preestabelecido pelo pesquisador, relacionadas às expectativas dos responsáveis pelas coleções no que se refere ao que deve ser feito para preservar seus documentos digitais.

Foram entrevistadas 86 pessoas entre aquelas identificadas no primeiro levantamento para o envio do questionário e as detectadas nas falas dos próprios respondentes durante a aplicação do primeiro grupo de entrevistas. No Quadro 10 visualiza-se a listagem completa de profissionais e pesquisadores contatados durante a pesquisa de campo.

Unidade	Setor	Cargo/Função	Nº Entrevistas
AEB	Arquivo Informática Comunicação Diretoria	Técnico Técnico Técnico Pesquisador	1 1 2 1
CBPF	Biblioteca Informática Comunicação	Bibliotecária Técnico Técnico	2 2 1
Ceitec	Comunicação	Pesquisador	1
CenPRA	Biblioteca Biblioteca Laboratório Comunicação	Bibliotecária Pesquisador Pesquisador Técnico	1 1 1 1
Cetem	Biblioteca Informática Comunicação Comunicação	Bibliotecária Cientista da Informação Pesquisador Técnico	1 1 1 1
Cgee	Centro de Inf. Centro de Inf. Informática Comunicação	Cientista da informação Técnico de informação Técnico de informática Técnico	1 2 1 1
Cnen	Biblioteca Centro de Informação Comunicação	Bibliotecária Pesquisador Técnico	1 1 1
CNPq	Biblioteca Centro da Memória Comunicação	Bibliotecária Pesquisador Técnico	1 1 1
Finep	Biblioteca Informática Comunicação	Bibliotecárias Técnico Técnico	2 1 1
Ibict	Biblioteca Informática Comunicação	Bibliotecária Técnico Técnico	1 1 1
IDSM	Serviço de Inf. Comunicação	Técnico de Informação Técnico	1 2
Impa	Biblioteca Comunicação	Bibliotecária Técnico	2 2
Inpa	Biblioteca Comunicação	Bibliotecária Técnico	1 2
Inpe	Biblioteca Depto. Científico Comunicação	Bibliotecária Pesquisador Pesquisador	3 1 1

Continua

Unidade	Setor	Cargo/Função	Nº Entrevistas
Insa	Comunicação Diretoria	Técnico Pesquisador	1 1
INT	Biblioteca Informática Comunicação	Bibliotecárias Técnico de informática Técnico	2 1 1
LNA	Biblioteca Biblioteca Informática Depto. Científico Comunicação	Bibliotecária Técnico de informação Pesquisador Pesquisador Técnico	1 1 1 1 1
LNCC	Biblioteca Informática Comunicação	Bibliotecária Técnico Técnico	1 1 1
LNLS-ABTLuS	Biblioteca Depto. Científico Comunicação	Bibliotecária Pesquisador Técnico	1 1 1
Mast	Biblioteca Informática Comunicação Arquivo	Bibliotecárias Técnico Pesquisador Técnico	2 1 1 2
Mpeg	Biblioteca Editoração	Bibliotecária Técnico	1 1
Nuclep	Comunicação Comunicação Informática	Pesquisadora Técnico Técnico	1 1 1
ON	Biblioteca/DID Comunicação	Bibliotecária Técnico	1 1
RNP	Biblioteca/DGI Informática Comunicação	Cientista da Informação Técnico Técnico	1 2 1

Quadro 10 – Listagem de respondentes das unidades de pesquisa do MCT entrevistados na fase exploratória e de pesquisa de campo

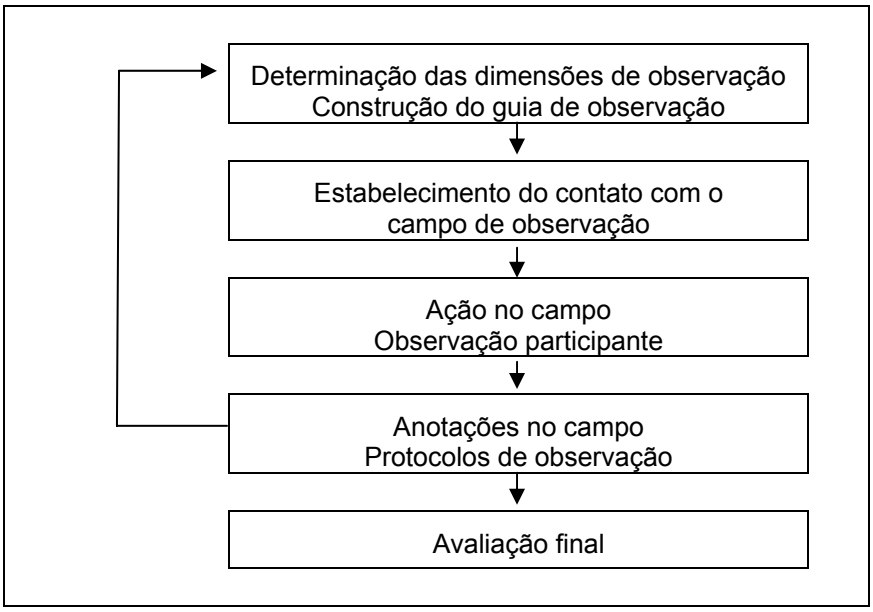
Os respondentes foram caracterizados como representantes de setores-chave das instituições procuradas, a saber: centro de informação, biblioteca, arquivo, informática, direção e departamentos de pesquisa. O critério de seleção desses respondentes decorreu de serem atores com algum grau de envolvimento formal na manutenção dos acervos digitais da instituição e cujas opiniões servissem para examinar as percepções sobre o tema e processos adotados.

Os dados obtidos com os entrevistados permitiram estabelecer um tipo de padrão no quesito necessidade de práticas de preservação digital não detectado nos questionários, e a definição de novos níveis de aprofundamento a serem alcançados por meio da observação e a análise etnográfica de conteúdo. O instrumento de coleta utilizado nas entrevistas está no Apêndice B.

3.3.5 Observação de curta duração

A coleta de dados por meio da observação oferece ao pesquisador a oportunidade de obter informação não apenas sobre o que os entrevistados asseguram que fazem ou pensam, mas de uma forma direta diante da realidade dos eventos descritos (DESCOMBE, 2003).

Para Mayring (2002), as dimensões da observação devem ser determinadas e concretizadas na construção de um guia de observação, estruturado segundo a formalização do contato estabelecido com os representantes do campo observado. Este plano de procedimento formulado por Mayring é constante e formalizado no que ele chama de protocolos de observação, que registram as ações empreendidas pelo pesquisador, até o momento da avaliação final dos dados coletados (Quadro 11).



Quadro 11 - Plano de procedimento de observação
Fonte: Mayring (2002, p. 132)

Para obter-se um diagnóstico sobre a situação da preservação digital no MCT, como complemento à técnica de entrevista, empregou-se o registro dos padrões de comportamento das pessoas por meio da observação direta de curta duração não participante (AAKER; KUMAR; DAY, 2004; GWATNEY, 2001). O objetivo foi levantar indícios relevantes sobre as práticas reais de preservação

digital, a fim de controlar o registro sistemático das ações executadas no tratamento dos documentos digitais e a caracterização das estratégias adotadas.

O critério considerado para a utilização desta técnica de observação foi o fato de ela não ser disruptiva e permitir a obtenção dos dados de forma semi-direta, segundo o comportamento dos sujeitos. Dessa forma, conseguiu-se obter uma imagem das práticas dos responsáveis pelos acervos digitais em determinado lugar, contexto e tempo, dependendo de uma sucessão de eventos.

4 ANÁLISE DOS DADOS

A organização dos dados foi resultado de um ciclo repetido de interações com as fontes bibliográficas e com os respondentes. A primeira foi determinada por dois objetivos específicos da pesquisa, que visavam caracterizar as práticas de preservação digital e resgatar os conceitos e aplicações usadas nas experiências nacionais e estrangeiras.

As interações com os respondentes por meio das entrevistas, a observação de curta duração buscaram identificar as necessidades de preservação digital da informação técnico-científica das unidades de pesquisa do MCT, assim como conduzir o estudo para uma abordagem mais qualitativa do tema, a aplicação da técnica de análise etnográfica de conteúdo. Dessa forma, foi possível desenvolver os critérios de preservação digital, propostos a partir de padrões encontrados nos dados coletados.

Segundo Pereira (2001, p. 67) “[...] a mensuração qualitativa é uma medida derivada, que não se realiza diretamente sobre o fenômeno de interesse, mas sobre as manifestações desse fenômeno [...]”. Com as três técnicas usadas procurou-se classificar individualmente os respondentes pelas suas funções e grupos de profissionais e descrever o conjunto de observações realizadas nos locais visitados.

As estratégias de análise mostraram como o esforço para sintetizar as informações coletadas precisou reduzir a dimensionalidade do fenômeno com fins de produzir conhecimento sobre o tema.

4.1. ANÁLISE BIBLIOMÉTRICA

O interesse na criação de indicadores bibliométricos sobre a produção científica vem acompanhando o crescimento da literatura publicada. No início do século XX apareceram as primeiras aplicações de métodos estatísticos à análise da literatura de áreas do conhecimento científico. Como uma nova prática estatística, e sem ser denominada ainda bibliometria, os resultados obtidos nos primeiros estudos proporcionaram um método de análise da prática científica (ALVARADO, 2007).

Um dos primeiros teóricos a definir o termo bibliometria, foi Pritchard como a “[...] aplicação de modelos matemáticos e estatísticos aos livros e a outros meios de comunicação escrita [...]” (PRITCHARD, 1969, p. 25). A técnica tem sido utilizada para identificar algumas características das publicações científicas, a produtividade, os paradigmas e as relações entre disciplinas científicas (BROOKES, 1973; IKPAAHINDI, 1985; POTTER, 2002).

Para Narin (1976), a bibliometria pode ser dividida em descritiva e avaliativa. A bibliometria avaliativa estaria dedicada a analisar os aspectos qualitativos da produção científica através da análise de citações e publicações. Já a bibliometria descritiva estuda as características específicas de distribuição geográfica, evolução temporal, entre outras. Narin também afirma que a contagem de artigos proporciona indicadores válidos da produção científica e o do desenvolvimento de temas relacionados e das instituições de origem.

Para fins desta pesquisa, a necessidade de entrevistas, narrativas históricas e memórias de fontes interessadas contemplou também, a apropriação de um método quantitativo para a compreensão das percepções dos envolvidos nas atividades de preservação digital. Os índices estatísticos utilizados serviram para observar o fenômeno da produção científica sobre preservação digital na variedade dos assuntos e na regularidade dos temas mais freqüentemente abordados.

A análise bibliométrica foi usada na identificação de ligações entre documentos e no mapeamento da literatura da área, fornecendo informações sobre o desenvolvimento de campos específicos e subsídios para formas de interpretação no contexto da realidade de instituições de pesquisa no Brasil.

4.1.1 Bibliometria descritiva da preservação digital

Para poder conhecer as abordagens teóricas e práticas que têm sido publicadas na última década, foi realizado um estudo bibliométrico descritivo, coletando dados de referências bibliográficas sobre o tema da preservação digital do período de 1996 até 2007. Na revisão da literatura foi constatada a ausência de um controle da produção bibliográfica sobre o tema na forma de bases de dados ou da

publicação de bibliografias da área. Existe apenas listagem de textos relevantes em *sites* de iniciativas e projetos de instituições nacionais e estrangeiras.

No período que compreendeu a elaboração desta pesquisa, a técnica de coleta adotada foi o acompanhamento nos últimos cinco anos das páginas de iniciativas e projetos de preservação digital onde são publicadas listagens e documentação relacionada ao assunto (ver Apêndice C). Cronologicamente, as referências foram selecionadas por serem bastante ilustrativas em relação ao desenvolvimento de alguns temas de interesse da área.

Tendo como foco a evolução temática da preservação digital como assunto de documentos publicados de 1996 a 2007, foram analisados 319 trabalhos apresentando os aspectos mais relevantes citados nessa produção, seus autores, tipo de autoria, tipo de documentos, origens e propostas. Também foi incluída a produção nacional nesse período registrada em artigos de periódicos, capítulos de livros, teses e dissertações.

Os trabalhos apresentados nos eventos da área não foram incluídos nesta análise devido a que muitos desses *sites* não permitiam o acesso aos documentos ou o texto integral não constava na relação de trabalhos apresentados nas páginas oficiais.

Pretendeu-se, com esta análise, responder às seguintes questões que envolvem o problema desta pesquisa:

1. Quais são os interesses dos autores em relação aos subtemas relacionados com a preservação digital?
2. Como se desenvolveram os principais subtemas relacionados com a preservação digital?
3. Quais foram as principais propostas e aspectos da preservação digital debatidos?
4. Qual é o tipo de documento que está sendo publicado?
5. De qual área do conhecimento eram as publicações que abordaram o tema da preservação digital?
6. Que periódicos científicos veicularam artigos sobre preservação digital?

As respostas a estas questões proporcionaram uma visão do assunto no contexto internacional e nacional. O tratamento bibliométrico realizado na bibliografia reunida tinha como objetivos:

a) identificar os autores e os tipos de documentos mais usados, os sub-temas, propostas, aspectos discutidos no período de 1996-2007;
b) conhecer a produtividade dos autores e o tipo de autoria;
c) identificar as áreas do conhecimento que mais publicaram sobre o tema.

A compilação das referências bibliográficas sobre o tema preservação digital permitiu inserção de dados dessas variáveis no programa estatístico SPSS. Com a aplicação da ferramenta SPSS – *Statistical Package for the Social Sciences* foi possível obter dados bibliométricos que permitiram a contextualização da produção científica da área nos anos 1996-2007. O levantamento incluiu a organização dos dados em dez variáveis, a saber:

1. autores: foram incluídos os nomes de autores e co-autores da bibliografia consultada;
2. autoria: os documentos foram registrados segundo o tipo (única e múltipla) e a ordem de autoria dos trabalhos;
3. área: no levantamento bibliográfico foram identificadas cinco áreas do conhecimento de onde partiram as publicações de trabalhos sobre o assunto (ciência da informação, arquivologia, biblioteconomia, ciência da computação e história);
4. subtema: foram definidos quatro subtemas segundo a abordagem geral dos trabalhos em relação ao tema principal e o enfoque que os autores destacaram (percepções, gerenciamento, estratégias e certificação);
5. proposta: esta variável se referiu ao assunto específico exposto e descrito de forma mais aprofundada pelos autores nos seus textos, sendo identificados 12 tipos de propostas;
6. aspecto: esta variável foi criada para identificar dentro dos subtemas os tópicos discutidos com maior freqüência nos trabalhos publicados, foram 14 questões discutidas resgatadas na análise dos documentos;
7. tipo de publicação: foram identificadas no levantamento 9 tipos de publicações (artigo; livro; tese; dissertação; relatório; capítulo de livro; texto *online*; norma, declaração); o tipo de documento “trabalho em evento científico” não foi incluído por causa da dificuldade de acessar os conteúdos da maioria dos trabalhos apresentados nos últimos anos;

8. periódico: este item foi destacado por se tratar da forma de publicação mais usada pelos autores da área para divulgar seus trabalhos; foram identificados 68 títulos (Tabela 8);
9. país: esta variável serviu para identificar a origem das publicações e sua predominância geopolítica (15 países);
10. ano: foi utilizada para contextualizar temporalmente a produção em preservação digital.

Essas variáveis foram submetidas a um tratamento estatístico descritivo (análise de freqüência) no qual foram cruzados os seguintes fatores:

- subtema *versus* autores;
- subtema *versus* autoria;
- subtema *versus* país
- subtema *versus* ano;
- proposta *versus* aspecto discutido;
- ano *versus* tipo de publicação;
- área *versus* tipo de publicação;
- país *versus* tipo de publicação;
- ano *versus* periódico.

4.1.2 Avaliação da produtividade

Quantificou-se a produção individual dos principais autores em diferentes períodos e seu impacto no desenvolvimento da área. A seguir apresentam-se as análises dos resultados seguindo as sete questões formuladas.

Qual é o interesse dos autores em relação aos subtemas relacionados com a preservação digital?

Os autores produziram documentos sobre quatro subtemas: Percepção, Gerenciamento, Estratégias e Certificação. A média de produções por autor é bastante baixa (Média = 1,29, DP = 0,79). A maioria dos autores (83,1%) teve apenas uma produção; 10,5% dos autores produziram dois documentos cada e 4% dos autores produziram três documentos cada. Um autor produziu quatro

documentos (H. M. Gladney), quatro autores produziram cinco documentos cada (K. Thomaz, M. Day, NLA, OCLC/RLG) e um autor produziu sete documentos (N. Beagrie).

A Tabela 1 apresenta a distribuição de autores em função dos subtemas relacionados à preservação digital. Conforme pode ser verificado, a maioria dos autores (67,1%) produziu documentos sobre Gerenciamento e Estratégias refletindo, desta maneira, maior porcentagem de publicações nestes temas (68,7%). A tabela também aponta uma distribuição semelhante do número de países envolvidos nestas publicações.

Tabela 1 – Produção dos autores em função dos subtemas de interesse

Subtema	Nº. autores	%	Nº. países	%	Nº. publicações	%
Certificação	35	12,4	7	20,0	38	11,9
Estratégias	89	31,4	9	25,7	104	32,6
Gerenciamento	101	35,7	8	22,9	115	36,1
Percepção	58	20,5	11	31,4	62	19,4
TOTAL	283	100	35	100	319	100

Conforme pode ser verificado na Figura 21, dos 16 países envolvidos na produção de documentos, apenas seis países produzem documentos em todos subtemas relacionados à preservação digital, a saber: Estados Unidos, Inglaterra, Brasil, Austrália, Alemanha e Holanda. Notadamente, os países que mais produziram documentos foram os Estados Unidos e a Inglaterra, sendo suas produções equivalentes a 56% e 22%, respectivamente, do total de publicações.

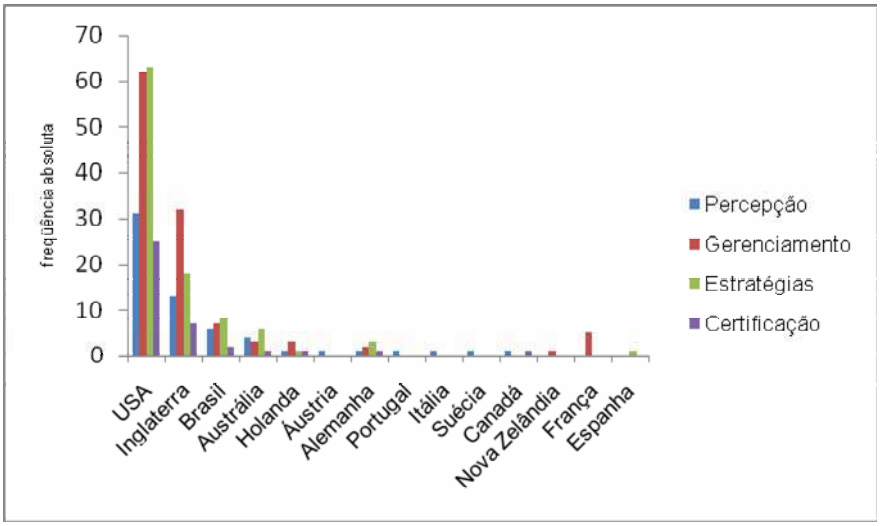


Figura 21 – Subtemas e locais das publicações

Como se desenvolveram os principais sub-temas relacionados com a preservação digital?

A Tabela 2 apresenta a distribuição de freqüência absoluta (fa_p) de trabalhos publicados em cada um dos subtemas e o tipo de autoria nas respectivas categorias.

Tabela 2 – Freqüência de publicações e autoria em função do subtema

Subtema	fa_p	%	Autoria única	%	Autoria múltipla	%
Certificação	38	11,9	29	12	9	11,7
Estratégias	105	33,0	77	32,0	28	36,4
Gerenciamento	116	36,5	90	37,3	26	33,8
Percepção	60	18,9	46	19,1	14	18,2
TOTAL	319	100	241	100	77	100

A Tabela 2 evidencia que a maioria dos trabalhos publicados tratou dos subtemas Estratégias e Gerenciamento, perfazendo um total de 69,5%. Poucos trabalhos se dirigiram a questões relacionadas à Certificação. Destaca-se que 75,8% do total de publicações (i.e., 241 trabalhos) corresponderam à autoria única, contra 24,2% (i.e., 77 trabalhos) de trabalhos publicados por mais de um autor.

Uma explicação do baixo índice de trabalhos que tratavam os assuntos relacionados com a certificação está no fato de que os resultados dos testes de

aplicação de indicadores de auditoria propostos nos atuais modelos de referência (*Repository Audit and Certification - RAC, Model Requirements for Electronic Records Management* - MoReq2 e o *Catalogue of Criteria for Truested Digital Repositories*) encontram-se em avaliação e revisão (CRL; 2007; KACZMARECK *et al.*, 2006; NESTOR, 2006).

Com relação à predominância da autoria única, pode-se afirmar que é motivada pela rápida publicação de resultados de pesquisas individuais, que regularmente demoram menos tempo em serem concluídas do que as que contêm resultados de grandes projetos e inciativas implementadas.

A Figura 22 apresenta a tendência de contribuição de cada subtema durante o período de tempo analisado.

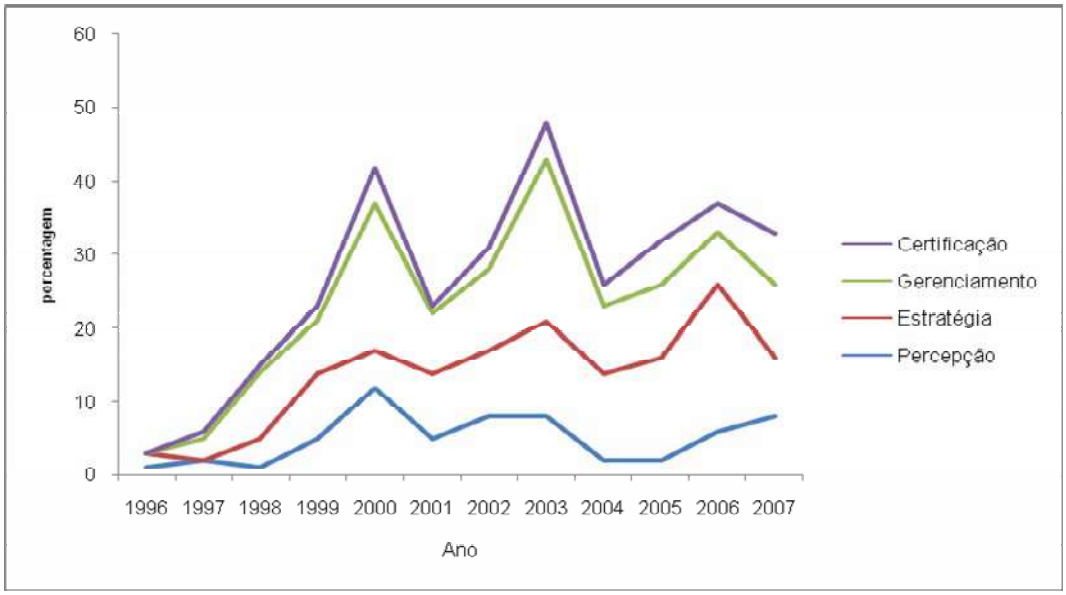


Figura 22 – Contribuição dos subtemas no período 1996 a 2007

A Figura 22 evidencia o aumento da contribuição da maioria dos subtemas a partir de 1997, excetuando o subtema Percepção, cuja contribuição permaneceu relativamente estável. Destaca-se que a participação de Gerenciamento e Certificação foi muito similar ao longo de todo o período, sendo que, nos anos de 2000 e 2003, houve aumento significativo de publicações nestas áreas. Isto pode estar relacionado ao fato de que, para essa época já haviam projetos estáveis em diversos países e que a pesquisa começou a enfocar o desenvolvimento de ferramentas e repositórios confiáveis (McGOVERN, 2007).

A tendência de contribuição do tipo de autoria pode ser vista na Figura 23, onde se observa aumento crescente no número de autorias únicas durante o período 1996-2007, sendo que, nos anos de 2000 e 2003, registrou-se número expressivo de trabalhos individuais. A freqüência de trabalhos publicados por mais de um autor aumentou modestamente em 1998, e tendeu a se estabilizar até 2004, quando pode ser verificado pequeno aumento deste tipo de autoria, que retornou aos níveis anteriores no ano de 2007.

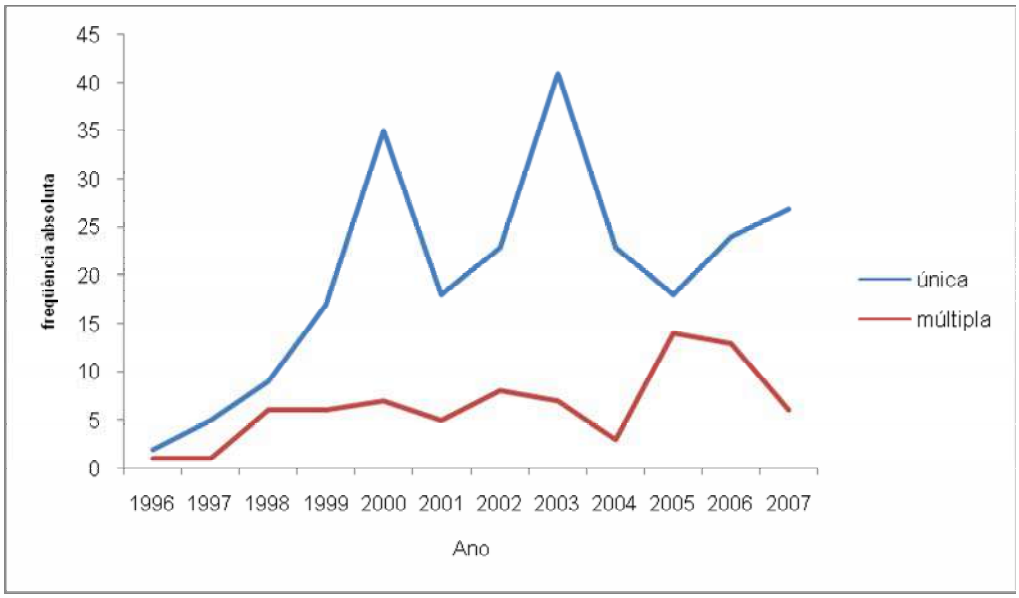


Figura 23 – Freqüência do tipo de autoria em função do ano de publicação

Quais foram as principais propostas e aspectos relacionados com a preservação digital debatidos?

Foram identificados 12 tipos de propostas no período analisado. Estas propostas estão relacionadas na Tabela 3. A categoria “recomendação” foi o tipo mais sugerido pelos pesquisadores (55,3%), seguido de “projetos” (15,6%). As dez categorias restantes corresponderam a 29,1% do total de propostas registradas, sendo “curso” e “catálogo” os tipos menos sugeridos pelos autores.

Tabela 3 - Tipos de propostas para preservação digital

Tipo	freqüência	%
recomendação	163	55,3
projeto	46	15,6
modelo	20	6,8
software	16	5,4
política	13	4,4
colaboração	12	4,1
manual	8	2,7
guia	6	2,0
tutorial	4	1,4
dicionário	4	1,4
curso	2	0,7
catálogo	1	0,3
TOTAL	295	100

Uma análise temporal da ocorrência de propostas, observada na Figura 24, mostra que foram apresentadas “recomendações” ao longo de todo o intervalo 1996 a 2007, e que o número de recomendações tendeu a aumentar a partir de 1997. Em 2000 e 2003 o número de recomendações foi superior aos demais anos do intervalo temporal considerado. Em segundo lugar se destacaram os “projetos” apresentados a partir de 1998, tendo ocorrido aumento significativo nos anos de 2000, 2003 e 2006.

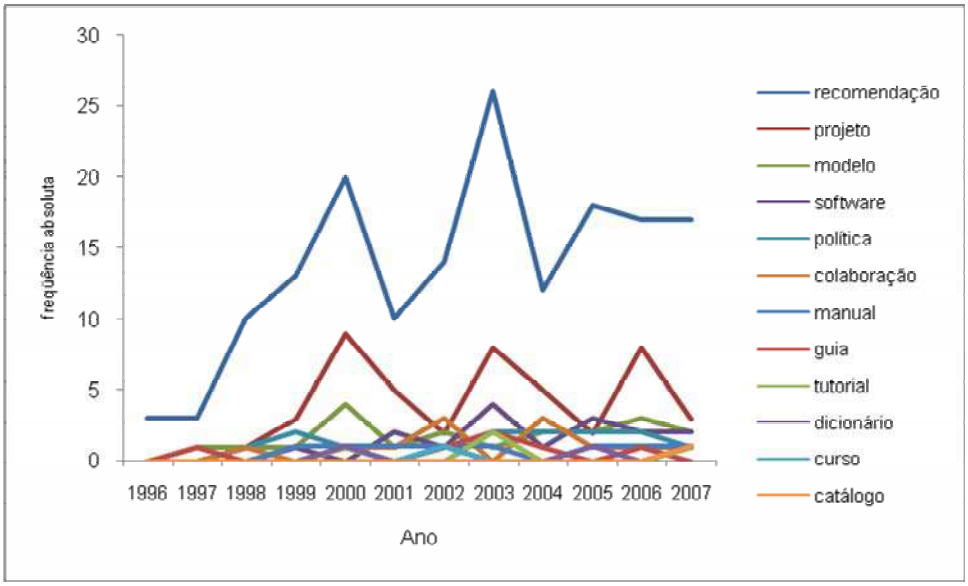


Figura 24 – Tipos de propostas apresentadas no período 1996 a 2007

Durante grande parte do período analisado também foram registradas propostas de manuais, *software*, políticas e colaborações. As propostas menos presentes ao longo do período foram tutoriais, cursos, dicionários e catálogos.

Uma análise qualitativa foi realizada para avaliar quais foram as questões debatidas no contexto das propostas apresentadas. Quatorze questões foram identificadas e podem ser vistas na Tabela 4, a seguir, que inclui a freqüência absoluta de ocorrência e a respectiva porcentagem de cada questão. Seis tipos de questões dominaram as propostas e juntas perfizeram 71% do total de questões: iniciativa, sistemas e ferramentas de arquivamento, coleções e recursos digitais, instituições e atributos e requisitos.

Tabela 4 - Questões debatidas no período 1996-2007

Questões debatidas	freqüência	%
iniciativa	14,1	13,9
sistemas e ferramentas de arquivamento	12,2	12,0
coleções e recursos digitais	11,9	11,7
objetos e informação digital	11,3	11,4
instituições	10,7	10,8
atributos e requisitos	10,7	10,8
metadados	6,9	7,0
publicações eletrônicas	6,3	6,3
padrões	6,0	6,0
custos	2,5	2,5
legislação	2,2	2,2
migração	1,9	1,9
acesso	1,9	1,9
emulação	1,6	1,6
Total	100	319

O Quadro 12, a seguir, retrata a presença de cada uma das questões debatidas no contexto das propostas ao longo do período 1996-2007. Nesta figura, as células preenchidas em cor cinza indicam que a questão foi debatida no respectivo ano. Conforme ilustra esse Quatro, nove questões estiveram presentes ao longo de praticamente todo o período focalizado: coleções e recursos digitais, objetos e informação digital, instituições, iniciativa, sistemas e ferramentas de arquivamento, atributos e requisitos, publicações eletrônicas, padrões e metadados. Dentre estas, registrou-se que questões relacionadas às coleções e recursos digitais, bem como objetos e informação digital foram as mais presentes no contexto

das propostas ao longo dos últimos 11 anos. Em contrapartida, observa-se que não estiveram muito presentes nas propostas discussões sobre questões relacionadas ao acesso, migração, legislação, emulação e custos durante o mesmo período.

Questões Debatidas	Ano											
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
acesso												
atributos e requisitos												
coleções e recursos digitais												
custos												
emulação												
iniciativa												
instituições												
legislação												
metadados												
migração												
objetos e informação digital												
padrões												
publicações eletrônicas												
sistemas e ferramentas de arquivamento												

Quadro 12 - Questões debatidas em cada ano do período

Por fim, a Tabela 5 traz a porcentagem de ocorrência de cada questão debatida no contexto de cada proposta apresentada no intervalo 1996-2007. A tabela mostra como o número de ocorrências de recomendações sobre todas as questões identificadas foi maior. Uma característica comum desses trabalhos esteve na forma como eles cobriam toda a variedade de assuntos relacionados ao tema e, a necessidade de apresentar descrições básicas sobre o gerenciamento dos objetos digitais.

Tabela 5 - Propostas e Questões Debatidas ao longo do período 1996-2007

Questões Debatidas	Propostas (%)											
	Rec	Pro	Mod	SW	Pol	Col	Man	G	Tut	Dic	Cur	Cat
iniciativa	6,46	3,74	0,00	0,00	0,68	1,70	0,00	0,00	0,34	0,00	0,34	0,00
coleções e recursos digitais	8,50	0,68	0,00	0,34	0,68	0,34	1,02	0,00	0,34	0,00	0,00	0,00
objetos e informação digital	9,52	0,00	0,34	0,68	0,00	0,00	0,00	0,34	0,34	0,00	0,00	0,00
instituições	6,46	2,38	0,68	0,00	0,68	0,68	0,00	0,00	0,00	0,00	0,34	0,00
atributos e requisitos	6,46	0,34	2,04	0,00	1,70	0,00	0,34	0,34	0,00	0,00	0,00	0,34
metadados	1,70	1,02	1,36	0,00	0,34	0,00	0,34	0,34	0,34	0,68	0,00	0,00
publicações eletrônicas	3,40	3,06	0,00	0,00	0,00	0,34	0,00	0,00	0,00	0,00	0,00	0,00
padrões	3,06	0,68	0,34	0,34	0,00	0,00	0,68	0,68	0,00	0,68	0,00	0,00
custos	2,38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
legislação	1,70	0,34	0,00	0,00	0,34	0,00	0,00	0,00	0,00	0,00	0,00	0,00
migração	0,68	1,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
acesso	1,36	0,00	0,34	0,00	0,00	0,34	0,00	0,00	0,00	0,00	0,00	0,00
emulação	0,34	0,34	0,68	0,00	0,00	0,00	0,00	0,34	0,00	0,00	0,00	0,00
sistemas e ferramentas	3,74	1,70	0,68	4,08	0,00	0,68	0,34	0,00	0,00	0,00	0,00	0,00

Legenda: Rec = recomendação, Pro = projeto, Mod = modelo, SW = software, Pol = política, Col = colaboração, Man = manual, G = guia, Tut = tutorial, Dic = dicionário, Cur = curso, Cat = catálogo

Qual é o tipo de documento que está sendo publicado?

A Figura 25 apresenta a tipologia de documentos verificada na amostra coletada. A porcentagem de cada tipo de documento pode ser observada junto a cada eixo da figura.

A tipologia mais freqüentemente utilizada pelos autores foi artigo (62%) e, em segundo lugar, textos *on-line*. Todos os demais tipos de documentos representaram 23% da tipologia observada, ou seja, os autores não têm preferido utilizar teses, declarações, normas, dissertações, capítulos de livro, livros ou relatórios para divulgar seus trabalhos na área de preservação digital.

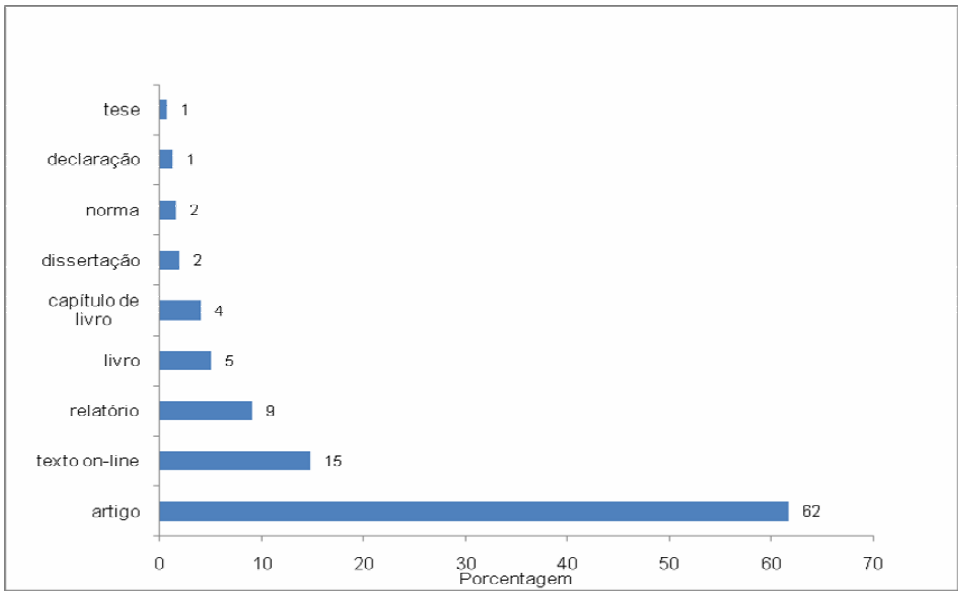


Figura 25 – Tipologia de documentos

Conforme pode ser constatado na Figura 26, a linha de tendência, delineada na cor preta e pontilhada, evidencia aumento crescente do número total de documentos produzidos no período de 1996 a 2007. Este aumento iniciou marcadamente a partir de 1997, sendo que nos anos de 2000 e 2003 a produção foi significativamente superior aos demais anos do período.

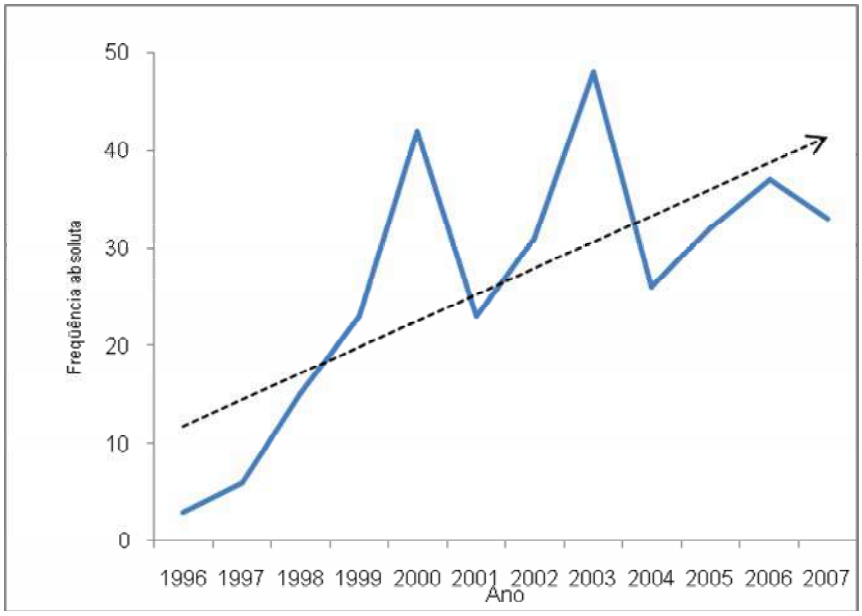


Figura 26 – Tendência da produção de documentos no período 1996 a 2007

É interessante observar o tipo de documento produzido no período de 1996 a 2007, conforme ilustra a Figura 27.

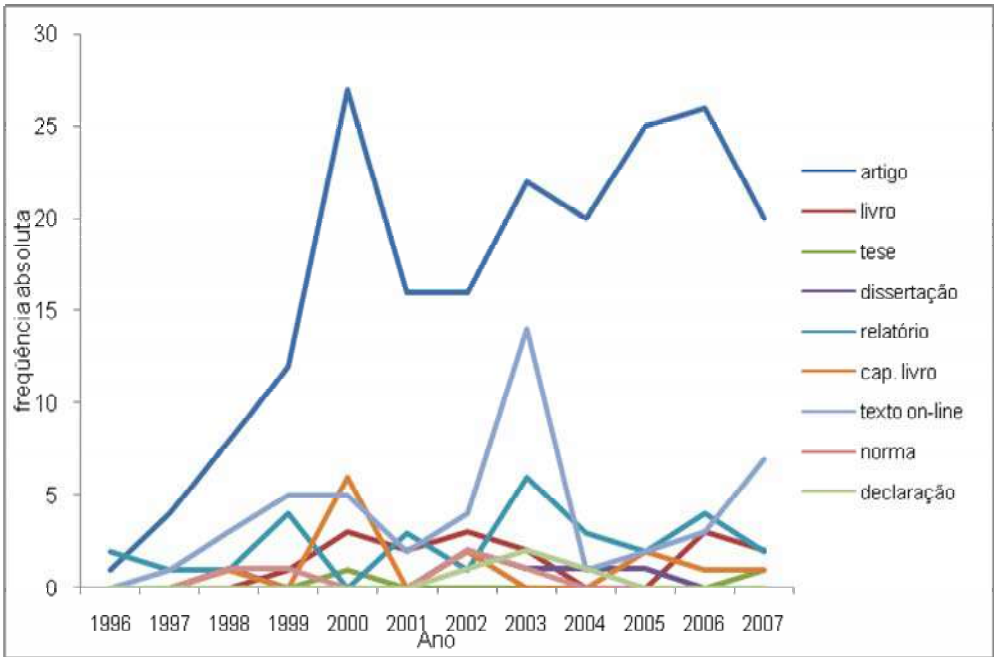


Figura 27 – Tipo de produção em função do ano de produção

A Figura 27 mostra que a categoria mais presente ao longo de todo o período foi “artigo”. Além disto, verifica-se tendência de aumento na quantidade de artigos publicados ao longo dos 11 anos de observação. A segunda categoria de publicação mais presente durante este intervalo de tempo foram os textos *on-line,* e a terceira categoria foram os relatórios. Nota-se, ainda, que os livros começaram a ser produzidos a partir do ano de 1999.

De qual área do conhecimento eram as publicações que abordaram o tema da preservação digital?

As publicações encontradas estavam associadas a uma das seguintes áreas: biblioteconomia, arquivologia, ciência da informação, ciência da computação e história. Conforme ilustra a Tabela 6, a maioria dos trabalhos foi produzida na área de biblioteconomia. Produções nas áreas de história ou ciência da computação são muito raras.

Tabela 6 – Distribuição de freqüência da área do conhecimento

Área	freqüência	%
biblioteconomia	256	80,3
arquivologia	34	10,7
ciência da informação	20	6,3
ciência da computação	5	1,6
história	2	0,6
sem/área	2	0,6
TOTAL	319	100

A Tabela 7, a seguir, apresenta a distribuição das publicações em função da área de conhecimento. Dois valores em cada célula da tabela são mostrados: os números ao lado dos parênteses indicam a freqüência de ocorrência de cada publicação na respectiva área de conhecimento, e o número entre parênteses indica a porcentagem da respectiva produção em relação ao total de produções.

Tabela 7 – Tipos de publicação em função da área de conhecimento

Tipo de Publicação	Área de Conhecimento					
	ciência da informação	arquivologia	biblioteconomia	ciência da computação	história	TOTAL
artigo	11 (3,5%)	19 (6,0%)	165 (52,1%)	1 (0,3%)	(0%)	61,8%
capítulo de livro	1 (0,3%)	4 (1,3%)	6 (1,9%)	(0%)	2 (0,6%)	4,1%
declaração	(0%)	1 (0,3%)	2 (0,6%)	(0%)	(0%)	0,9%
dissertação	5 (1,6%)	(0%)	1 (0,3%)	(0%)	(0%)	1,9%
livro	1 (0,3%)	4 (1,3%)	12 (3,8%)	(0%)	(0%)	5,4%
norma	(0%)	3 (0,9%)	1 (0,3%)	(0%)	(0%)	1,3%
relatório	29 (9,1%)	(0%)	(0%)	(0%)	(0%)	9,1%
tese	2 (0,6%)	(0%)	(0%)	(0%)	(0%)	0,6%
texto on-line	(0%)	3 (0,9%)	40 (12,6%)	4 (1,3%)	(0%)	14,8%

Conforme mostra essa tabela foram produzidos artigos em quatro áreas de conhecimento, sendo que a maior parte ocorreu em biblioteconomia. A maioria dos livros, capítulos de livros, textos *on-line* e declarações também estavam associados a esta área do conhecimento.

A maior parte da produção na área de ciência da informação consistiu de relatórios, mas também foram registrados alguns artigos, livros, tese, dissertação e capítulo de livro.

Na área de arquivologia foram encontrados os seguintes tipos de publicação: artigo, livro, capítulo de livro, texto *on-line*, norma e declaração. Na ciência da computação foram registrados poucos textos *on-line* e artigos, e na área de história somente encontrou-se a produção do tipo capítulo de livro.

Que periódicos científicos veicularam artigos sobre preservação digital?

Como pode ser observado na tabela 8, a seguir, a concentração de números de periódicos que mais publicaram artigos sobre preservação digital esteve nos Estados Unidos e na Inglaterra. Existe alta concentração de artigos em pequeno número de periódicos, a maioria da área da biblioteconomia e ciência da informação.

Tabela 8 - Título e país de origem dos periódicos com artigos sobre preservação digital

Título	País	Artigos
ACM Transactions on Computer Systems	Estados Unidos	1
ACM Transactions on Information Systems	Estados Unidos	1
Alexandria	Inglaterra	1
American Archivist	Estados Unidos	1
American Libraries	Estados Unidos	2
Ariadne	Inglaterra	13
ARL: A Bimontly Report on Research Library Issues and Action	Estados Unidos	1
Arquivística.net	Brasil	3
Aslib Proceedings	Inglaterra	1
Australian Academic & Research Libraries	Austrália	2
Change	Estados Unidos	1
Chronicle of Higher education	Estados Unidos	1
Ciência da Informação	Brasil	1
Collection Management	Estados Unidos	3
College & Research Libraries	Estados Unidos	2
Communications of ACM	Estados Unidos	1
Computers and the Humanities	Holanda	2
Computers in Libraries	Estados Unidos	1
Data Science Journal	França	2
DataGramaZero	Brasil	2
Digital Library Network	Estados Unidos	1
D-Lib Magazine	Estados Unidos	35
Documents Numerique	França	1
EDUCAUSE Review	Estados Unidos	2
El Profesional de la Información	Espanha	1
First Monday	Estados Unidos	2
Hi Energy Physics Libraries Webzine	Suíça	1
Information Services & Use	Holanda	2
International Journal of Digital Curation	Inglaterra	4
International Journal on Digital Libraries	Alemanha	3
Issues in Science and Technology Librarianship	Estados Unidos	1

Continua

Tabela 8 - Continuação

Título	País	Artigos
Journal of Academic Librarianship	Holanda	1
Journal of American Society for Information Science and Technology	Estados Unidos	1
Journal of Archival Organization	Estados Unidos	1
Journal of Digital Information	Estados Unidos	2
Journal of Electronic Publishing	Estados Unidos	4
Journal of Library Administration	Inglaterra	1
Journal of Research of the National Institute of Stand Tech	Estados Unidos	2
Journal of the Association of Moving Image Archivists	Estados Unidos	1
Learned Publishing	Inglaterra	4
Libraries and the Academy	Estados Unidos	3
Library and Information Research News	Inglaterra	1
Library Hi Tech	Inglaterra	2
Library Journal	Estados Unidos	1
Library Journal, Net Connect	Estados Unidos	1
Library Management	Inglaterra	1
Library Resources & Technical Services	Estados Unidos	1
Library Review	Inglaterra	2
Library Trends	Estados Unidos	6
Multimedia Information and Technology	Inglaterra	1
New Review of Academic Librarianship	Inglaterra	4
OCLC Systems & Services	Estados Unidos	3
On-line Journal for Cultural Studies	Austria	1
Perspectivas em ciência da informação	Brasil	1
Program	Inglaterra	2
Reference Librarian	Estados Unidos	1
Research Information	Estados Unidos	2
Revista IP	Brasil	1
RLG DigiNews	Estados Unidos	27
SciCom Info	Suécia	1
Science & Technology Libraries	Estados Unidos	1
Serials	Inglaterra	3
Serials Librarians	Estados Unidos	4
Serials Reviews	Estados Unidos	11
Technicalities	Estados Unidos	1
The Botton Line: Managing Library Finances	Inglaterra	1
Update	Inglaterra	1
Vine	Inglaterra	1

Os exemplos citados refletem a comunidade de prática definida por representantes de arquivos, bibliotecas, museus e outras instituições de ensino e pesquisa. Foi constatada a origem geográfica da maior parte das informações divulgadas, as recomendações e descrições dos aspectos técnicos da preservação digital publicadas por autores em periódicos no exterior, os quais, de alguma forma, estavam relacionados com iniciativas e projetos de instituições no Primeiro Mundo.

4.1.3 Temáticas e questões discutidas

As referências citadas formaram um contingente de documentos selecionados, que produziram a necessidade de delimitar o escopo da análise bibliométrica. Todos os trabalhos retratam uma amostragem da produção intelectual sobre os assuntos relacionados ao tema principal desta pesquisa.

O estudo apontou como resultados que a definição sobre o conceito de preservação digital não é mais uma preocupação para os autores e que os aspectos relacionados com a implantação e certificação de repositórios digitais estão no centro do debate nos últimos anos. Os autores revelaram a preocupação com a necessidade de conhecimento sobre os procedimentos de armazenamento e acesso aos documentos digitais.

Também, o estudo permitiu avaliar a produção científica em categorias específicas, levando em consideração os debates centrais propostos pelos autores. A definição das variáveis foi o resultado da revisão da literatura que envolveu um dos objetivos específicos desta pesquisa, que é identificar os critérios para o estabelecimento de sistemas de gestão de preservação digital de informação científica.

Identificados os principais autores e fontes de informação da área, as questões discutidas por eles focalizaram as características das iniciativas e as características das coleções e objetos digitais com que eles trabalham. O aparecimento de alto número de trabalhos em publicações periódicas nacionais e estrangeiras divulgando apenas recomendações é uma indicação do estágio de desenvolvimento de estudos sobre o assunto.

Em vários livros da área da ciência da informação a preservação digital aparece como um tópico relativo ao gerenciamento de informação em bibliotecas e repositórios digitais. Também é rara a publicação de teses ou dissertações dedicadas ao tema da preservação digital. Pode afirmar-se que a área da preservação digital está na sua fase de institucionalização e legitimação.

4.2 ANÁLISE DOS DADOS DOS QUESTIONÁRIOS

Examinar a amostra da população das unidades de pesquisa do MCT que possuem centros de informação e biblioteca implicou uso de métodos de coleta de dados para constituir uma lógica de análise. A aplicação de questionários e entrevistas permitiu a coleta direta de informação com as pessoas envolvidas com as atividades institucionais relacionadas com a preservação dos documentos digitais.

Como mostraram Fink e Kosecoff (1985), os questionários e entrevistas são métodos válidos, exeqüíveis e que transferem credibilidade aos resultados das pesquisas, havendo apenas a necessidade de se indagar a quem as perguntas serão dirigidas, onde e com que freqüência, assim como a abordagem de análise e a interpretação dos dados.

O questionário utilizado nesta pesquisa (Apêndice A) foi produto da necessidade de contextualizar as práticas de preservação dos documentos digitais e do interesse por parte da Biblioteca de Política em Ciência e Tecnologia (BPCT) da Coordenação-Geral de Gestão e Inovação (CGGI) do Ministério da Ciência e Tecnologia (MCT). A biblioteca registrava, entre seus projetos para o ano de 2004, o levantamento das atividades que as unidades de pesquisa realizavam com relação ao tratamento dos seus documentos bibliográficos, em qualquer suporte.

O questionário foi encaminhado pela responsável da BPCT via correio eletrônico para os endereços oficiais dessas instituições, as respostas foram analisadas conjuntamente com a BPCT e as informações coletadas puderam ser confirmadas posteriormente, por meio das entrevistas nas sedes das unidades.

A aplicação do questionário foi uma escolha justificada pela expectativa de poder identificar o maior número de pessoas envolvidas com os trabalhos de preservação de documentos em todos os formatos. No quesito sobre o tratamento dos documentos digitais, esperava-se encontrar dados que pudessem auxiliar na localização de informações oficiais sobre esses objetos digitais.

Os dados dos questionários evidenciaram o contexto dos centros de informação do MCT, o perfil das suas equipes e as características do manejo da informação digital no âmbito do Ministério.

4.2.1 Características dos centros de informação

A maioria das unidades de pesquisa do Ministério da Ciência e Tecnologia conta com centros de informação ou setor de biblioteca onde são tratados todos os materiais bibliográficos das suas áreas de competência, produzidos por pesquisadores internos e externos à instituição.

Foram recebidos questionários com respostas de responsáveis por centros de informação e bibliotecas de 20 unidades de pesquisa do MCT. Os dados foram classificados para análise em sete grupos:

1) setor responsável pelo tratamento dos materiais;
2) tipos de materiais tratados nos centros;
3) tipo de aplicação de *software* usado;
4) ferramenta para indexação dos conteúdos;
5) tipo de acesso;
6) tratamento da produção científica local;
7) tratamento dos materiais digitais.

As respostas das instituições apontaram as peculiaridades em relação ao tipo de tratamento dado aos registros da produção científica. O contexto histórico de cada unidade pode ser observado no Apêndice D.

O Quadro 13, a seguir, apresenta os locais das unidades de pesquisa e os materiais bibliográficos tratados pelas respectivas unidades. Como pode ser verificado, seis unidades de pesquisa não responderam a esta questão, quais sejam: Inpa, Mast, CNPq, Cetem, Impa e Cnen.

Para rastrear a localização dos arquivos digitais, foi necessário perguntar sobre o setor responsável pelo tratamento segundo o tipo de material. Baseados nas respostas de treze unidades, constatou-se que, segundo a percepção dos responsáveis pelos centros de informação e biblioteca, eram eles os encarregados por esse trabalho.

Isso pode ser verificado no Quadro 13, uma vez que a maioria das unidades de pesquisa atribuiu às bibliotecas a responsabilidade do respectivo tratamento.

Unidade de pesquisa	Tipo de material								
	Monografias (livros/ anais/ teses/ dissertações)	Analíticas (capítulos de livros)	Periódicos	Artigos	Comunicações em eventos	Multimeios (filmes, CDs, DVDs, fotografias etc.)	TCC Trabalhos Experimentais	Produção docente	Outros
CBPF		Biblioteca	Biblioteca	Biblioteca	Biblioteca	Biblioteca	Biblioteca	Biblioteca	
CenPRA	Divisão de Comunicação e Gabinete DCOG		DCOG		DCOG	DCOG			
CGEE	Informação	Informação	Informação	Informação	Informação	Informação			
Finep	Biblioteca		Biblioteca	Biblioteca		Biblioteca			
Ibict	Processos Técnicos		Aquisição	Processos Técnicos					
IDSM	Biblioteca	Biblioteca	Biblioteca	Biblioteca		Biblioteca			
Inpe	Processamento Técnico de Livros		Processamento Técnico de Periódicos		Processamento Técnico de Livros	Processamento Técnico de Livros		Memória Técnico-Científica	
INT	Biblioteca								
LNA	Biblioteca		Biblioteca						
LNCC	Biblioteca		Biblioteca						
LNLS	Biblioteca	Biblioteca	Biblioteca	Biblioteca	Biblioteca	Biblioteca	Biblioteca	Biblioteca	
Mpeg	Biblioteca	Biblioteca	Biblioteca			Biblioteca e Arquivo			Biblioteca (coleção de livros raros)
ON	Divisão de Informação e Documentação								
RNP									Centro de Informações

Quadro 13 - Tipo de material segundo e setor responsável no centro de informação e biblioteca

O Cnen contava com 14 unidades vinculadas, algumas delas com bibliotecas, mas unicamente a biblioteca sede foi incluída nesta pesquisa. Um centro de informação que não foi incluído foi a Coordenação de Sistemas de Informação da unidade Indústrias Nucleares do Brasil (INB). A resposta recebida ao envio do questionário explicava que:

> - [...] o acervo de nossa biblioteca é constituído basicamente de normas técnicas nacionais e estrangeiras para atender às necessidades das duas fábricas de combustível nuclear. A produção intelectual interna – se existe – não está acessível, nem consta do acervo da nossa biblioteca, tendo em vista o caráter restrito de parte das atividades da INB. (Bibliotecária R2)

A segunda comparação dos resultados ocorreu com relação aos tipos de materiais e aos *software* utilizados para o registro em bases de dados locais. O Quadro 14, a seguir, apresenta as bases de dados usadas pelos centros de informação e bibliotecas para o registro dos materiais bibliográficos. O Quadro inclui as respostas de 18 unidades de pesquisa. Cetem e CNPq não responderam a este item do questionário.

No Quadro, as células preenchidas com a palavra “Não” significam que a unidade não utiliza o sistema de base de dados aplicado em outros materiais pela própria unidade.

Existe uma variedade de soluções em funcionamento nas bibliotecas e centros de informação, representando um modelo que não permite a integração das informações de todas as instituições. As soluções encontradas eram bases de dados proprietárias, provedores de dados de arquivos abertos, repositórios digitais e sistemas de publicação eletrônica. A falta de compartilhamento das soluções adotadas deixou em evidência as características de sistemas de gestão da informação pouco integradoras dos centros de informação e bibliotecas.

Unidade de pesquisa	Tipo de material								
	Monografias (livros, anais, teses/ dissertações)	**Analíticas (capítulos de livros)**	**Periódicos**	**Artigos**	**Comunicações em eventos**	**Multimeios (filme, CD, DVD, fotografia, etc.)**	**TCC / Trabalhos Experimentais**	**Produção docente**	**Outros**
CBPF	Sistema Próprio	Sistema Próprio	Sistema Próprio	Sistema Próprio	Sistema Próprio	Sistema Próprio	Sistema Próprio	Sistema Próprio	
CenPRA	SICTEC								
CGEE	Próprio	Próprio	Próprio	Próprio	Próprio	Próprio			
Cnen	Microisis com interface desenvolvida localmente		Secs – Bireme (local)		Anais de eventos – sistema próprio (base Anais)	Microisis com interface desenvolvida localmente)			
Finep	Sysbibli		Sysbibli	Sysbibli		Sysbibli			
Ibict	Microisis – PHL 8.0	Não	Caribe - PHL 8.0	Microisis – PHL 8.0 [1]	Microisis – PHL 8.0	PHL 8.0 [2]		Gerenciado no DEP/RJ	
IDSM	Caribe	Caribe	Caribe	Caribe		Caribe			
Impa	Horizont		Horizont						
Inpa	BIBLIOBASE					BIBLIOBASE	BIBLIOBASE	BIBLIOBASE	
inpe	Sophia	Não	Sophia	Não	Sophia	Sophia	Não	URLib Service [3]	
INT	SYSBIBLI/PHL	SYSBIBLI/PHL				CD de Anais	Científica	Defesas de teses	
LNA	Access for Windows		Access for Windows	Não	Dentro de monografias				
LNCC	Pergamum		pergamum						
LNLS	RM Biblios	Não	RM Biblios	Não	RM Biblios	RM Biblios	RMBiblios	Delphi / SQL Server	
Mast	Microisis	Microisis	Peripuc		Microisis	Microisis	Microisis	Microisis	
Mpeg	Caribe	Caribe	Caribe			Caribe			Caribe (coleção livros raros)
ON	Microisis (sendo implantado) Pergamum	Não	ReriPuc (sendo implantado)	Não	Não	Microisis (sendo implantado)		Sim	
RNP	Não	Não	*web*	Não	Não	Não	Não	Não	Access Windows

Quadro 14 – Bases de dados usadas nos centros de informação e bibliotecas para o registro dos materiais bibliográficos

As ferramentas de indexação utilizadas pelos centros e bibliotecas das unidades, segundo o tipo de material, podem ser vistas no Quadro 15. Somente duas unidades de pesquisa não responderam a esta solicitação do questionário (ON e CNPq). Neste Quadro, as células que contêm a palavra “Não” significam que a ferramenta de indexação escolhida não foi aplicada para esse tipo de material.

A maioria das unidades de pesquisa do MCT conta com algum tipo de ferramenta de indexação dos registros bibliográficos em suas bases de dados. Esse tipo de mecanismo permite que seja realizada a correta descrição da produção científica internacional e local. Os dados do Quadro 15 atestam que a forma tradicional de descrição bibliográfica continua prevalecendo nessas instituições.

Unidade de pesquisa	Tipo de material								
	Monografias (livros/ anais/teses/ dissertações)	Analíticas (capítulos de livros)	Periódicos	Artigos revistas	Comunicações em eventos	Multimeios (filme, CD, DVD, Fotografia, etc.)	TCC/Trabalhos Experimentais	Produção docente	Outros
CBPF	INSPEC		Não						
CenPRA	AINFO-NTIA (EMBRAPA)		AINFO-NTIA (EMBRAPA)			AINFO-NTIA (EMBRAPA)			
Cetem	Cabeçalho de assunto Bibliodata								
CGEE	Termos livres	Termos livres	Termos livres	Termos livres	Termos livres	Termos livres			
Cnen	Vocabulário próprio				Não				
Finep	Vocabulário sistematizado		Vocabulário sistematizado	Vocabulário sistematizado		Vocabulário sistematizado			
Ibict	Tesauro da Ciência da Informação – TCI e o Spines para áreas correlatas		TCI	TCI					
IDSM	INTERVOC	INTERVOC	INTERVOC	INTERVOC		INTERVOC			
Impa	Classificação de assunto do IMPA	Classificação de assunto do IMPA	Classificação de assunto do IMPA	Classificação de assunto do IMPA	Classificação de assunto do IMPA	Classificação de assunto do IMPA		Classificação de assunto do IMPA	
Inpa	Library of Congress (LC) e Biblioteca Nacional (BN)		LC e BN	LC e BN			LC e BN		
Inpe	Tesauro da NASA		Não		Tesauro da NASA	Tesauro da NASA		Tesauro da NASA	
INT	Vocabulário controlado								
LNA		Não	Não	Não	Não				
LNCC	Subject Heading da LC		Subject Heading da LC						
LNLS	Assunto	Não	Assunto	Não	Assunto	Assunto	Assunto	Ano de Publicação	
Mast	Vocabulário controlado	Vocabulário controlado			Vocabulário controlado	Vocabulário controlado	Vocabulário controlado		
Mpeg	CDD20ed. SUBJECT, 23ed. THESAURUS ESPECIALIZADOS	CDD20ed. SUBJECT, 23ed. THESAURUS				CDD			CDD (coleção de livros raros)
RNP	Vocabulário livre	Vocabulário livre	Vocabulário livre	Vocabulário livre	Vocabulário livre	Vocabulário livre	Vocabulário livre	Vocabulário livre	

Quadro 15 - Ferramenta de indexação utilizada segundo o tipo de material

O Quadro 16, a seguir, indica o tipo de acesso disponibilizado pelas unidades de pesquisa do MCT, segundo o tipo de material. O Mast e INT não responderam a este item do questionário.

Como observado na revisão da literatura, uma das preocupações centrais dos envolvidos na aplicação de critérios de preservação digital é o tipo de acesso e a inserção de dados nos repositórios (CCSDS, 2003; DAY, 2002). Nos casos em que ainda não foi realizada nenhuma forma de padronização em relação a metodologias de preservação digital, esperava-se que existisse uma postura institucional que estabelecesse as condições mínimas para o funcionamento de algum sistema de preservação digital. Entretanto, esta expectativa não se confirmou.

No caso das unidades de pesquisa do MCT, os materiais analógicos e digitais permitem dois tipos de acesso: o bibliográfico, que disponibiliza as informações catalográficas do registro; e aquele que permite o acesso ao texto completo.

Observou-se que, nas instituições que permitiam algum tipo de acesso ao texto completo, isso era realizado por meio da indicação do *site* da revista na qual o mesmo estava disponível. Conforme demonstrado na revisão da literatura, esse tipo de prática não pode ser considerado de preservação digital, pois se a biblioteca apenas disponibiliza o endereço da revista ou uma cópia do texto da revista com acesso livre na Internet, não há qualquer garantia de preservação deste material.

Unidade de pesquisa	Tipo de material								
	Monografias (livros/ anais/ teses/ dissertações)	**Analíticas (capítulos de livros)**	**Periódicos**	**Artigos revistas**	**Comunicações em eventos**	**Multimeios (filme, CD, DVD, fotografia, etc.)**	**TCC/ Trabalhos Experimentais**	**Produção docente**	**Outros**
Cenpra	Bibliográfico		Bibliográfico		Bibliográfico	Bibliográfico	Bibliográfico		
Cetem	Texto completo								
CGEE	Bibliográfico e Texto completo	Bibliográfico e Texto completo	Bibliográfico e Texto completo	Bibliográfico e Texto completo	Bibliográfico e Texto completo	Bibliográfico e Texto completo			
Cnen	Bibliográfico		Bibliográfico		Bibliográfico				
Finep	Bibliográfico		Bibliográfico	Bibliográfico		Bibliográfico			
Ibict	Bibliográfico		Bibliográfico	Bibliográfico					
IDSM	Bibliográfico	Bibliográfico	Bibliográfico	Bibliográfico		Bibliográfico			
Impa	Bibliográfico *On-line*		Texto completo on-line						
Inpa	Texto completo				Texto completo	Texto completo	Texto completo	Texto completo	
Inpe	Bibliográfico		Bibliográfico		Bibliográfico	Bibliográfico		Bibliográfico e Texto completo	
LNA	Bibliográfico		Bibliográfico						
LNCC	Bibliográfico		Bibliográfico						
LNLS	Bibliográfico Intranet/ *On-line*		Bibliográfico Intranet / alguns textos completos	Bibliográfico *On-line*	Bibliográfico Intranet /*On-line*	Bibliográfico Intranet / *On-line* / alguns textos completos	Bibliográfico Intranet / *On-line*	Bibliográfico Intranet / *On-line* / alguns textos completos	
Mpeg	Bibliográfico	Bibliográfico				Bibliográfico			
ON	Bibliográfico		Bibliográfico						
RNP	Bibliográfico	Bibliográfico	Bibliográfico	Bibliográfico	Bibliográfico	Bibliográfico	Bibliográfico	Bibliográfico	Texto completo

Quadro 16 – Tipo de acesso disponibilizado pelas unidades de pesquisa do MCT segundo o tipo de material

As formas de tratamento da produção científica das diferentes unidades de pesquisa podem ser observadas no Quadro 17, a seguir. Quatro unidades não responderam a esta questão: INT, ON, CNPF e LNA.

Conforme mencionado anteriormente, requisito indispensável para o estabelecimento de um sistema de preservação digital nas instituições é a necessidade de definir as formas de armazenamento e metodologias adequadas para a conservação dos materiais analógicos e digitais (FUNARI, 2005; LYNCH, 2003; PETERS, 2002).

O Quadro evidencia que as unidades de pesquisa aplicam formas de classificação, armazenamento e metodologias de conservação da produção científica em qualquer suporte. Além disso, destaca-se a falta de atenção que a preservação de longo prazo tem merecido nessas unidades e a necessidade da análise dos requisitos, planejamento, concretização e avaliação de soluções tecnológicas adequadas para a preservação dos recursos digitais.

As práticas de preservação continuam sendo moldadas pelas atividades tradicionais de preservação para o material impresso, sendo que a preocupação com a preservação digital ainda não é central para a maioria dessas instituições. O tipo de documento que apresenta alguma implementação de preservação digital é o periódico eletrônico. Relatórios técnicos em formato digital e outros tipos de literatura cinzenta estão começando a ser coletados, mas sem enfrentar diretamente a preservação de longo prazo.

Unidade de pesquisa	Classificação	Forma de armazenamento	Metodologia para conservação (em qualquer suporte)
CBPF	Específico para a coleção p/ ano: Classificação Decimal Dewey p/ livros	Específico para a coleção em caixas: estantes	Específico para a coleção: cópias em papel e digitalizadas: higienização e desinfestação por firma especializada
CenPRA	Igual a do acervo geral: papel e digital	Igual a do acervo geral: papel e digital	Igual a do acervo geral: banco de dados
Cetem	CDD		
CGEE	Igual a do acervo geral	Igual a do acervo geral	Igual a do acervo geral
Cnen	Igual a do acervo geral	Igual a do acervo geral. Específico para a coleção: a parte mais recente está armazenada em arquivos PDF e é disponibilizada via intranet. O armazenamento dos arquivos está sendo feita num servidor onde ficam os serviços bibliográficos	Igual a do acervo geral
Finep	Específico para a coleção: seqüencial	Específico para a coleção: seqüencial	Específico para a coleção: microfilme
Ibict	Igual a do acervo geral CDU	As monografias são armazenadas em estantes metálicas, em ordem de classificação; as coleções de periódicos são armazenadas em estantes metálicas em ordem alfabética de título.	
IDSM	Igual a do acervo geral: CDD		
Impa	Específico para a coleção: tabela específica Classificação de assunto do IMPA		
inpa	Igual a do acervo geral	Igual a do acervo geral	Igual a do acervo geral
Inpe	Específico para a coleção (Duas seqüências: 1. geral, 2. pelo tipo de publicação)	Específico para a coleção: arquivos (HTML, Word, PDF)	Específico para a coleção: armazenamento digital em disco rígido; CD-ROM; *Optical Disc*
LNCC		Igual a do acervo geral	Igual a do acervo geral
LNLS	Igual a do acervo geral: Publicações Internas, Comunicação Técnica–CT, Especificações Técnicas –ET, Manual Técnico – MT, Memorando de Projeto –MeT, Notas de cursos e Workshops – NCW, *Preprint*, Relatório de Bolsistas de Verão –RBV, Relatório de Bolsistas –RB, Relatório de Estagiário –RE	Igual a do acervo geral: o documento é registrado na Intranet, recebe um número anual, o texto é anexado e transformado em PDF e disponibilizado para consulta dos usuários. O material é impresso e encadernado, arquivado por ano (ainda não disponibilizamos p/ consulta via *Web*)	
Mast	Igual a do acervo geral	Igual a do acervo geral	Igual a do acervo geral
Mpeg	Igual a do acervo geral e específico para a coleção: acrescenta-se sigla para identificar MGCOL	Base de dados igual a do acervo geral e específica para a coleção	Igual do acervo geral: alguns em armários deslizantes
RNP	Igual a do acervo geral: Classificação Decimal de Dewey, 21. Ed. Categorias documentais, de acordo com nossas necessidades e especificidades de assuntos e documentos.	Igual a do acervo geral: Sistemático, por ordem de classificação.: estrutura de arquivos independentes, em diretórios e disponíveis na intranet e no *website*.	Manutenção de organização dos arquivos eletrônicos armazenados, segundo critérios de categoria, versão e visão.

Quadro 17 – Formas de tratamento da produção científica da unidade de pesquisa

O Quadro 18 indica se as unidades de pesquisa aplicam métodos de preservação aos materiais bibliográficos. As seguintes unidades não responderam a este questionamento: Impa, LNA, MPEG, Mast, Finep e IDSM.

Os responsáveis pelos centros de informação e bibliotecas foram interrogados com relação ao tratamento dado aos materiais em formato digital, uma vez que, nas unidades de pesquisa, são produzidos objetos digitais. Esses podem ser pela conversão de formatos ou pela produção de documentos originalmente digitais.

A pesquisa identificou que aproximadamente metade das instituições realiza alguma forma de tratamento no acervo retrospectivo com vistas a formar coleções digitais para o acesso por parte dos seus usuários.

A digitalização pode ser considerada como o início da vida útil de um documento digital, e por esse motivo as práticas de preservação de objetos digitais poderiam se estender ao planejamento da conversão inicial de formatos. Apesar de que nenhum procedimento de preservação digital pode ser aplicado a materiais analógicos.

Unidade de Pesquisa	Aplicação de metodologia de digitalização	Aplicação de estratégia de preservação	Participação em iniciativa de preservação digital
CBPF	Sim	Não	Não
CenPRA	Sim	Sim	Não
Cetem	Sim	Não	Não
CGEE	Não	Não	Não
Cnen	Sim	Não	Não
CNPq	Sim	Não	Não
Ibict	Sim	Não	Não
Inpa	Não	Não	Não
Inpe	Não	Sim	Sim
INT	Sim	Sim	Não
LNA	Não	Não	Não
LNLS	Não	Não	Não
ON	Não	Não	Não
RNP	Não	Não	Não

Quadro 18 - Métodos de preservação aplicados aos materiais bibliográficos nas unidades de pesquisa

Os dados do Quadro 18 apontam que é reduzido o número de estratégias de preservação sendo implementadas nas agências de pesquisa governamentais.

Mas a complexidade destas estratégias não é um sinal de maturidade na forma como o problema está sendo abordado.

Em 2004, o Serviço e Documentação e Acervo do CNPq informou no questionário que a biblioteca encontrava-se nas condições como foi recebida do MCT, sem nenhuma classificação ou arranjo. Com a criação do Centro da Memória, o CNPq iniciou, naquele ano, um processo de digitalização das atas do Conselho Deliberativo.

Em 2006, o Cetem digitalizou as séries esgotadas das suas publicações em versão impressa. No mesmo ano, a unidade de pesquisa CBPF começou a desenvolver um projeto de preservação digital, que partiu do interesse dos funcionários da Coordenação de Documentação e Informação Científica (CDI), mas a iniciativa não recebeu continuidade. O objetivo do projeto era desenvolver estratégias de preservação de informação digital técnico-científica usando metadados.

Outros projetos de digitalização de acervos estão sendo iniciados nas unidades, mas a organização das novas coleções digitais retrospectivas ainda não foi avaliada. O fato de os centros e bibliotecas usarem soluções proprietárias sugere que a maioria das instituições pesquisadas não demonstra preocupação com o estabelecimento de políticas de preservação digital nem no desenvolvimento de Pacotes de Disseminação de Informação (PDI), como a norma OAIS recomenda. O Inpe foi a única unidade que adotou esse modelo no planejamento da sua biblioteca digital.

Os dados evidenciaram, também, que a informação técnico-científica nessas instituições carece de padronização com relação às atividades de preservação digital. Além disso, a pesquisa mostrou que os documentos digitais estão armazenados em bases de dados proprietárias, sem nenhum plano estratégico que garanta a longevidade da informação neles contida.

Desde 2002, o Ibict trabalha com sistemas de informação que servem para a preservação digital. Já em 2003, o instituto reuniu vários especialistas para discutir o tema da preservação digital e a criação de um comitê. O comitê foi composto por representantes do Ibict, Biblioteca Nacional, Programa Sociedade da Informação, Bireme, Capes, Abec, Unesco e especialistas convidados. Foram criados vários GTs, entre os quais estava o de Certificação Digital. Entretanto, a iniciativa não teve prosseguimento devido a mudanças administrativas e de diretoria.

O Ibict continua interessado em identificar as tendências e apontar as atividades de preservação digital que podem ser aplicadas no contexto nacional de C&T. Foram realizados estudos sobre ferramentas para repositórios institucionais (CAFÉ *et al,* 2003), concretamente, as pesquisas tiveram início mediante a avaliação das ferramentas mais adequadas para essa finalidade. Algumas das ferramentas analisadas foram: Archimede da Layal University Library, o CDSware do CERN, o Fedora das Universidades de Virginia e Cornell e o DSpace do MIT.

O Ibict acompanhou a primeira implementação do DSpace em língua portuguesa, em 2002, realizada pela Universidade do Minho, em Portugal, que resultou na criação do seu repositório institucional (RODRIGUES *et al.*, 2004), o RepositoriUM. A partir de então, o Ibict iniciou atividades de experimentação com este *software*. Em 2004, a Biblioteca do Superior Tribunal de Justiça (STJ) recebeu assessoramento do Ibict para a criação do repositório institucional BDJur, (BASEVI, 2004). Além disto, o Ibict estabeleceu uma parceria com a Universidade de São Paulo (USP) que resultou na criação do Reposcom – Repositórios Institucionais em Ciências da Comunicação (FERREIRA; SOUTO, 2005).

A utilização do *software* DSpace pelo Ibict surgiu com o projeto-piloto denominado Ridi (Repositório Institucional Digital do Ibict) que visava recuperar a documentação de todas as atividades de pesquisa, publicando os conteúdos que se encontravam dispersos em vários departamentos. Em 2007, o Instituto estava trabalhando em uma nova versão do DSpace totalmente em português (Brasil) e baseada na versão mais atual deste *software*. O produto incluiu sua revisão e a criação do “pacote” que está sendo repassado a outras instituições.

Internacionalmente faz-se reconhecida a importância das parcerias na realização de atividades de preservação digital (FEENEY, 1999; LAVOIE, 2008). Elas têm possibilitado melhor compreensão dos benefícios no compartilhamento de custos e avaliações de esforços não redundantes no estabelecimento de estratégias efetivas de preservação.

Nas instituições pesquisadas, a construção da infra-estrutura considerada necessária para esse tipo de projeto ainda não foi analisada, assim como não se estimou a necessidade de contar com o apoio de instituições de fomento e de pesquisa. Essa realidade reproduz, em menor escala, as conclusões do estudo realizado por Hedstrom e Montgomery para o RGL em 1998, ao afirmarem que as instituições participantes desse consórcio não tinham desenvolvido políticas nem

práticas de preservação digital e nem assumido a responsabilidade da preservação dos materiais em formato digital.

A separação entre as propostas de preservação digital que surgem nas iniciativas internacionais e o *status* real da preservação digital em instituições como as unidades de pesquisa começa a ser percebida, mas a insuficiência de recursos e o planejamento inadequado das práticas para permitir o acesso de longo prazo são o maior obstáculo para o estabelecimento de sistemas confiáveis de preservação digital.

4.2.2 Perfil profissional das equipes

O primeiro contato com as equipes dos centros e unidades de pesquisa do MCT se deu por meio da listagem interna que a biblioteca do Ministério dispõe de todos os funcionários desse órgão. Após o envio do questionário, na forma de um documento “circular” da biblioteca do Ministério, as respostas começaram a serem enviadas para o pesquisador e os possíveis respondentes identificados.

O segundo contato ocorreu durante os eventos das associações de ciência e tecnologia de 2004, 2005 e 2007, nos quais se reuniram representantes das unidades de pesquisa. Nesses encontros foi possível ter uma primeira visão do contexto das unidades e dos profissionais que dispunham de informações oficiais sobre os setores de biblioteca do MCT.

As entrevistas realizadas durante os eventos de C&T com os representantes das unidades complementaram as informações coletadas no questionário, mais especificamente sobre os profissionais que realizavam algum tipo de manutenção dos documentos digitais nas unidades.

O levantamento de dados sobre o perfil das equipes dos centros de informação e biblioteca do MCT pretendia identificar as pessoas que fariam parte da amostra de gestores da informação em formato digital do Ministério. Como mostra a Tabela 9, as 20 instituições que enviaram suas respostas definiram suas equipes como sendo integradas por bibliotecários, auxiliares técnicos, analistas e arquivistas, assim como pessoal administrativo.

Embora o questionário apenas permitisse a opção de escolha de resposta entre bibliotecários e auxiliares, os respondentes manifestaram a presença dos profissionais da área da ciência da computação como sendo uma parte importante para o desenvolvimento das atividades desses centros.

O grupo de participantes da pesquisa começou a ser definido após a análise do questionário. Especificamente com relação à população, percebeu-se que havia a necessidade de agrupar os respondentes em três categorias: o primeiro formado pela predominante presença de bibliotecários; o segundo formado pelos auxiliares técnico-administrativos; por último, o grupo dos técnicos de informática e estagiários (estudantes de biblioteconomia, arquivologia e ciência da computação), que apresentaram maior mobilidade com relação ao tempo de atuação nos locais visitados.

Tabela 9 – Servidores nas bibliotecas das unidades de pesquisa do MCT

Unidade de pesquisa	Tipo de biblioteca e servidor				
	Tipo de Biblioteca	Servidores (Bibliotecários)	Servidores (Auxiliares)	Servidores encarregados (Outros)	Total
CBPF	Especializada	4	5	1 (analista)	10
CenPRA	Abrangência geral	1	2	1 (analista)	4
Cetem	Especializada	1	1	2 (analistas)	4
CGEE	Especializada	1	0	1 (analista) 2 (técnicos de informação)	4
Cnen	Especializada	2	1	1 (indexador)	4
CNPq	Central/ Especializada	1	1	1	3
Finep	2 bibliotecas / 2 arquivos	4	3	1 (analista)	8
Ibict	Especializada	2	2	1 (analista)	5
IDSM	Especializada	1	1	0	2
Impa	Especializada	2	2	1 (analista)	5
Inpa	Especializada	1	0	0	1
Inpe	Central / Especializada	6	2	2 (analistas)	10
INT	Especializada	3	0	1 (analista)	4
LNA	Especializada	0	2	0	2
LNCC	Especializada	1	3	1 (analista)	5
LNLS	Especializada	1	0	1 (estagiário)	2
Mast	Especializada	1	1 (administrativo)	1 (arquivista)	3
Mpeg	Especializada	6	4	0	10
ON	Especializada	2	1	1 (analista)	4
RNP	Central/ Especializada	1 (apoio do responsável pela Biblioteca da Embrapa)	2	1 (assistente) 1 (analista)	5

Além dos itens identificados na Tabela 9, nenhuma das unidades mencionou a presença, em suas equipes, de especialistas em preservação digital ou a contratação de consultores para a realização dessas atividades. Isto pode ser explicado pela pouca divulgação no Brasil de consultorias dedicadas especificamente à preservação digital de acervos bibliográficos.

Existe uma única empresa brasileira dedicada a prestar esse tipo de serviços. A empresa TECH-IN, liderada por pesquisadores da área da ciência da informação e arquivologia com sede na cidade de Nova Lima, Minas Gerais, realiza diagnósticos e formula propostas para sistemas de gestão de preservação de

documentos digitais. Os serviços de consultoria de preservação digital que a TECH-IN oferece, estão baseados nos modelos de referências aqui citados e engloba o trabalho de “aculturização” prévia sobre o tema dos profissionais das instituições interessadas (THOMAZ, 2008).

Como será mencionado na análise das entrevistas, algumas mudanças organizacionais das unidades de pesquisa foram decisivas para a formação de profissionais que atendessem às necessidades de preservação dos materiais digitais nos espaços de suas bibliotecas e centros de informação.

4.3 ANALISE DOS DADOS DAS ENTREVISTAS

Os objetivos específicos para a fase de pesquisa de campo propostos neste trabalho foram:

OE3) identificar as necessidades de preservação digital da informação técnico-científica das unidades de pesquisa do Ministério da Ciência e Tecnologia;

OE4) resgatar os conceitos e aplicações usados nessas organizações que caracterizam suas práticas de preservação, tendo por parâmetro as condições locais que interfiram no entendimento do significado e propósito da implantação das metodologias de preservação de documentos digital.

Os resultados obtidos foram a identificação das necessidades de preservação digital da informação científica das unidades de pesquisa do MCT, bem como conceitos usados para caracterizar suas práticas de preservação, segundo os principais envolvidos e as condições locais.

A aplicação do questionário tinha por meta estabelecer um contato inicial com os gestores dos centros de informação e bibliotecas e suas equipes. A abordagem dos dados coletados passou a ser mais qualitativa, assumindo-se que, o entendimento do significado e propósito das atividades de preservação presentes nesses institutos, seria mediado pela interpretação do pesquisador e o seu marco de referencial teórico.

Como Richardson *et al.* mencionam (1999, p. 210), algumas das vantagens da técnica da entrevista estão na postura do entrevistado, que responde com um grau maior de liberdade (diretividade da entrevista) e a expressão de respeito por parte do entrevistador à manifestação da opinião do entrevistado.

As narrativas históricas das memórias das fontes contatadas seguiram uma ordem cronológica que dependeu de fatores específicos de agendamento por parte do entrevistador e dos entrevistados. As entrevistas foram gravadas em fita magnética e MP3 em encontros agendados durante eventos científicos nacionais e em visitas às sedes das unidades de pesquisa.

O primeiro roteiro de entrevista usado nos encontros com representantes das instituições, durante os três eventos científicos nacionais em 2004[44], 2006[45] e 2007[46], pretendia identificar os respondentes e avaliar o conhecimento que eles detinham sobre o tratamento técnico dado aos arquivos digitais nas suas instituições. Também se registraram as opiniões sobre o tipo de material que deve ser preservado, bem como foram identificados os responsáveis localmente.

O roteiro de entrevista com os respondentes nas sedes dos institutos passou a ser orientado segundo o perfil da pessoa contatada. Como mencionado no item 4.2.2, o questionário apontou que os dois grupos a serem abordados nesta fase seriam os profissionais da informação (bibliotecários, arquivistas, técnicos) e os da computação.

A aplicação do segundo roteiro de entrevista foi, portanto, dividido em duas partes dirigidas aos dois setores (biblioteca e informática) representados pelos respectivos profissionais. Nos casos em que a pessoa contatada não era totalmente ligada às atividades de tratamento dos materiais digitais, aplicava-se o primeiro roteiro semi-estruturado. O segundo roteiro era usado apenas nos casos em que a interação se realizou com os respondentes previamente identificados.

Nas respostas aos questionamentos propostos, pretendeu-se detectar novas variáveis que não houvessem sido contempladas no quadro conceitual que a revisão da literatura sugeriu. Os temas e assuntos discutidos encontravam-se relacionados, em menor escala, às expectativas dos responsáveis pelas coleções

[44] 56ª Reunião Anual da SBPC, 2004, Cuiabá, MT.
[45] XIV SNBU Seminário Nacional de Bibliotecas Universitárias, 2006, Salvador, BA.
[46] 58ª Reunião Anual da SBPC, 2007, Belém, PA.

digitais consideradas como materiais importantes para a instituição, considerando o perfil científico e tecnológico.

O Quadro 19 mostra o local das entrevistas com os representantes das 24 unidades de pesquisa e os tipos de entrevistas realizadas no período de 2004 a 2007.

Unidade de pesquisa	Características da entrevista		
	Local da entrevista	Entrevistado	Entrevista
AEB	Arquivo (Brasília, DF) Informática (Brasília, DF) Diretoria (Brasília, DF) SBPC (Cuiabá, MT) SBPC (Cuiabá, MT)	Técnico Técnico Pesquisador Técnico Técnico	Individual Individual Individual Em grupo Em grupo
CBPF	SBPC (Cuiabá, MT) Biblioteca (Rio de Janeiro, RJ) Biblioteca (Rio de Janeiro, RJ) Biblioteca (Rio de Janeiro, RJ) Biblioteca (Rio de Janeiro, RJ)	Técnico Bibliotecária Bibliotecária Técnico Técnico	Individual Em grupo Em grupo Em grupo Em grupo
Ceitec	SBPC (Belém, PA)	Pesquisador	Individual
CenPRA	Biblioteca (Campinas, SP) Biblioteca (Campinas, SP) Biblioteca (Campinas, SP) Biblioteca (Campinas, SP) SBPC (Cuiabá, MT)	Bibliotecária Pesquisador Responsável Pesquisador Pesquisador Técnico	Em grupo Em grupo Em grupo Em grupo Individual
Cetem	Biblioteca (Rio de Janeiro, RJ) Informática (Rio de Janeiro, RJ) Informática (Rio de Janeiro, RJ) SBPC (Cuiabá, MT) SBPC (Cuiabá, MT)	Bibliotecária Responsável Profissional de CI Técnico Pesquisador Técnico	Individual Individual Individual Em grupo Em grupo
CGEE	Centro de Informação (Brasília, DF) Centro de Informação (Brasília, DF) Centro de Informação (Brasília, DF) Centro de Informação (Brasília, DF) SBPC (Cuiabá, MT)	Profissional de CI Técnico Responsável Técnico Técnico Técnico	Em grupo Em grupo Em grupo Em grupo Individual
Cnen	Biblioteca (Rio de Janeiro, RJ) Biblioteca (Rio de Janeiro, RJ) SBPC (Cuiabá, MT)	Bibliotecária Responsável Pesquisador Técnico	Individual Individual Individual
CNPq	Biblioteca (Brasília, DF) Centro da Memória (Brasília, DF) SBPS (Cuiabá, MT)	Bibliotecária Responsável Pesquisador Responsável Técnico	Individual Individual Individual
Finep	Biblioteca (Rio de Janeiro, RJ) Biblioteca (Rio de Janeiro, RJ) Informática (Rio de Janeiro) SBPC (Cuiabá, MT)	Bibliotecária Responsável Bibliotecária Técnico Técnico	Em grupo Em grupo Individual Individual
Ibict	Biblioteca (Brasília, DF) Informática (Brasília, DF) SBPC (Cuiabá, MT)	Bibliotecária Responsável Técnico Técnico	Individual Individual Individual
IDSM	SBPC (Belém, PA) SBPC (Cuiabá, MT) SBPC (Cuiabá, MT)	Técnico Responsável Técnico Técnico	Individual Em grupo Em grupo
Impa	Biblioteca (Rio de Janeiro) Biblioteca (Rio de Janeiro) Biblioteca (Rio de Janeiro) SBPC (Cuiabá, MT)	Bibliotecária Responsável Bibliotecária Técnico Técnico	Individual Individual Individual Individual
Inpa	SBPC (Belém, PA) SBPC (Belém, MT) SBPC (Cuiabá, MT)	Bibliotecária Responsável Técnico Técnico	Em grupo Em grupo Individual
Inpe	SNBU (Salvador, BA) e Biblioteca (São José dos Campos, SP) Biblioteca (São José dos Campos, SP) Biblioteca (São José dos Campos, SP) SPBC (Cuiabá, MT)	Bibliotecária Responsável Bibliotecária Pesquisador Pesquisador	Individual Em grupo Em grupo Em grupo Individual

Continua

Unidade de pesquisa	Características da entrevista		
	Local da entrevista	Entrevistado	Local da entrevista
Insa	SBPC (Belém, PA) SBPC (Belém, PA)	Técnico Pesquisador	Individual Individual
INT	Biblioteca (Rio de Janeiro, RJ) Biblioteca (Rio de Janeiro, RJ) Informática (Rio de Janeiro, RJ) SBPC (Cuiabá, MT)	Profissional de CI Responsável Bibliotecária Técnico Técnico	Individual Individual Individual Individual
LNA	Biblioteca (Itajubá, MG) Biblioteca (Itajubá, MG) Biblioteca (Itajubá, MG) Biblioteca (Itajubá, MG) Biblioteca (Itajubá, MG) e SBPC (Cuiabá, MT)	Bibliotecária Técnico Responsável Pesquisador Pesquisador Técnico	Em grupo Individual Em grupo Em grupo Em grupo Individual
LNCC	Biblioteca (Petrópolis, RJ) Informática (Petrópolis, RJ) SBPC (Cuiabá, MT)	Bibliotecária Responsável Técnico Técnico	Individual Individual Individual
LNLS-ABTLuS	Biblioteca (Campinas, SP) Depto. Científico (Campinas, SP) SBPC (Cuiabá, MT)	Bibliotecária Responsável Pesquisador Técnico	Individual Individual Individual
Mast	Biblioteca (Rio de Janeiro, RJ) Biblioteca (Rio de Janeiro) Biblioteca (Rio de Janeiro, RJ) SBPC (Cuiabá, MT) Arquivo (Rio de Janeiro, RJ) Arquivo (Rio de Janeiro, RJ)	Bibliotecária Responsável Bibliotecária Técnico Pesquisador Técnico Técnico	Individual Individual Individual Individual Em grupo Em grupo
Mpeg	Biblioteca (Belém. PA) SBPC (Cuiabá, MT) SBPC (Belém, PA)	Bibliotecária Responsável Técnico Técnico	Individual Individual Individual
Nuclep	Comunicação (Itaguaí, RJ) Comunicação (Itaguaí, RJ) Informática (Itaguaí, RJ)	Pesquisadora Técnico Técnico	Em grupo Em grupo Individual
ON	Biblioteca (Rio de Janeiro, RJ) SBPC (Cuiabá, MT)	Bibliotecária Técnico	Individual Individual
RNP	Sala de Reuniões (Rio de Janeiro, RJ) Sala de Reuniões (Rio de Janeiro, RJ) Sala de Reuniões (Rio de Janeiro) SBPC (Cuiabá, MT)	Profissional de CI Responsável Técnico Técnico Técnico	Em grupo Em grupo Em grupo Individual

Quadro 19 – Listagem de entrevistados das unidades de pesquisa do MCT

Foram entrevistados 19 responsáveis de 18 centros de informação e biblioteca sendo que, um instituto contava com dois responsáveis, um pela biblioteca e outro pelo Centro de Memória (CNPq). Apenas dois responsáveis pelas bibliotecas e centros de informação visitados não foram encontrados para a realização das entrevistas.

Uma característica das 90 entrevistas realizadas é que 54 delas foram individuais e 36 em grupo. Quatorze dos 19 responsáveis pelos centros de informação e bibliotecas foram entrevistados individualmente, sendo que um deles foi entrevistado também em grupo (Inpe). Os outros cinco responsáveis foram abordados em entrevistas em grupos reunidos durante os encontros científicos e nas visitas aos locais de trabalho das 88 pessoas participantes da pesquisa.

O desenvolvimento das cinco entrevistas em grupo não foi definido previamente pelo pesquisador devido, a que elas aconteceram de forma espontânea por decisão dos próprios entrevistados na hora e local de abordagem.

Apesar de 11 dos entrevistados de quatro das unidades de pesquisa visitada não possuírem centro de informação ou biblioteca (AEB, Nuclep, Insa e Ceitec), os respondentes foram caracterizados como representantes de setores-chave das instituições procuradas, a saber: diretorias, arquivo, informática e comunicação.

4.3.1 Perfil profissional dos entrevistados

A caracterização dos entrevistados tem em comum o escopo das suas atividades voltado ao tratamento da informação técnico-científica, embora houvesse casos em que as funções de alguns pudessem ser classificadas em um ou mais setores. Por exemplo, um pesquisador ocupar um cargo na área de informática e ao mesmo tempo colaborar ativamente no desenvolvimento de projetos de gerenciamento de informação (Inpa e CNEN) ou, no caso inverso, um bibliotecário lidar com a infra-estrutura da montagem de um servidor de dados de pré-*prints* na instituição (Impa).

Como mostrado no item sobre o universo da pesquisa (3.2), foram entrevistados 19 responsáveis pelas bibliotecas e centros de informação de 20 unidades que contam com esse serviço (o CNPq conta com dois responsáveis, um pela biblioteca e outro pelo Centro de Memória). Estes estavam incluídos no grupo de 88 pessoas entrevistadas neste estudo.

O Quadro 20 disponibiliza a listagem completa de profissionais e pesquisadores abordados durante a pesquisa de campo, ordenados segundo o setor onde atuam.

Unidade de pesquisa	Setor do instituto	Profissão/função
AEB	Arquivo Informática Comunicação Comunicação Diretoria	Técnico Técnico de informática Técnico Técnico Pesquisador
CBPF	Biblioteca Biblioteca Informática Informática Comunicação	Bibliotecária Bibliotecária Técnico de informática Técnico de informática Técnico
Ceitec	Comunicação	Pesquisador
CenPRA	Biblioteca Biblioteca Pesquisa Pesquisa Comunicação	Bibliotecária Pesquisador Pesquisador Técnico Técnico
Cetem	Biblioteca Informática Informática Comunicação Comunicação	Bibliotecária Cientista da Informação Técnico de informática Pesquisador Técnico
CGEE	Centro de Informação Centro de Informação Centro de Informação Informática Comunicação	Cientista da informação Técnico de informação Técnico de informação Técnico de informática Técnico
Cnen	Biblioteca Informática Comunicação	Bibliotecária Pesquisador Técnico
CNPq	Biblioteca Centro da Memória Comunicação	Bibliotecária Pesquisador Técnico
Finep	Biblioteca Biblioteca Informática Comunicação	Bibliotecária Bibliotecária Técnico de informática Técnico
Ibict	Biblioteca Informática Comunicação	Bibliotecária Técnico de informática Técnico
IDSM	Serviço de Informação Comunicação Comunicação	Técnico de informação Técnico Técnico
Impa	Biblioteca Biblioteca Informática Comunicação	Bibliotecária Bibliotecária Técnico de informática Técnico
Inpa	Biblioteca Comunicação Comunicação	Bibliotecária Técnico Técnico
Inpe	Biblioteca Biblioteca Pesquisa Comunicação	Bibliotecária Cientista da Informação Pesquisador Pesquisador
Insa	Comunicação Diretoria	Técnico Pesquisador
INT	Biblioteca Biblioteca Informática Comunicação	Bibliotecária Bibliotecária Técnico de informática Técnico
LNA	Biblioteca Biblioteca Informática Pesquisa Comunicação	Bibliotecária Técnico de informação Pesquisador Pesquisador Técnico
LNCC	Biblioteca Informática Comunicação	Bibliotecária Técnico de informática Técnico

Continua

Unidade de pesquisa	Setor do instituto	Profissão/função
LNLS-ABTLuS	Biblioteca Depto. Científico Comunicação	Bibliotecária Pesquisador Técnico
Mast	Biblioteca Biblioteca Informática Comunicação Arquivo Arquivo	Bibliotecária Bibliotecária Técnico de informática Pesquisador Técnico Técnico
Mpeg	Biblioteca Editoração Comunicação	Bibliotecária Técnico Técnico
Nuclep	Comunicação Comunicação Informática	Pesquisadora Técnico Técnico
ON	Biblioteca/DID Comunicação	Bibliotecária Técnico
RNP	Biblioteca/DGI Informática Informática Comunicação	Cientista da Informação Técnico de informática Técnico de informática Técnico

Quadro 20 – Entrevistados por setor das unidades de pesquisa do MCT

No decorrer das entrevistas tornou-se evidente que os respondentes trabalhavam em setores diferentes do mesmo instituto e em projetos relacionados com o tratamento da informação científica. Além dos setores de biblioteca e arquivo, que seriam os lugares indicados para os trabalhos de preservação digital, os respondentes eram oriundos dos setores de informática, comunicação, editoração, pesquisa relacionada à temática da instituição e diretorias.

A variedade de locais de trabalho dos participantes mostra que a técnica usada para obter informação procurou se beneficiar dos anos de experiência dos profissionais e conseguir, dessa maneira, uma perspectiva mais ampla do que a dos responsáveis pelos acervos digitais.

Um grupo de pessoas que não haviam sido identificadas por meio do questionário foi o formado por 14 pesquisadores. Tratava-se de pessoas que tinham desenvolvido ou realizavam projetos relacionados com o setor de biblioteca da instituição chegando, em alguns casos, a serem responsáveis pelo centro de informação (CenPRA e CNPq). Os demais mantinham participação nos setores de informática, comunicação e pesquisa. Eles contribuíram com muitas informações contextuais, mas para este estudo foram utilizadas apenas aquelas ligadas ao tema da preservação digital.

Os dados coletados e as opiniões dos entrevistados permitiram estabelecer um tipo de padrão no quesito necessidade de práticas de preservação

digital, que não havia sido detectado nos questionários, e a definição de novos níveis de aprofundamento que pudessem ser alcançados por meio da observação e da análise etnográfica de conteúdo.

4.3.2 Análise dos conteúdos das entrevistas

A metodologia de análise do conteúdo das entrevistas levou em consideração a necessidade de organizar os respondentes em grupos para que, dessa forma, fosse detectado o eixo comum nas respostas dadas aos questionamentos. No Quadro 21 estão apresentados os respondentes, segundo a função e a instituição de origem.

Função	Unidade de Pesquisa	Total
Arquivista	Mast, AEB	3
Bibliotecário	Impa, Inpe, CBPF, Mast, ON, CenPRA, Finep, LNA, INT	10
Bibliotecário responsável	Ibict, Impa, Mpeg, Inpe, Cnen, LNCC, LNLS, Mast, Finep, CNPq, Cetem, Inpa	12
Pesquisador	Inpe, Cnen, LNLS, Mast, CenPRA, LNA, Cetem, AEB, Nuclep, Insa, Ceitec	14
Pesquisador responsável	CenPRA, CNPq	2
Profissional de CI	Cetm, CGEE	2
Profissional de CI responsável	RNP, INT	2
Técnico	Ibict, Impa, Mpeg, CBPF, Cnen, LNCC, LNLS, ON, CenPRA, Finep, CNPq, LNA, RNP, Cetem, INT, Inpa, CGEE, IDSM, AEB, Nuclep, Insa	24
Técnico de biblioteca	Impa, Mast, CGEE	3
Técnico de informática	Ibict, CBPF, LNCC, Mast, Finep, RNP, Cetem, INT, CGEE, AEB, Nuclep	13
Técnico responsável	LNA, CGEE, IDSM	3

Quadro 21 - Função dos entrevistados nas unidades de pesquisa

As respostas dos 88 entrevistados foram agrupados em cinco categorias de respondentes definidas, segundo a esfera de ação profissional de origem, a saber:

1. grupo 1: formado por bibliotecários responsáveis pelas bibliotecas, bibliotecários auxiliares, profissionais da ciência da informação e técnicos de biblioteca (total = 28 pessoas);
2. grupo 2: integrado por técnicos responsáveis pelas bibliotecas e técnicos em geral de diferentes áreas (total = 27 pessoas);
3. grupo 3: formado por pesquisadores de várias áreas do conhecimento responsáveis por bibliotecas e pesquisadores em geral (total = 17 pessoas);
4. grupo 4: com técnicos de informática (total = 13 pessoas);
5. grupo 5: formado por 3 arquivistas.

Constatou-se que a maioria dos respondentes do grupo 2, que tinham funções e cargos de técnicos, foram os que evidenciaram maior participação nas entrevistas em eventos científicos. Os dois técnicos responsáveis (CGEE e LNA) e os membros dos outros quatro grupos foram abordados principalmente nas entrevistas nas sedes das instituições. As respostas das entrevistas em grupo foram transcritas, e as manifestações de cada um dos respondentes separadas de acordo com essa tipologia.

4.3.2.1 Sobre as próprias funções e atividades de preservação

Na cronologia das entrevistas foi possível constatar que os contatos com os grupos ficaram muito dirigidos às características locais, e que os representantes desses grupos de entrevistados não tiveram dificuldade em mencionar algumas das suas preocupações com relação ao material científico que deveria ser preservado e sobre as necessidades da existência de projetos de preservação digital.

> - Os projetos do instituto têm muita documentação que não está registrada. São informações que são importantes para a história da instituição. Nós estamos preocupados não com o número, mas na preservação da memória. Os editais agora mencionam a preservação da memória, tem projetos no CNPq, no Mast e no INT, mas você tem que ter profissionais com essa visão da necessidade. No Mast estão preocupados não apenas em preservar, mas em ensinar às pessoas sobre a preservação. (Bibliotecária responsável 1)

> - Não temos atividade de preservação. A gente até ficou preocupada. Nós atuamos em uma área em que tínhamos um grande especialista, mas essa pessoa faleceu e muitas coisas ficaram com ele, as informações não foram repassadas. Hoje a preocupação existe em termos não de guardar, mas que exista um processo de continuidade do conhecimento, de a gente repassar o que produziu. (Pesquisador geral 2)

> - Não temos essa discussão de preservação de longo prazo. Essa competência foi distribuída para cada secretaria de coordenação; o pesquisador pertence a essa secretaria, isso é repassado para as bibliotecas da área e nossa intenção é que estejam disponibilizados eletronicamente outros trabalhos dos pesquisadores, por isso uma comissão foi constituída para ter mínimas regras dessa distribuição, é uma comissão de pesquisadores. Se cada instituição tiver uma definição de como vai trabalhar com isso, ela vai ter as coisas mais arrumadas. (Técnico geral 12)

> - Não é necessário fazer nenhuma atividade de preservação do material científico publicado por que ele pertence a grandes bancos internacionais e não precisamos mantê-lo. A única preocupação que eu tenho é de manter o registro da aparição do Laboratório na mídia, a parte da pesquisa não é necessária. Mas nós temos tudo muito bem catalogado, porque quando um astrônomo pede tempo de uso de telescópio, nós fazemos uma triagem de como tem sido o aproveitamento desse tempo, e se ele está publicando com esses dados ou não. A biblioteca faz essa triagem e as comissões de pedido de tempo têm uma secretaria que concentra num banco de dados todo o uso de todos os pesquisadores envolvidos em projetos. Temos outro banco de dados de manuais técnicos preservado na nossa instituição. (Pesquisador geral 3)

> - De 2000 para cá já está tudo digitalizado, as teses e os artigos, os trabalhos apresentados em colóquios são entregues em formato digital. A política foi de que, a partir de 2000, tudo tem de estar em formato digital. O que for antes disso, decidiu-se que não ia se fazer nada, apenas guardar em papel. Porque achou-se que não valia a pena. Não existe nenhum projeto de preservação, até houve orçamento numa época, feito para uma empresa que estava interessada em fazer. Mas eles não se interessaram, acharam que as coisas com mais qualidade acabam sendo publicadas em revistas e que são preservadas de alguma forma. (Técnico geral 13)

> - Aqui no Centro estamos caminhando para todo mundo ter os mesmos formatos e catálogo entre as unidades vinculadas. Mas aqui não temos nenhuma coleção da biblioteca que esteja digitalizada, tudo está em papel. O projeto Memória está começando, tudo o que foi escrito pelos pesquisadores daqui e os que os estrangeiros que escreveram sobre o Centro. O portal que vai ser lançado neste ano vai ter um *link* para a produção científica, com o que seria uma base de dados com referências aos trabalhos desses autores. (Bibliotecária responsável 3)

> - Em termos de preservação das coleções eletrônicas nunca pensei nisso, ninguém nunca me cobrou nada a esse respeito. (Técnico geral 12)

A percepção e a confirmação de que o lugar das atividades de preservação seria na biblioteca estiveram presentes nas falas, mesmo naquelas em que esse setor não existia.

> - O responsável pela preservação dos relatórios dos projetos é o Setor de Documentação. (Técnico geral 2)

> - A responsável pela preservação é a biblioteca. (Técnico geral 4)
>
> - A responsável pela guarda da produção é a Diretoria Técnico Científica. (Técnico geral 5)
>
> - A responsável pela preservação da produção científica é a Biblioteca. Cada ano temos um ciclo de cursos e palestras. (Técnico geral 11)
>
> - Eu acho que tudo que nós fazemos deve ser preservado, a memória técnica é muito importante. As coisas mais antigas ainda estão em papel, isso deveria ser tratado e disponibilizado, se bem que algumas coisas não deveriam ser disponilizadas porque são projetos terceirizados. Mas é importante que eles sejam preservados pela biblioteca porque até dentro do próprio desenvolvimento do centro isso não pode ser disponibilizado em congressos, mas elas podem estar disponíveis para outros pesquisadores que querem utilizar essas informações para determinado estudo. (Pesquisador geral 2)

Como outras pesquisas internacionais mostraram (WHEATLEY; BEAGRIE; BRADLEY, 2004), as organizações não garantem recursos específicos para a preservação digital, sendo isso considerado uma das maiores ameaças à permanência dos documentos digitais. Os comentários sobre a realidade das bibliotecas e centros de informação apontaram as condições do serviço público como obstáculo para a continuidade dos trabalhos de preservação. Isso ficou evidente nas observações de 80% dos entrevistados.

> - Não existe um compromisso de guardar informações. Principalmente hoje, com esse nível de êxodo que está atingindo o serviço público, não se consegue repor o pessoal para ter o controle melhor das coisas. Cada ano que passa parece que mais pessoas têm de fazer o papel de várias pessoas ao mesmo tempo. Isso prejudica o trabalho, a qualidade do trabalho e o nível da informação que está sendo preservada. (Técnico geral 3)
>
> - No Museu temos 30 anos de acervo do CNPq, nós disponibilizamos, criamos um livro e um CD-ROM, pessoas que querem saber a história do CNPq consultam lá. Mas somos poucas pessoas para processar tanta informação. (Técnico de biblioteca 1)
>
> - A biblioteca foi desativada faz tempo, mais ou menos seis anos, ela está sendo reativada aos poucos, mas perdeu-se muita coisa dentro dessa paralisação no referente à renovação de publicações, por exemplo. (Técnico geral 3)
>
> - Nós não temos biblioteca, quem cuida das informações é a área da comunicação social, quem guarda é cada diretoria, a guarda não é centralizada. Tudo está em formato digital. Dentro da área da administração temos uma pessoa que é responsável por guardar todas as informações em papel. (Técnico geral 10)
>
> - No Centro cada grupo de pesquisa tenta conseguir financiamento, e eles são responsáveis pela preservação dos equipamentos e pela documentação desses projetos. (Técnico geral 8)

As atitudes dos entrevistados com relação a essas barreiras apresentam uma noção geral de que a informação produzida nessas unidades deveria ser acessada, também identificaram os desafios de trabalhar em colaboração com outras unidades e centros internacionais de pesquisa.

> - O que importa é que de fato há uma preocupação em atender de maneira rápida qualquer pedido de artigo científico de uma pesquisa que tenha sido realizada no Laboratório. (Técnico geral 4)
>
> - A política do Instituto sempre foi de divulgação de seus resultados. Recebemos 12 mil pessoas anualmente. (Bibliotecário responsável 1)
>
> - O projeto da Memória da Ciência do CNPq foi do vice-presidente do CNPq, que na época esteve no Museu antes de montar esse projeto. Ele esteve conversando com os pesquisadores, pediu textos, e com essas idas ao Museu e textos ele fortificou a idéia para desenvolve-la no CNPq. Se esse trabalho hoje existe, a base dele foi toda pega no Museu de Astronomia. (Técnico de biblioteca 1)
>
> - Nós interagimos com pesquisadores, eles entram no nosso site e encontram informações que já temos armazenadas em fita. Quando existir o observatório Virtual Brasileiro, que vai estar conectado ao Observatório Virtual Mundial, aí sim vai haver instituições colaborando, mas será um trabalho gigantesco porque vai exigir que padronizemos dados com outros países. (Bibliotecária auxiliar 1)
>
> - Todos os memorandos técnicos, além de estarem no formato em papel e digital, *backups* estão disponíveis em outras instituições. Toda a produção técnico-científica está em linha; os cientistas deixam sua produção depositada na biblioteca. (Técnico geral 4)
>
> - A biblioteca está sendo informatizada para digitalizar o acervo de teses; ainda não temos isso, não temos orçamento, nem gente suficiente. O Observatório sedia e coordena um banco de dados ambientais para a indústria, em um projeto que envolve outras instituições de ensino e pesquisa. (Técnico geral 11)

Em entrevistas com pesquisadores e técnicos participantes de projetos de preservação digital, fica evidente a dificuldade encontrada por eles, pelo fato de ser uma iniciativa que começou por interesse pessoal.

> - Em 1995, um pesquisador localizou um *software* para montar um simpósio e descobriu que ele podia ser usado para uma biblioteca digital. Migramos todos os registros que estavam no Microisis com toda a coleção dos anos 1960 a 2006, com 12.400 registros; 6 mil eram texto completo. Alguns foram digitalizados a partir de 2000. Com a obrigatoriedade de depositar a tese na Biblioteca Digital, já temos 583 teses. Agora a Biblioteca Digital, além de ser um repositório que guarda a memória científica, serve para o planejamento de gestão do instituto. Temos tudo digitalizado de cada área do Programa Espacial. E agora estamos entrando na gestão dos arquivos. (Bibliotecária responsável 1)
>
> - Todo o meu tempo, como você pode imaginar, foi absorvido para a montagem desse *software*; agora talvez, com muita coisa finalizando, eu vou ter mais tempo para fechar uma espécie de pacote; ele já esta disponível na *web* desde o início, o problema é personaliza-lo para isso que

> você precisa. A pessoa precisa ter conhecimento do sistema operacional, se vai instalar em Linux, porque em Linux existem problemas de permissões, e isso é delicado. Depois o *software* foi desenvolvido em linguagem TCL, o próprio banco de dados eu fiz no TCL uma linguagem muito simples e limpa, comparada com PERL ou PHP, ela é muito enxuta e rápida, mesmo sendo uma linguagem interpretada que normalmente é mais lenta. Eu não esperava que fosse funcionar, quando comecei a usar TCL, eu não tinha idéia do potencial, mas até hoje está sendo bem útil. (Pesquisador geral 1)
>
> - No Instituto temos outros profissionais que trabalham com preservação digital. Tem o professor Bergamini, que trabalhou com a professora Katia Thomaz sobre a norma OAIS. Esse trabalho foi o apêndice da tradução da norma ISO. O professor Banon fez a revisão que faz parte do apêndice. (Bibliotecária responsável1)
>
> - O Laboratório começou para valer a fazer ciência em julho de 1997, quando o equipamento entrou em operação. As áreas que vieram depois são muito recentes, tem a biologia molecular estrutural e as nanociências, todas muito novas e com pouco acervo. (Técnico geral 4)
>
> - A área é muito específica, a biblioteca não tem uma demanda muito grande do público externo, quem procura são poucas pessoas e as bibliotecárias dão conta de fazer esse serviço. Fazem o controle interno da memória técnica quanto aos pedidos dos pesquisadores. Elas são responsáveis pela divulgação da produção científica (Técnico geral 6)
>
> - Eu acho que uma das coisas mais difíceis de você determinar, igual em uma casa, o que é que você vai guardar, o que é prioritário. Como são produções de indivíduos, a minha produção sempre vai ser mais importante para mim, para te dizer a verdade. É uma coisa humana. (Técnico geral 12)

4.3.2.2 Sobre o tratamento dos objetos digitais na instituição

Os entrevistados informaram os formatos dos arquivos que os pesquisadores usam mais frequentemente na produção dos documentos digitais, os quais seriam: PDF, HTML, *Word*, PPS, LATEX, JPG, GIF. As referências a algum tipo de tratamento desses objetos digitais evidenciaram sugestões de soluções tecnológicas próprias, sendo que a maioria delas não considera a necessidade de concordância com os modelos de preservação digital OAIS e TDR.

> - Os documentos estão em um banco de dados administrado por uma equipe de pesquisadores e coordenadores. A Diretoria Técnico-Científica é a encarregada desse arquivamento da produção para ser publicada também. Ela tem diversas linhas de pesquisa e cada programa de pesquisa tem seu arquivo. Tudo está em formato digital, impresso e em arquivos em CD e disquete. (Técnico geral 5)
>
> - Toda a documentação que se produz na área de engenharia é armazenada em um sistema nosso de documentação. Na área da segurança também acontece, um trabalho operacional e de rotina é

> documentado e armazenado no nosso sistema interno. A gente quer resgatar algumas informações da história da Rede e publicar, porque para nós isso faz muita falta, e os próprios usuários desconhecem o conteúdo do sistema. (Técnico geral 7)

> - Na biblioteca do Centro nós temos um sistema que é a memória técnica desenvolvido por um analista que esteve aqui como bolsista, e todos nossos relatórios técnicos, depois de digitalizados, passam pela biblioteca . O pesquisador entrega em formato Word e a secretária da coordenação transforma em formato PDF e chega para a biblioteca para padronização. A memória técnica começou em 2002, mas está apenas na intranet. Os dados referenciais desde 1978 foram migrados, mas o texto completo ainda não. Tínhamos um projeto de digitalização, mas nunca chegou verba para isso. (Bibliotecária responsável 2)

> - Antes tudo era em papel. Agora a maior parte da produção científica está em formato digital. O Instituto tem uma biblioteca digital com registros desde 1984, onde temos todas as publicações editadas pelo Instituto. Temos um arquivo dos dados desde que o Instituto foi criado. Ele nasceu como um repositório, um serviço de divulgação. No final dos anos 1980, usamos o Microisis da Bireme para melhorar aquela base, nela a produção científica passou a ter resumo e referências. Nessa mesma época passamos a trabalhar toda a produção retrospectiva. Desde 2001, temos uma tabela que todo documento depositado precisa seguir; ele vai pela área, pelo grupo e pelo tipo. Assim, o que vai aparecer para o usuário é o artigo e a apresentação do material audiovisual, caso seja um trabalho em evento. (Bibliotecária responsável 1)

Como destacado na revisão da literatura, as abordagens tecnológicas de preservação estão direcionadas ao acesso permanente (THIBODEU, 2002). No Brasil, o Inpe se destaca pelo número de soluções propostas para a preservação digital implementada, tanto por tipo de material como pelo tipo de mecanismo que garante o acesso permanente aos documentos.

> - Quando quis integrar o trabalho dos grupos de pesquisa, eu imaginei o identificador e a partir daí construí a biblioteca digital, se não tivesse esse identificador não teria essa biblioteca. Primeiro usamos o *e-mail* como identificador, mas agora passou a ser usado o nome do computador. A diferença com o DOI e com o Handle é que você não depende de acordo com ninguém. Você pode identificar seus identificadores e não existe conflito. Qualquer um pode usar o identificador persistente, apenas deve seguir uma regra muito simples, basta seguir o nome do domínio, nome da máquina, *e-mail* da pessoa ou domínio do computador. Com isso você cria uma identidade para cada trabalho depositado. O sistema permite a estruturação dos acervos. Podem ser reagrupados dois acervos. Você pode encapsular, por exemplo, teses com dissertações. Com os identificadores, pode-se achar o documento que foi encapsulado. (Pesquisador geral 1)

> - A Biblioteca Digital está guardada hoje em 12 computadores; compramos um servidor Sun em 1998, com orçamento da Fapesp, e a partir daí decidimos centralizar os servidores que tem cada um dos serviços da BD, sendo criada a subdivisão dos serviços. Não temos espelhamento, propomos isso ao MCT desde 2004, mas não houve resposta. Temos uma máquina de segurança que o Grupo de Segurança do Inpe teve a preocupação de criar. (Bibliotecária responsável 1)

> - Também temos o acervo de pre-*prints,* que é uma tendência interessante; estamos fazendo o trabalho e ele fica preservado. Se o autor decide o que vai fazer com ele, primeiro deposita uma cópia pré-*print*, e ela vai ser preservada quando ele for publicado. (Bibliotecária responsável 1)
>
> - As fotografias estão em parte digitalizadas, a história oral está guardada em DVD e estamos com um estudo pronto para a produção multimídia. (Bibliotecária responsável 1)

Para evitar que as informações sejam corrompidas, as técnicas de preservação devem iniciar no momento da criação do objeto digital (HODGE, 2000). Nas unidades de pesquisa, a sistematização dos processos de tratamento dos arquivos digitais começa a ser aplicado, ainda que a maioria contemple apenas a parte que concerne ao depósito de documentos.

> - Toda a produção está em formato digital. (Técnico geral 1)
>
> - Em termos de publicações, a única coisa que fizemos é a partir de certo momento começar a digitalizar tudo o que podia. (Pesquisador geral 2)
>
> - A gente não recebe nada em papel tudo vem através de formulário eletrônico desde 2002. Temos "n" procedimentos só via formulário na página eletrônica. Não aceitamos que ninguém entregue disquete ou pelo correio. Pelo menos já evoluímos bastante, os pesquisadores do laboratório já fazem tudo no formato Latex específico para essa área, o que sistematizou bem o depósito dos trabalhos. (Técnico geral 4)
>
> - Todas as séries estão em formato digital e os livros estão sendo escaneados e convertidos para PDF e disponibilizados nas páginas do Centro. Os relatórios estão sendo armazenados em disquete, em CD. (Técnico geral 6)
>
> - No CNPq não se trabalha mais com papel, todas as informações de resultados de pesquisa ficam armazenadas em formato eletrônico. (Técnico geral 9)

Constatou-se, também, a noção geral existente entre os técnicos de informática, de que a preservação digital seria apenas alguma coisa relacionada com a manutenção de cópias de segurança.

> - O Serviço de Documentação e Arquivo recebe e a Informática faz uma cópia de segurança. Os responsáveis pela preservação seriam as áreas técnicas, mas futuramente passará para a Informática. (Técnico informática 9)
>
> - Os *backups* são feitos basicamente em fitas, e elas são fisicamente guardadas em outro local não perto do servidor. Vai ter um prédio para fazer isso, temos uma outra máquina que faz um *backup on-line* e está em outro bloco do instituto. (Técnico informática 1)
>
> - Para preservação temos duas máquinas, uma na biblioteca, outra fora do prédio, na área de segurança de TI. Três vezes por semana são feitas as atualizações (*backup*) que ficam em outro prédio. (Técnico informática 10)

> - Quem cuida do banco de dados tem a obrigação de efetuar todo o procedimento de segurança e migrar para outro banco de dados. O servidor já faz automaticamente duas vezes por dia, e uma por mês, e nós, em cima disso, fazemos periodicamente, colocando em outras máquinas. Deveria ser em outro prédio, mas hoje fica tudo na nossa divisão. Estamos fazendo cópias em CD e em DVDs. (Técnico informática 11)

A necessidade de sistematizar os processos que envolveriam a preservação da informação científica nas unidades foi percebida em poucos implementadores dos sistemas de gerenciamento de informação entrevistados. Também, alguns profissionais da informação estão preocupados com o uso extensivo de tecnologias e com seu registro em soluções cada vez mais complexas. No Inpe, a concepção desses sistemas esteve influenciada pela percepção de repositórios que seguem o modelo OAIS, segundo os principais participantes dos projetos.

> - Nossa biblioteca Digital pode ser exportada por meio do protocolo OAI. Isso completaria toda a parte de gerenciamento de uma biblioteca digital, a segurança, a identificação persistente e a interoperabilidade. Quando eu tomei conhecimento, primeiro foi um empurrão do Ibict, recebemos um carta do Ibict para ver se a gente poderia importar os dados, aí eu fui ver o *site* do OAI e lá estava bem explicado. A gente conseguiu implementar isso em um mês. Desde 2003, fomos dos primeiros a integrar a Biblioteca Digital de Teses e Dissertações (BDTD). O conteúdo da BD poderia ser coletado também pelo coletador de repositórios OASIS.BR do Ibict. (Pesquisador geral 1)

> - A gente ainda está em busca do modelo ideal de sistematização de informação da produção científica e de como indexar essa informação, porque os próprios pesquisadores às vezes adotam modelos americanos, europeus. (Técnico geral 4)

> - O *software* do Inpe tem muito a ver com o modelo OAIS. Eu observei que o OAIS tinha muito conceito que já existia na nossa estrutura de biblioteca digital, e estamos reutilizando essa metodologia. Nosso sistema é parte do OAIS. O OAIS é muito sofisticado. (Pesquisador geral 1)

Como foi mostrado na revisão da literatura (LAVOIE; DEMPSEY, 2004), os sistemas de preservação digital estariam divididos em níveis, sendo que um deles incluiria os serviços relacionados com permissões de acesso, disseminação e tudo que fosse ligado às políticas. Nas unidades de pesquisa do MCT, a necessidade de políticas relacionadas com a preservação dos documentos digitais foi percebida apenas no Inpe.

> - A parte da submissão foi criada com políticas de prazos a cumprir, pois temos prazos para responder a perguntas estratégicas; mas a busca da produção não é uma coisa simples. Embora tenhamos essa política a cultura não se estabelece, você tem que ir no corpo a corpo com o pesquisador e sempre tem que estar perguntando ou checando o Web of Science. A política não é garantia de que o cientista vai depositar seu documento. Estamos criando um serviço personalizado no *site* onde está

> uma planilha usada na biblioteca digital para ele fazer o depósito lá. Nós estabelecemos os campos das informações daquilo que não tirar o tempo do pesquisador, depois a biblioteca completa as informações que faltam. (Bibliotecária responsável 1)
>
> - O Inpe sempre teve políticas de editoração e de preservação, com normas técnicas para publicação e depósito. Nós (Inpe) temos uma política de preservação, sempre existiu, mas ela depende do Conselho de Editoração. A nova versão que está sendo aprovada menciona três coisas: a biblioteca digital, o manual de editoração e a obrigatoriedade de depósito para os autores. Essa obrigatoriedade já existia, mas não era explicita. (Pesquisador geral 1)

Nas instituições pesquisadas, além de poucas contarem com pessoal especializado no tratamento de arquivos digitais, foi detectada a ausência de um consenso claro sobre quais seriam as estratégias e os métodos de preservação digital a serem aplicados; esses fatores aparecem como barreiras para a efetivação de ações concretas dentro do Ministério.

As explicações para o desenvolvimento lento de projetos de preservação digital, na área do gerenciamento da informação científica em agências governamentais no Brasil, estariam associadas à falta de uma aplicação direta de políticas e tecnologias que sigam padrões internacionais e da formação de recursos humanos.

Finalmente, por meio das entrevistas foi possível confirmar a existência da iniciativa individual e pioneira do Inpe, mesmo com a insuficiência de recursos e o planejamento fragmentado de atividades de preservação digital dirigidas pelo seu Serviço de Informação e Documentação.

4.4 ANÁLISE ETNOGRÁFICA DE CONTEÚDO

O acompanhamento das diversas etapas administrativas e técnicas que envolvem o gerenciamento dos registros digitais permitiu observar as condições sociais da sua produção e sua comparação com os projetos de preservação digital propostos na literatura. Tal análise pode ser efetuada no âmbito da ciência da informação usando uma técnica de outra área, como mais uma evidência da sua convergência com outras disciplinas e atividades.

Dentre as ciências sociais, a antropologia é aquela que tem oferecido mais exemplos de como a imprecisão permeia todas as formas de avaliação da realidade (GEERTZ, 1988). O levantamento mais perfeito, o mais rigoroso teste estatístico depende do significado que as palavras, categorias e até números transmitem. A pesquisa etnográfica é uma das técnicas de análise qualitativa mais aproveitada por diversas áreas do conhecimento na busca por interpretações científicas da realidade (CULYBA; HEIMER; PETTY, 2004).

A técnica Análise Etnográfica de Conteúdo (AEC) foi concebida na área de ciências da comunicação, como abordagem para o estudo da origem e uso de conceitos formulados nas mídias. Na área de jornalismo, as pesquisas que aplicaram a técnica assumiram como propósito a formulação de conceitos em notícias sobre eventos transmitidos nas mídias como jornais. Os documentos ou materiais divulgados são definidos como representações simbólicas de informações ligadas a situações complexas.

Por ser um tipo de Análise Qualitativa de Documentos (AQD), a AEC envolve exemplos teóricos e emergentes de documentos de bases de informação, desenvolvimento de protocolos para análise sistemática e constantes comparações. Diferentemente da análise de conteúdo tradicional, a AEC interage com as informações sem ficar restrita a amostras.

A pesquisa qualitativa de análise etnográfica de conteúdo usa métodos de investigação indutivo e dedutivo (ALTHEIDE, 1987), combina a observação com outros métodos para “triangulação” dos dados, para garantir a completa e acurada descrição de relações e interações que ocorrem entre membros de um grupo (EWING *et al*., 2003). Entre as vantagens dessa técnica, citam-se as seguintes:

a) observações não precisam ser estruturadas ao redor de uma hipótese;
b) em termos de validade, junto com a observação, ela pode servir para coletar informação profunda sobre determinado tema ou comportamento;
c) preocupa-se com uma realidade que não pode ser quantificada, buscando a compreensão e não a explicação.

A análise etnográfica de conteúdo, definida como “[...] *the reflexive analysis of documents* [...]*”* (PLUMMER, 1983), faz parte do grupo de modalidades de pesquisa qualitativa que surgiu na segunda metade do século XX, junto com a

observação participante e a *grounded theory* (DENZIN; LINCOLN, 2000). Nessas duas técnicas, o pesquisador quer chegar o mais perto possível do seu objeto, no intuito de abrir nova perspectiva dentro da situação observada (GLASER, 1978).

Por tratar o processo de amostragem como uma comparação constante, cujas descobertas são usadas para delinear categorias específicas e narrativas descritivas, a análise etnográfica de conteúdo está enquadrada no campo da pesquisa etnográfica. Nessa abordagem, os dados coletados seguindo alguma forma de protocolo estruturado podem trazer à tona questões cruciais e temas não-aparentes em uma primeira aproximação, permitindo o retorno aos dados quando novas perguntas vão surgindo. Também, segundo Altheide (1987), a análise etnográfica de conteúdo pode ser contrastada com a análise de conteúdo, no sentido de esta partir de pressuposições positivistas de objetividade e da amostragem estática dos dados.

Seguindo a proposta deste trabalho e um dos seus objetivos específicos que é o de resgatar os conceitos e aplicações usadas para caracterizar as práticas de preservação digital (Objetivo específico 4), utilizou-se a técnica de análise etnográfica de conteúdo para examinar as percepções dos envolvidos e, dessa forma, descobrir temas decorrentes, padrões emergentes e processos enfatizados e adotados.

Tanto na análise bibliométrica como na pesquisa de campo, os documentos publicados por pesquisadores de três unidades de pesquisa do MCT foram englobados, considerando seus trabalhos como contribuições na formulação de definições e recomendações gerais, para os interessados no tema de preservação digital. Cabe destacar que, apenas os trabalhos de uma dessas unidades partiram da aplicação prática de projetos de preservação digital.

4.4.1 Observações de campo

Teóricos como Selltiz *et al.* (1975, p. 70) recomendam, como técnicas de coleta de dados apropriadas a estudos exploratórios, o exame de registros de conhecimentos existentes, a entrevista não estruturada e a observação. Para fins

desta pesquisa e partindo de uma abordagem essencialmente qualitativa, cada objetivo específico deste estudo precisou de métodos e técnicas adequadas.

O objetivo da observação das práticas de preservação digital implementadas nessas instituições de pesquisa públicas brasileiras era coletar dados complementares sobre as percepções e ações relacionadas às necessidades de preservação digital e comparar os resultados com os obtidos em outros estudos citados na bibliografia.

A pesquisa de campo, que consiste na observação atenta de um objeto (ou fenômeno) de interesse do pesquisador, envolve técnicas de coleta e apresentação de dados e discussões qualitativas dos resultados. A pesquisa de campo nas sedes dos centros de informação confirmou e incrementou os dados recebidos nos questionários enviados via correio eletrônico.

Foram coletadas fotografias, cópias de textos impressos sobre as instituições e depoimentos gravados. As observações nos locais visitados foram registradas em forma de notas e diários.

Durante as visitas às 20 sedes das unidades de pesquisa do MCT, tornou-se possível observar, por curto período de tempo, os responsáveis pelos centros de informação e suas equipes, em grupo e individualmente, procurando obter maior quantidade de dados, apoiando-se no interesse dos entrevistados pelo tema da pesquisa.

As visitas permitiram uma forma de interação com a maioria das pessoas identificadas, sendo que todas elas foram agendadas e os respondentes avisados sobre o teor das visitas. Os encontros com os entrevistados se deram nas instalações das bibliotecas e centros de informação, nas salas dos pesquisadores e de reuniões. A maioria dos responsáveis pelos centros entrevistados mostrou as instalações e apresentou suas equipes. Na maioria dos casos, foi possível visitar os setores de arquivo, informática e as diretorias.

Nos setores de arquivo visitados constatou-se, também, não haver nenhuma atividade de preservação digital expressa por parte das equipes abordadas. Nos setores de informática foi possível observar o local de armazenamento dos arquivos digitais, algumas das rotinas de cópias de segurança e uso de bases de dados. De maneira geral foi percebida: a ausência da aplicação de um plano estratégico de preservação digital; o uso mínimo de padrões

internacionais de interoperabilidade e inexistência de mecanismos que garantissem a inalterabilidade dos registros digitais.

Nas unidades de pesquisa, as atividades relacionadas com a migração de formatos e com as cópias de segurança foram as mais apontadas pelos analistas de sistemas e de suporte técnico. Com exceção de uma unidade, nenhuma havia adotado alguma solução tecnológica especificamente voltada à preservação dos documentos digitais.

As práticas de migração para as novas versões de *software* que foram mencionados refletiram a dependência nas soluções proprietárias. Nos casos em que houve adoção de *software* livre e que existia preocupação na aquisição de algum grau de interoperabilidade entre os sistemas, os responsáveis pelos centros manifestaram seu interesse em que essas soluções pudessem chegar a ser usadas na aplicação de estratégias de preservação de longo prazo da informação científica (Protocolo OAI, LOCKSS, DSpace, SEER).

Registraram-se algumas contribuições por parte de grupo técnico e de representantes das diretorias, ainda que não lidassem diretamente com a preservação dos documentos digitais. Elas mostraram outras visões do problema relacionadas à conservação de bens culturais ou de aparelhos tecnológicos usados e soluções encontradas para o acompanhamento gerencial de projetos, a saber:

> - Todos os projetos e relatórios que passam pelo crivo da Fundação deveriam ser preservados (Técnico geral 2).
>
> - Os trabalhos dos pesquisadores devem ser preservados. Eu acho que o levantamento histórico feito do acervo do Museu, que conta a história não só da astronomia, mas das ciências no Brasil, um acervo que veio do Observatório Nacional e que foi todo tratado, e isso tem de ser digitalizado. (Técnico geral 1).
>
> - Tudo deveria ser preservado, principalmente a área de arqueologia, uma das áreas mais fortes do Museu. Os curadores das coleções nos departamentos e serviços ligados às coordenações são os que lidam com a conservação. (Técnico geral 1).
>
> - O que deveria ser preservado no Centro são as diretrizes de atuação, os guias de procedimentos de trabalho (Técnico geral 3).
>
> - O Centro criou desde 1992 um sistema de gestão de acompanhamento de projetos, um sistema que permite o controle do fluxo da informação no centro gerenciado pela área de planejamento. (Técnico geral 3).

Outro aspecto constatado nas visitas foi a pouca menção a padrões de metadados que contemplem a preservação digital nas soluções tecnológicas

adotadas. Mesmo nas instituições que denotaram algum grau de preocupação com a preservação dos documentos digitais o foco se dirigia para a aquisição de material digital e a promoção de algum tipo de acesso.

> - Esse é o compromisso básico de todo pesquisador que tem um projeto aprovado no Laboratório, ele não pode se negar a apresentar o resultado do trabalho dele e deixar uma cópia na biblioteca. (Técnico geral 11)
>
> - O acesso aos documentos na nossa memória técnica depende do tipo de documento, pode ser restrito; a informação aparece na nossa intranet, mas a permissão chega na biblioteca e a gente consulta pesquisadores se podem ser acessados. (Bibliotecário responsável 5)
>
> - Os textos completos não são colocados na página do Laboratório porque são muito pesados. Quando alguém pede um artigo, a biblioteca envia o documento. (Técnico geral 6)

Com a observação nas unidades de pesquisa esperava-se encontrar percepções e ações que mostrassem a visão local sobre a preservação digital. A documentação recolhida nas instituições visitadas permitiu acompanhar a alocação de recursos materiais e humanos nos centros de informação e bibliotecas, e suas atividades relacionadas com as coleções digitais de informação científica. Também, foi constatada a natureza dos projetos planejados (Ibict, CBPF) e implementados (Inpe), com os efeitos das intermitências das políticas institucionais nas prioridades de suas administrações.

4.4.2 Os documentos

Na pesquisa de campo foram obtidos materiais publicados nas unidades de pesquisa e do MCT, procurando encontrar conceitos ligados ao contexto da preservação digital. Identificaram-se apenas três tipos de documentos: de divulgação, oficiais e bibliográficos.

4.4.2.1 Materiais de divulgação

As unidades com mais de uma década de existência têm publicado materiais sobre o histórico da instituição. Durante a pesquisa de campo se teve acesso a esses materiais. Percebeu-se que além da ausência de um tratamento padronizado no uso e divulgação de informações consideradas históricas, os temas relacionados com a preservação digital não estiveram presentes.

Além das informações nas páginas dos portais das unidades na internet, as instituições já publicaram algum tipo de informação descrevendo a história da fundação e suas transformações no tempo. Essas informações estão contidas em livros (Inpa), teses e dissertações (MPEG, LNLS, Cetem, LNA, Ibict) que mencionam fatos históricos e pessoas destacadas na instituição.

Nos materiais de divulgação analisados a preservação de suportes de informação analógicos e instrumentos de pesquisa foram citados, mas não foi encontrada nenhuma descrição de procedimentos relacionados à manutenção de arquivos digitais.

Existe um único material distribuído pela Coordenação de Documentação em História da Ciência do Mast (MUSEU DE ASTRONOMIA E CIÊNCIAS AFINS, 2007), mencionando o oferecimento do quinto curso de segurança de acervos culturais em maio de 2007. Ele destaca que, no terceiro dia, a estrutura do curso incluiria conteúdo sobre preservação digital. A palestra abordou: definições, obsolescência das mídias digitais, metadados e projetos internacionais em andamento.

4.4.2.2 Documentos oficiais

Em termos de publicações oficiais que mencionam o tema, há um documento (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2003) assinado pela diretora do Ibict, instituindo um comitê técnico de estudos sobre preservação de documentos em arquivos digitais. Nele são relacionadas as necessidades de estabelecer propostas para a criação de políticas e

diretrizes nacionais de preservação de documentos digitais no âmbito da Biblioteca Digital Brasileira (BDB), programa institucional mantido pelo órgão.

O comitê teve apenas uma reunião geral, mas foram criados grupos de trabalho para discussão de assuntos ligados à preservação digital. O único documento encontrado como resultado de reuniões foi o do GT de Certificação digital, no qual são propostas a criação de um sistema de certificação digital e a expedição de um selo brasileiro de preservação digital (INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA, 2003b).

No Inpe foi localizado o documento que caracteriza a política de editoração e preservação da produção intelectual do instituto. Em 2005, em forma de resolução, o diretor da instituição, assessorado pelo Conselho de Editoração e Preservação da Produção Intelectual do Inpe, determinou que o Serviço de Informação e Documentação definisse as "[...] características do sistema de *software* no que se refere, em particular, ao item preservação por longo prazo [...]", assim como "[...] supervisionar submissões e migrar/atualizar conteúdos." (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS, 2005, p. 2).

4.4.2.3 Material bibliográfico

O levantamento bibliográfico inicial da pesquisa ajudou a identificar os autores brasileiros que escreveram sobre o tema no período de 1996 a 2007, entre os quais servidores do Cnen, do Inpe e do Ibict. Na seleção dos documentos produzidos nas unidades, foram incluídos trabalhos apresentados em eventos que não foram tratados na análise bibliométrica (item 4.1). A Tabela 10 mostra os materiais bibliográficos coletados nas três instituições referidas.

Tabela 10 - Produção bibliográfica sobre preservação digital no MCT (1998-2007)

Unidade de pesquisa	Tipo de documento				
	Artigo	Capítulo	Dissertação	Livro	Trabalho em evento
Cnen	1	1	0	0	2
Ibict	2	0	1	0	3
Inpe	4	0	0	1	6
TOTAL	7	1	1	1	11

Percebeu-se que os trabalhos em eventos foram os materiais bibliográficos a que mais vezes recorreram os autores para divulgar os resultados das suas pesquisas. Esta evidência pode ter origem na rapidez da divulgação que esse formato permite. Em segundo lugar estão os artigos que conseguem ter uma abrangência mais ampla de subtemas e com número maior de propostas. Por último, com um único registro cada, estão os capítulos de livro, dissertações e livros, sendo que o número de casos pode ser atribuído ao próprio tipo de formato que requer período de tempo considerável para sua produção (ver Apêndice E).

Seguindo a caracterização que Altheide (1987) propõe para o que seria um formato de apresentação dos conteúdos, os registros encontrados que mencionavam assuntos ligados à preservação digital, continham poucas referências aos subtemas e propostas mais frequentemente citados na literatura da área. Os assuntos discutidos não saíram dos limites das recomendações gerais, com apenas uma solução tecnológica aplicada localmente.

4.4.3 Questões discutidas

Seguindo a abordagem da AEC, falar de um padrão de aparecimento e uso dos conceitos relacionados com a preservação digital precisaria envolver os contextos onde eles surgiram.

Os dados bibliográficos expostos podem ser agrupados em duas categorias, segundo o contexto da abordagem do tema por parte dos pesquisadores das três unidades de pesquisa do Ministério. A primeira categoria seria aquela em que a preservação digital aparece como parte de uma discussão teórica que expõe a carência de propostas nacionais para abordar a problemática. A segunda categoria

estaria conectada com as características locais próprias de uma implementação de soluções de preservação de documentos digitais.

O primeiro grupo foi composto por trabalhos provenientes de duas unidades que ainda não implementaram metodologias de tratamento de preservação de objetos digitais. No Ibict, a etapa de estudos e prospecção de tecnologias ainda não deu continuidade à implementação de estratégias reconhecidas internacionalmente, tais como, consórcios de repositórios DSpace, rede de bibliotecas LOCKSS ou a preservação desagregada de repositórios EPrints, entre outras.

No Cnen não há ainda o planejamento para a criação, armazenamento e acesso às coleções digitais locais. Como informou um pesquisador dessa instituição:

> - Tem as coisas que são nossas que estamos colocando na coleção digital e que não tem planejamento em termos de preservação digital; a única coisa que tem é que estamos guardando em formato pdf, não tem mais nada. Em termos de segurança, temos um servidor novo e as cópias de *backup*. Temos um protótipo de memória técnica que certamente vai conter uma política de preservação digital no futuro. O importante agora é dar acesso. (Pesquisador geral 5)

As decisões sobre a implementação desse tipo de projeto são tomadas pelas diretorias desses institutos. O respondente apontou que:

> - Fica até entranho, porque na verdade eu trabalho aqui, lido com esse negócio, mas é uma coisa isolada, é uma formação acadêmica que acaba sendo não aproveitada. (Pesquisador geral 5)

De forma parecida, na única instituição do grupo que implementou um projeto de preservação digital e que contou com o apoio da diretoria, o pesquisador que em primeiro lugar propôs um sistema de preservação digital manifestou o caráter de iniciativa individual na busca da solução. Segundo ele:

> - O *software* não era para essa missão, foi desenvolvido de forma independente em 1998, ele começou a ser usado para gerenciar as teses e os eventos. Não estava muito definido o que era. Ele é um *software* gratuito para as instituições públicas, já houve um interesse privado, mas esperamos que, com o projeto de lei sobre o acesso livre que torna obrigatório o desenvolvimento de repositórios, o interesse seja maior. (Pesquisador geral 1)

Nos três institutos onde o tema da preservação digital foi abordado, as formulações dos subtemas e aspectos discutidos estiveram presentes em todos os tipos de documentos analisados: artigo, dissertação, capítulo de livro, livro, assim como em trabalhos em eventos.

O que poderia ser considerado como padrão de surgimento e uso dos termos relacionados com a preservação digital estaria na menção deles dentro de definições, propostas de recomendações e detalhamento técnico de aplicações destacados na produção bibliográfica desses pesquisadores, como mostra o Quadro 22.

Unidade de pesquisa	Documentos	Experiências citadas	Subtemas e propostas	Aspectos discutidos
Cnen	Artigos, capítulo de livro, trabalho em evento	Iniciativas e projetos internacionais	Definições, estratégias, recomendações	Sistema de arquivamento, padrões, acesso
Ibict	Artigos, dissertação, trabalhos em eventos	Iniciativas e projetos internacionais	Definições, estratégias, recomendações	Objetos digitais, instituições, legislação
Inpe	Artigos, livro, trabalhos em evento	Projeto local	Descrição de projeto, *software*, manual, política	Iniciativa, sistema de arquivamento, padrões

Quadro 22 – Conteúdo dos documentos produzidos por servidores do MCT sobre preservação digital

No âmbito do MCT surgiram algumas das primeiras definições de preservação digital em português. No mesmo período já haviam sido publicados trabalhos em Portugal relacionados com o tema. Em ambos os países, os conceitos foram traduções de formulações feitas por pesquisadores de língua inglesa (FERREIRA, 2006, RODRIGUES, 2003; BORBINHA, 2001).

As recomendações mencionadas nos textos analisados também seguem as propostas de autores em publicações estrangeiras. Elas vão desde aquelas relacionadas à formação dos responsáveis, até a indicação de procedimentos e tecnologias para o gerenciamento da preservação de documentos digitais.

Finalmente, os contextos mencionados em que surgem formulações locais sobre o tema, são áreas do Ministério nas quais se buscam soluções para os problemas que os documentos digitais apresentam. A pergunta sobre como estava sendo percebida a preservação digital nas unidades de pesquisa do MCT pode ser respondida afirmando que o problema está na inexistência de experiências práticas e na ausência de decisões institucionais, que não têm emergido por parte das diretorias.

Uma amostra da produção sobre o tema e de como o Inpe conseguiu dentro da instituição estabelecer o primeiro projeto e política de preservação digital do Ministério, está disponível no Apêndice E.

4.5 DETERMINAÇÃO DE CRITÉRIOS

Um critério é uma idéia a partir da qual decidimos ou valorizamos alguma coisa ou alguém. Isso implica requerer de indicadores que ajudem a visualizar algumas ações concretas para poder avaliar se o critério está sendo atendido ou não. Para o dicionário Houaiss ele seria “uma norma de confronto, avaliação e escolha”.

A definição mais usada de critério é que ele sirva de base para que pessoas e coisas possam ser comparadas ou julgadas. Por isso, um critério de preservação digital é uma regra que permite a fundamentação racional da escolha, decisão e deliberação de uma entidade, com relação ao serviço de informação de preservação que ele presta à sua comunidade.

O objetivo central desta tese foi propor um conjunto de critérios de preservação digital que possam ser usados na elaboração de projetos de serviços de gestão de informação de caráter científico nas unidades de pesquisa do MCT. Considerando os argumentos expostos na revisão de literatura, configurou-se um conjunto de pressupostos para tentar entender os problemas críticos relacionados à permanência das informações digitais. Assim a formulação dos critérios de preservação digital de informação científica partiram, num primeiro momento dos seguintes pressupostos:

1) a gestão dos documentos digitais é um contínuo integrado de processos que envolvem todo o ciclo de vida documental;
2) existência de graus de completude dos sistemas organizacionais com relação à gestão da preservação dos documentos digitais;
3) a percepção de que as necessidades de preservação digital são determinantes na definição desses sistemas;
4) os critérios de preservação digital são plausíveis de serem adequados ao contexto das organizações.

A primeira etapa de uma possível adequação de critérios básicos de preservação digital observou a definição dos tópicos tratados na literatura que revelaram a importância do tema (ver item 2.7). A produção científica das áreas de arquivologia, biblioteconomia e ciência da informação indicou alguns dos tópicos que podem ser abordados no estudo da preservação digital, entre os quais:

a) alcances das estratégias (THIBOBEAU, 2002; BEAGRIE, 2003);
b) arquitetura das soluções (LAVOIE; GARTNER, 2005; LAZORCHAK, 2005);
c) metodologias apropriadas (HODGE; CARROLL, 1999);
d) motivações e necessidades dos usuários (WOODS, 2002);
e) percepções e ações dos gestores (WEBB, 2000; GLADNEY, 2006).

O contexto explorado na pesquisa de campo evidenciou que qualquer implementação de soluções para a preservação de documentos digitais precisa contemplar todas as atividades que envolvem o ciclo de vida do objeto digital. A literatura da área mostrou que, já na fase inicial da criação das coleções digitais (tenham sido nascidas digitais ou digitalizadas), devem ser observados alguns critérios de preservação ao estabelecer as primeiras ações para sua permanência:

a. avaliação dos conteúdos;
b. identificação dos requisitos de metadados;
c. avaliação da capacidade dos metadados serem coletados;
d. cotejamento da solução selecionada com as funções dos modelos de referência existentes;
e. seleção de *software* e aplicativos automatizados de estratégias de preservação.

Em um primeiro momento, gestores dos centros de informação e bibliotecas, como as do MCT, precisariam decidir quais são os critérios de preservação digital que devem ser observados nas suas unidades para garantir a persistência dos documentos digitais em seus sistemas de informação. O julgamento de qual solução adotar pode surgir da observância desses critérios, determinando, assim, o tipo de resultado que se quer obter.

Na literatura se encontra a definição de diversos tipos de elementos, princípios, propriedades, atributos, requisitos e critérios contemplados, quando se deseja criar um tipo de serviço de informação orientado à preservação de longo prazo (BROWN, 2003; CONARq, 2006; CORNELL UNIVERSITY LIBRARY, 2003;

CRL, 2007; RLG/NARA, 2006; 2007; RLG/OCLC, 2002). Como descrito, os grupos de pesquisa que fizeram esse tipo de formulação vêm testando e propondo regras baseados em modelos de referência (OAIS, TDR, TRAC, Moreq, DRAMBORA, e-Arq) nas suas próprias iniciativas e projetos.

Todos os esforços de elaboração de normas para definir padrões de preservação digital têm o único propósito de que a informação contida nos sistemas de informação permaneça acessível por longo período de tempo. Quando se estabelece um serviço de informação orientado a esse propósito, a instituição deve informar quais são os critérios que ela está seguindo.

O levantamento realizado nas unidades de pesquisa do MCT deu a conhecer que, com exceção do Inpe, seus centros de informação e bibliotecas não estão aplicando nenhuma metodologia de preservação digital no tratamento dos seus documentos digitais de informação científica.

Comparando com outros estudos no Brasil, e dos grupos de estudo que começam a surgir em instituições de ensino superior (PUC-Rio, UNICAMP, UFPB, UniRio, UFMG), a situação não é diferente, como demonstra a pesquisa de Thomaz (2004), ao analisar o ambiente de arquivamento de informação digital em organizações públicas brasileiras de médio e grande porte.

No caso das unidades de pesquisa do MCT, o conjunto de critérios para o estabelecimento de sistemas de informação orientados à preservação digital a ser proposto pode ser usado para avaliar as atribuições de responsabilidade, a confiança que ele deve possuir e a acessibilidade que esse tipo de sistemas de informação deve permitir.

A determinação da lista de critérios foi elaborada segundo as principais formulações de requisitos básicos existentes e que estão sendo testados, assim como a predisposição que os sistemas devem mostrar à auditoria externa para confirmação da concordância com os modelos de referência reconhecidos internacionalmente (BULLOCK, 1999; DIGITAL CURATION CENTRE; DIGITAL PRESERVATION EUROPE, 2007; MoReq2, 2007; OCLC-CRL, 2007; RLG, 2007; RLG/OCLC, 2002).

Os dois modelos de referência reconhecidos internacionalmente para serem seguidos pelos sistemas de gerenciomento de informação científica em objetos digitais são OASIS e TDR. Com base neles estabeleceu-se o conjunto de

critérios que englobam os aspectos gerenciais e institucionais necessários para o funcionamento desses sistemas.

4.5.1 Critérios institucionais

Os critérios institucionais são aqueles que servem para informar os filtros que o sistema de preservação digital possui, com relação a garantir sua permanência organizacional e de recursos constantes para seu funcionamento. Eles representam uma referência que os usuários têm de que existe uma entidade institucional que responderá pela guarda, manutenção e acesso aos documentos depositados nesses sistemas.

Ao mesmo tempo, o grupo de critérios institucionais de preservação digital refere-se à completude da estrutura funcional do sistema de informação estabelecido, segundo seus componentes tecnológicos e de papéis necessários a seu funcionamento, assim como à garantia da consistência das informações arquivadas.

As seis categorias de critérios institucionais propostas, lideradas pelo critério de confiabilidade são as seguintes:

1. Confiabilidade
2. Responsabilidade política
3. Sustentabilidade econômica
4. Inclusão em repositórios digitais
5. Transparência
6. Acessibilidade de longo prazo

Os critérios institucionais, que englobam as características necessárias para o funcionamento de um sistema de preservação digital, permitem determinar as práticas de preservação e as condições institucionais que afetam a forma como é enfrentada a preservação da informação em formatos digitais.

Os indicadores dos critérios institucionais de preservação do conhecimento científico, ao serem sistematizados e incorporados às funções dos gestores da informação digital local, podem servir para medir a produtividade dos

centros de informação e bibliotecas, focalizando a utilização da informação e reutilização do conhecimento das novas gerações de cientistas.

4.5.1.1 Confiabilidade

Um critério institucional que determina os aspectos que a comunidade observa quando se estabelece um serviço de preservação digital é o da confiabilidade. Ele envolve questões sobre os formatos aceitáveis para depósito, armazenamento e esquema de segurança aplicado, temas que devem ser tratados abertamente com todos os participantes do projeto de preservação, incluindo seus usuários potenciais (BENNETT, 1997).

O critério da confiabilidade está intimamente relacionado com a adoção de padrões internacionais, usados como elementos para a certificação de que as coleções digitais observam todos os fatores condicionantes de preservação. Estão contidos no critério confiabilidade os indicadores que expressam as garantias que a instituição provê de que o serviço prestado é adequado às necessidades da comunidade-alvo. O cumprimento deste critério é conferido algumas vezes aos serviços terceirizados ou uso de soluções não oriundas da instituição.

Observa-se o estabelecimento de uma dependência física na manutenção dos acervos digitais quando a confiança de que o serviço cumpre funções de preservação de longo prazo não compete ao centro de informação ou biblioteca. Nesses casos, independentemente da solução adotada, a instituição deve estabelecer os graus de responsabilidade e informar o critério que está sendo cumprido.

Cabe aos gestores a tarefa de identificar e revisar as informações de empacotamento dos dados dentro dos sistemas tecnológicos adotados e sua concordância com o modelo de referência OAIS (CCSDS, 2002). O armazenamento estaria determinado pela maneira como foram organizadas as informações de conteúdo, a informação descritiva e as permissões de acesso.

Um sistema de preservação digital deve possuir, em sua estrutura, dispositivos que permitam a guarda das informações sobre a segurança, direitos autorais, tipos de acessos e possibilidades de auditorias. A confiabilidade também deve ser

considerada nas medidas de segurança, observadas na construção dos repositórios digitais. Essas medidas podem garantir que os materiais depositados manterão sua autenticidade no decorrer do tempo.

4.5.1.2 Responsabilidade política

A apreciação do critério institucional “responsabilidade política”, efetuada por meio da avaliação das políticas de implementação ou pelo tipo de parcerias que ela envolve, pode ajudar a identificar as variáveis organizacionais que determinam a presença ou não de ações concretas ante aos riscos de perdas da produção científica local.

A partir das resoluções oficiais divulgadas como política de preservação digital, que incluem as funções e atributos que cabem às instituições ou setores responsáveis por adotar os critérios de preservação assumidos, é possível detectar também a importância da informação registrada nesses sistemas. A verificação da observância de critérios pode confirmar se a instituição tem na sua missão manter repositórios de informação confiáveis (TDR, 2002).

A natureza da declaração da política, resolução ou mandato de preservação digital é evidente nas primeiras bibliotecas nacionais a determinarem suas políticas, como a National Library of Australia ou a British Library. Nesta última, a formulação dos critérios para o estabelecimento do elemento preservação nas soluções adotadas está contida nos seguintes indicadores:

1) desenvolvimento de um sistema de armazenamento e gerenciamento de materiais digitais a longo prazo;
2) características das funções do sistema a ser desenvolvido e do material a ser depositado;
3) propriedades do sistema para ser considerado desejável aos propósitos para os quais foi criado (tipo de dados, automação dos processos, formatos armazenáveis, tipos de *back-ups*);

4) utilização de soluções temporárias, até o estabelecimento do sistema definitivo.[47]

No caso dos repositórios institucionais de acesso livre, a preservação digital não é prioridade em sua criação, mas quando os critérios de preservação são observados, os repositórios adquirem caráter de responsabilidade que as organizações não manifestam quando disponibilizam um serviço de acesso livre.

Ao publicar sua política de preservação digital, as instituições informam que seus repositórios possuem a habilidade técnica para manter e demonstrar a autenticidade e integridade dos objetos preservados. Afirmam, também, seu compromisso em criar serviços de preservação digital como, por exemplo, os serviços que estão sendo acompanhados pelo *Directory of Open Access Repositories* (OpenDOAR)[48] *da University of Nottingham.*

4.5.1.3 Sustentabilidade econômica

O critério de sustentabilidade econômica é útil na avaliação do grau de responsabilidade de um sistema de preservação, segundo o tipo de apoio que recebe internamente e das instituições de fomento. Este critério inclui considerações sobre a parte do orçamento anual dedicado a atividades de preservação digitais, tais como migração, atualização, criação de metadados, entre outras.

Os possíveis modelos de negócios dos sistemas de preservação digital ainda estão sendo estruturados, influenciados principalmente pela natureza das instituições de ensino e pesquisa, e pela forma como estão sendo percebidos pelas agências oficiais, responsáveis pelo fortalecimento da capacidade de cada país na área de ciência e tecnologia.

Algumas técnicas para reduzir custos com os sistemas de preservação digital apontam para a necessidade de identificar os componentes do sistema que requerem maior quantidade de investimento, em cada uma das etapas do ciclo de vida dos objetos digitais (CHAPMAN, 2003; ROSENTHAL *et al*, 2005).

---

[47] NA: versão traduzida e adaptada da Política da *British Library* disponível em: http:// Acesso em 12/set/2008)

[48] URL: http://www.opendoar.org. Acesso em 20/out/2008.

A união de experiências entre as instituições que desenvolvem projetos de preservação digital representa parte fundamental do critério de sustentabilidade econômica. A importância das parcerias é reconhecida internacionalmente (JIS/NPO, 1995), porque elas possibilitam melhor compreensão dos benefícios no compartilhamento de custos e avaliações de esforços não redundantes (CLOONAN, 2001; NATIONAL LIBRARY OF AUSTRALIA, 2003).

Devido ao grande volume de informação e à natureza da tecnologia digital, as parcerias entre criadores de materiais digitais podem definir qual a parcela de responsabilidade e as atribuições que lhes correspondem (BESSER, 2007).

4.5.1.4 Inclusão em repositórios digitais

Um dos primeiros passos na direção da preservação dos objetos digitais é a sua transferência para um repositório digital. A responsabilidade desse processo é compartilhada pelo depositante e pela instituição que oferece esse serviço.

No modelo de referência OAIS, os materiais são transmitidos aos repositórios por meio dos Pacotes de Submissão de Informação (PSI), como forma de garantir a persistência dos documentos. Com eles são implementadas as decisões sobre as características essenciais dos materiais digitais, sua preservação e acesso, independentemente da tecnologia aplicada (LeFURGY, 2002). Seria recomendável que os sistemas de gestão da informação científica incorporassem todas as regras e descrição da consistência dos materiais que estão nos PSI em relação à tecnologia adotada na criação dos documentos digitais.

O critério de “inclusão” pressupõe que a instituição possui uma coleção própria de documentos digitais que contém a produção científica local. Esse indicador essencial do caráter científico dos dados proporciona validade aos conteúdos do sistema adotado, certificando a conformidade desses registros com as normas e padrões. Tais padrões devem ser reconhecidos nas políticas estabelecidas para inclusão e na declaração das políticas de direitos autorais adotadas.

Na estrutura de arquivamento do OAIS, o critério inclusão estaria representado dentro do Pacote de Submissão de Informação (PSI) como a descrição

que serve para avaliar os mecanismos adotados para garantir a autenticidade dos objetos digitais. As técnicas de assinatura digital e de marca d´água digital também estão sendo usados com essa finalidade.

A coleção digital a ser preservada deverá conservar o seu formato original e informar qualquer modificação ou transformação futura (LYNCH, 2000). O serviço de preservação digital deve oferecer garantias de que os formatos das coleções nele depositadas serão atualizados periodicamente, usando *software* livre e minimizando o acesso físico às mídias. A habilidade de manter a completitude dos dados pode ser detectada usando técnicas de verificação dos pacotes de *software* para repositórios como o DSpace (MOORE; SMITH, 2007) e o LOCKSS.

4.5.1.5 Transparência

Os repositórios digitais são ferramentas necessárias para os produtores, disseminadores e usuários de documentos digitais. Projetos direcionados à preservação digital em repositórios insitucionais devem atender a pré-requisitos e serem parte de um processo transparente de disponibilização de longo prazo dos objetos digitais. Os responsáveis pela aplicação dos padrões e normas reconhecidos internacionalmente são os desenvolvedores desses projetos.

A formulação do critério “transparência” tem como propósito a observância dos processos básicos de um sistema de gestão da preservação digital nas instituições. Desde a adoção de uma solução de *software* livre, como a Unesco (2007) e o projeto *Open Access Repositories in New* Zealand (OARINZ) recomendam (CATALYST, 2006), até o uso compartilhado de esquemas de metadados de preservação.

Em alguns casos, a transparência na recuperação de dados depositados nos sistemas de preservação digital pode depender da solução tecnológica adotada (JAMES *et al*., 2003). Contar com a documentação dos procedimentos é garantia de que todos os arquivos serão recuperados de forma organizada.

Uma forma de garantir a autenticidade dos registros e dar certeza aos usuários de que o texto recuperado está inalterado é considerar os componentes de sistemas usados para extração de metadados e auditoria de documentos (ex.: *Data*

Time Stamping ou DTS), ou recorrer a mecanismos de auditoria em rede, como o LOCKSS realiza (ROSENTHAL *et al.*, 2005).

Atender à necessidade de que os sistemas contenham mecanismos de auditoria e que sejam compatíveis com as propostas de certificação valoriza o trabalho dos gestores responsáveis e expõe a preocupação com a memória institucional (RLG/OCLC, 2002).

Parte integrante do critério transparência são os processos de certificação. Eles são um dos itens indispensáveis quando se planeja realizar trabalhos colaborativos com redes de repositórios e bibliotecas digitais que usem os mesmo padrões de metadados e de intercâmbio de dados (WATERS; GARRETT, 1996). O desenvolvimento das redes de preservação digital mostra que, sem essa condição, não poderá haver uma integração e colaboração entre projetos de preservação fora e dentro das instituições (OCL/CRL, 2007).

Certificar um sistema de preservação sob a norma OAIS significa que a instituição mantedora revela suas interfaces internas nas especificações de cada um dos pacotes de informação definidos por uma norma (SIP, AIP, e PDI). Projetos de preservação digital, como o *Nestor* da Alemanha, o *Digital Curation Center* na Inglaterra e os da RLG e NARA apontam para uma aplicação colaborativa dos critérios de forma a expor qualquer risco e o registro das decisões assumidas.

A medição do indicador certificação encontra-se em discussão entre os especialistas da área, para quem a certificação pode representar ou não um processo em vários níveis que refletem os graus de competência e confidencialidade da instituição (AMBACHER, 2007).

Com o crescimento do número de repositórios em funcionamento, se faz necessário o estabelcimento de programas nacionais de certificação de repositórios digitais de informação científica apoiem seus mecanismos de avaliação e medição em normas internacionais.

4.5.1.6 Acessibilidade de longo prazo

Os gestores dos sistemas de preservação digital deverão incluir informações sobre o tipo de tratamento específico para cada coleção e se ele

precisará de um aplicativo especial para ser reusado. O planejamento da manutenção da acessibilidade deve informar qual estratégia será considerada nos casos de mudança de *software* e os termos e condições para uso dos materiais.

O item acesso em um repositório digital direcionado pelo modelo de referência OAIS ajuda a identificar os serviços e processos que possibilitam aos usuários visualizar os acervos. A acessibilidade do sistema de preservação digital permite medir as facilidades de uso que ele proporciona aos usuários, a dimensão das tarefas realizadas e o grau de conhecimento que eles têm das funções do sistema.

Nos projetos de preservação digital em funcionamento é freqüente encontrar um tipo de guias e manuais de usuário, oferecendo material referencial básico que visa à obtenção da certificação da acessibilidade manual ou automática (BEAGRIE; JONES, 2006; KACZMAREK *et al.,* 2006).

Como mencionado, a acessibilidade de um sistema de preservação digital depende, também, da sua capacidade de ser interoperável. Essa condição é alcançada por meio do uso de ferramentas de *software* livres, possibilitando a alteração de qualquer parte do sistema sem nenhuma conseqüência nas suas funcionalidades.

A interoperabilidade representa um desafio para os repositórios digitais, pois oferece a possibilidade de trabalhar colaborativamente com outros serviços de preservação digital, trazendo ganhos para as organizações ao transferir conteúdos entre repositórios solucionando, dessa forma, suas necessidades de cópias de segurança remotas e replicações.

O critério de acessibilidade de longo prazo pode ser medido por meio de indicadores de desempenho que proporcionam os dados sobre a eficiência do sistema no atendimento de usuários e suas expectativas de tempo e tipo de resposta. Também, o contexto técnico de conectividade entre *links* (*linkage*) impõe a preservação da integridade dos objetos digitais, levando em consideração suas relações com outros recursos e criando contextos mais amplos que identificam características da rede a que pertencem.

5 CONCLUSÕES E SUGESTÕES

Considerando a proposta geral desta pesquisa de formular um conjunto de critérios necessários à implementação de sistemas de gestão de preservação de informação científica em documentos digitais, pode-se afirmar que esse objetivo foi alcançado na definição dos critérios institucionais de preservação digital para as unidades de pesquisa do MCT.

A definição desses critérios foi resultado do estudo das principais propostas de requisitos, padrões e normas internacionais encontradas na literatura das áreas da ciência da informação e arquivologia. O levantamento bibliográfico possibilitou ressaltar a natureza multidisciplinar da preservação digital, envolvendo essas duas ciências com outras teorias das áreas humanas e aplicadas.

Quatro foram os objetivos específicos, desta pesquisa, a saber:

1. caracterizar as práticas de preservação digital da informação científica;
2. identificar os critérios para sistemas de gestão da preservação digital;
3. identificar as necessidades de preservação digital;
4. identificar os conceitos e aplicações usadas que podem interferir ou não na implementação de procedimentos padronizados de preservação digital da informação científica.

Os dois primeiros objetivos foram alcançados na análise das experiências com projetos de preservação citados na literatura. O levantamento bibliográfico proporcionou uma aproximação aos resultados de experiências internacionais de projetos e pesquisas da área, levando à constatação da origem das práticas e da identificação dos fatores que compõem os critérios necessários para o funcionamento de um sistema de preservação digital.

Os dois objetivos restantes foram alcançados por meio da pesquisa de campo, que possibilitou a detecção das necessidades que os setores de informação devem preencher e as condições nas quais surgiu a primeira experiência com sistemas de preservação digital no MCT.

Como a pesquisa envolveu uma população restrita – responsáveis por centros de informação e bibliotecas do MCT e suas equipes – não é possível generalizar os resultados obtidos. Contudo, a pesquisa possibilitou trazer à consideração a abordagem dos critérios de preservação digital, envolvendo os

sistemas de informação científica. Não se pode deixar de apontar que o nível de tratamento administrado às possíveis variáveis contidas nos indicadores desses critérios foi consoante ao limitado número de experiências dentro do Ministério da Ciência e Tecnologia

A verificação das práticas de tratamento dos materiais digitais levou à seleção dos critérios dirigidos, principalmente, aos dirigentes e responsáveis pelos centros de informação e bibliotecas do MCT. Resumindo, os critérios de preservação propostos são:

1) confiabilidade, integrada pelo conjunto de requisitos técnicos e gerenciais que garante a integridade dos formatos, a permanência do armazenamento dos dados e a segurança em todas as etapas do serviço;
2) responsabilidade política, que obriga a instituição a assumir uma parcela de controle na manutenção dos acervos digitais;
3) sustentabilidade econômica, que define as ações necessárias para a continuidade do serviço;
4) inclusão em repositórios digitais, que estrutura instâncias de validação de dados, seu caráter científico e da abrangência dos serviços;
5) transparência, que explicita e documenta as especificações técnicas para a recuperação, auditoria e certificação dos conteúdos;
6) Acessibilidade de longo prazo, que visa a manutenção técnica do sistema, suas condições de interoperabilidade, desempenho e *linkage* com outros objetos e serviços em rede.

Como observado na literatura, a importância relativa que é dada ao tema da preservação digital pode variar em função da área ou subárea científica considerada. Por exemplo, a arquivologia e a ciência da informação podem ter diferenças com relação ao foco dos trabalhos, mas mostram as mesmas necessidades de esquemas e modelos de preservação digital.

Na revisão da literatura observou-se que uma porcentagem elevada de trabalhos científicos, baseados em experiências com projetos de preservação digital, dedica-se a publicar, ainda, recomendações sobre todas as opções existentes de tratamento de materiais digitais. Entre as explicações possíveis para esse fato, está a recentidade de projetos e iniciativas e as incertezas sobre a durabilidade desses sistemas.

Tecnicamente, o tratamento adequado dos documentos digitais corresponde, ainda, à descrição detalhada dos seus componentes físicos e imateriais e à padronização no uso dos seus sistemas e formatos. Tal afirmação pode ser confirmada na consulta à bibliografia deste trabalho e no resumo das atividades de preservação em vários países do mundo nos últimos dez anos escrita por pesquisadores, como McGovern em 2007.

As alternativas de tratamento dos objetos digitais apresentadas precisam ser avaliadas e discutidas para que os gestores desses acervos possam chegar a uma conclusão adequada sobre qual estratégia seguir.

A literatura apresentada proporcionou alternativas para discutir os atributos, critérios e normas de funcionamento de sistemas de informação para a preservação digital. As experiências internacionais apontam a necessidade urgente de estratégias de preservação aplicadas aos materiais que estão sendo produzidos nas instituições preocupadas com a memória patrimonial e a herança cultural.

A maior dificuldade registrada nesses projetos e iniciativas está concentrada na parte organizacional e tecnológica das soluções encontradas. A formação de recursos humanos necessários para o funcionamento desses sistemas começa a ser definida em instituições de ensino e pesquisa, onde repercute o interesse para que os cursos e ementas de disciplinas das áreas de ciência da informação, arquivologia e museologia abordem essa necessidade nos seus conteúdos.

A constatação da reduzida relevância que as instituições conferem às atividades de preservação digital pode explicar porque não foram identificadas experiências semelhantes na área do governo brasileiro que lida com a informação científica e tecnológica, assim como o pequeno número de trabalhos que sirvam de referência para a implementação desse tipo de serviço no Brasil.

A fim de avaliar as noções que os gestores dos acervos digitais dessas instituições de pesquisa tinham sobre a preservação digital, foram selecionados alguns depoimentos coletados na pesquisa de campo realizada com essa finalidade, assim como se propôs uma classificação do seu conteúdo adaptada do modelo da análise etnográfica de conteúdo sugerida por Altheide (1987).

Os dados coletados diretamente com os responsáveis dos centros de informação e suas equipes, em datas e lugares diferentes, por meio de questionários e entrevistas, confirmaram a carência de conhecimentos sobre o tema. Pode-se

conjecturar que a percepção dessas pessoas sobre a preservação de documentos digitais corresponde àquela mantida pelo respectivo órgão.

O Inpe foi a única unidade de pesquisa que desenvolveu uma solução tecnológica nessa direção, incluindo a formulação de uma política para padronizar as atividades de depósito e acesso.

O Inpe adotou o modelo de arquitetura de arquivamento digital OAIS em seu projeto de biblioteca digital e apresentou preocupação quanto ao desenvolvimento de Pacote de Disseminação de Informação (PDI), como a norma recomenda. Nas outras unidades, houve apenas a confirmação da existência de atividades relacionadas com a migração de formatos e com as cópias de segurança, mostrando que os setores responsáveis não estão seguindo nenhum critério de preservação digital.

Com base no exposto, constatou-se que a manifestação da necessidade de preservação a longo prazo dos objetos digitais e de maior integração entre os sistemas de informação científica aponta para a adoção institucional de normas e padrões de preservação digital. Os participantes do projeto do Inpe expressaram que a aderência a esses padrões deve apoiar-se na adequação de critérios para o desenvolvimento de repositórios digitais confiáveis.

Pode-se afirmar que, considerando a ampla divulgação de ferramentas de *software* livre para a construção de repositórios digitais e de exemplos relevantes como a do próprio Inpe e das iniciativas apoiadas pelo Conarq, a postura dos responsáveis pelos centros de informação e biblioteca começa a mudar, assim como a percepção sobre o tratamento das coleções digitais.

Posde-se afirmar, que a preservação digital é mais do que um processo técnico, ela é um processo social e cultural pelo fato de que nela se aplicam critérios. É também um procedimento legal porque define os direitos e privilégios necessários para a manutenção permanente dos registros científicos.

Considerando as observações anteriormente expostas, pode-se concluir que os objetivos desta tese foram alcançados. Espera-se que esta pesquisa ofereça subsídios para que instituições que lidam com informação científica em formato digital possam identificar os critérios mínimos que podem ser observados nos programas, políticas e projetos de preservação digital.

A confirmação da mudança na percepção dos gestores dos centros de informação e bibliotecas sobre os assuntos ligados à preservação digital poderia se

constituir em tema de estudos a serem realizados no futuro. Nestes poderiam ser comparadas informações oficiais e os projetos, além de administrar metodologias semelhantes àquelas desenvolvidas na presente tese.

Como foi observado, percepções individuais e institucionais podem ser analisadas e sua constatação comparada em períodos de tempo consideráveis. Isto permite acompanhar a variação da percepção dos gestores e as aplicações das políticas nos diferentes setores das instituições.

Existe, nas áreas da ciência da informação e arquivologia no Brasil, a necessidade de estudos que tratem não apenas da natureza dos objetos digitais, mas também de definição de políticas institucionais que incluam a preservação desses objetos como bens materiais de conteúdo científico relevante.

É necessário desenvolver pesquisas sobre a aprendizagem e aplicação de normas e padrões a serem cumpridos pelos integrantes das equipes que gerenciam os sistemas de informação. Estas pesquisas poderiam adotar perspectivas da ciência da informação sobre as competências dos profissionais da área ante os novos recursos de informação digital, especialmente nas tarefas e rotinas de funcionamento de sistemas de preservação.

O Brasil precisa, nas áreas que lidam com informação e atividades de desenvolvimento científico, de canais oficiais para a identificação, comunicação e avaliação da preservação digital de longa duração. O modelo distribuído de centros de preservação digital no Brasil pode ser semelhante àqueles desenvolvidos pelas iniciativas observadas na comunidade internacional durante a última década: formulação de políticas que envolvam o uso de tecnologias e padrões testados e adaptados em projetos de coleções digitais. A organização desses centros não pode desconsiderar a sua dimensão internacional, pois, no momento em que se escolhe determinado conjunto de critérios de preservação, ele deve ser integrado aos outros repositórios que adotem critérios semelhantes.

O desenvolvimento de conjuntos de critérios para a retenção de longo prazo de materiais produzidos por pesquisadores de instituições públicas deverá incluir uma discussão detalhada sobre a implementação de políticas para repositórios digitais, metadados de preservação para formatos específicos de documentos e considerações sobre o gerenciamento dos direitos autorais no ambiente digital.

Recomenda-se, finalmente, que projetos, programas e políticas de gestão de informação científica a serem estabelecidos nos órgãos públicos identifiquem sua orientação, segundo critérios reconhecidos internacionalmente.

REFERÊNCIAS

AAKER, D. A.; KUMAR, V; DAY, E. G. S. **Marketing research**. Hoboken , NJ : John Wiley & sons, 2004.

ABRAMO, P. Pesquisa em ciências sociais. In: HIRANO, S. (Org.). **Pesquisa social**: projeto e planejamento. São Paulo: T. A. Queiroz, 1979. p. 21-88.

ALEMNEH, D. G.; HASTINGS, S. K.; HARTMAN, C. N. A metadata approach to preservation of digital resources: the University of North Texas Libraries: experience. **First Monday**, v. 7, n. 8, Aug. 2002. Disponível em: <http://firstmonday.org/issues/issue7_8/alemneh/index.html>. Acesso em: 8 ago. 2006.

ALTHEIDE, D. Ethnographic content analysis. **Qualitative Sociology**, v. 10, n. 1, p. 65-77, Spring 1987.

ALVARADO URBIZAGÁSTEGUI, R. A bibliometria: história, legitimação e estrutura. In: TOUTAIN, L. M. B. B. (Org.). **Para entender a ciência da informação**. Salvador: EDUFBA, 2007. p. 185-217.

AMBACHER, B. Government archives and the digital repository audit checklist: modern record program. **JoDI**, v. 8, n. 2, 2007.

AMERICAN INSTITUTE OF PHYSICS. **Publishing services**. 2003. Disponível em: <http://www.aip.org/publishing/services/cs_archleg.html>. Acesso em: 18 jun. 2006.

ANDERSON, S.; PINFIELD, S. Preserving e-prints: scaling the preservation mountain. In: DPC-CURL SEMINAR ON INSTITUTIONAL REPOSITORIES, 2004, London. **Proceedings**… London: [s. n.], 2004.

ARMS, W. **Digital libraries**. Cambridge: MIT Press, 2000.

BANON, G. J. F. **Biblioteca Digital da Memória Técnico-Científica do INPE**. São José dos Campos: INPE, 2006. Disponível em: <http://urlib.net/dpi.inpe.br/banon-pc2@1905/2005/12.07.19.19>. Acesso em: 24 out. 2008.

_____; BANON, L. C. Ferramentas on-line associadas a uma biblioteca digital para publicação em eventos: Projeto XII SBSR. In: SIMPÓSIO INTERNACIONAL DE

BIBLIOTECAS DIGITAIS, 3., 2005, São Paulo. **Anais**... São Paulo: INPE, 2005. Disponível em: <http://urlib.net/sid.inpe.br/iris@1912/2005/12.02.12.24>. Acesso em: 24 out. 2008.

BASEVI, T. BDJur Consortium: Juridical Digital Library: implementing DSpace in the Brazilian Judiciary. In: ICCC CONFERENCE ON ELECTRONIC PUBLISHING, 9., 2005, Leuven , Be. **Proceedings**… Leuven, Be: Katholieke Universiteit de Leuven, 2005.

BEAGRIE, N. **A continuing Access and digital preservation strategy for the Joint Information Systems Committee (JISC)**. London: JISC, 2002. Disponível em: <http://www.jisc.ac.uk/index.cfm?name=pres_continuing>. Acesso em: 14 jan. 2007.

_____. **National digital preservation initiatives**: an overview of developments in Australia, France, the Netherlands, and the United Kingdom and of related international activity. Washington: Council on Library and Information Resources and the Library of Congress, 2003. Disponível em: <http://www.clir.org/pubs/reports/pub116/pub116/pub116.pdf>. Acesso em: 28. out. 2006.

_____; GREENSTEIN, D. **A strategic policy framework for creating and preserving digital collections**. London: King's College, 1998. Disponível em: <http://ahds.ac.uk/managing.htm>. Acesso em: 28 out. 2006.

BEARMAN, D. **Electronic evidence**: strategies for managing records in contemporary organizations. Pittsburgh: Archives and Museum Informatics, 1994.

BELKIN, N. J. Information concepts of information for Information Science. **Journal of Documentation**, London, v. 34, n.1, p. 55-85, 1978.

BELLOTTO, H. L. **Arquivos permanentes**: tratamento documental. 2. ed. rev. e ampl. Rio de Janeiro: FGV, 2004.

BENNETT, J. C. **A framework of data types and formats, and issues affecting the long term preservation of digital material**. London: British Library, 1997. (JISC/NPO studies on the preservation of electronic materials, 50). Disponível em: <http://www.ukoln.ac.uk/services/papers/bl/jisc-npo50/bennet.html>. Acesso em: 25 mar. 2007.

BERNE Convention for the Protection of Literary and Artistic Works, of September 9, 1886. Disponível em: <http://www.wipo.int/treaties/en/ip/berne/index.html>. Acesso em: 30/out/2002

BESSER, H. Collaboration for electronic preservation. **Library Trends**, v. 56, n.1, p. 216-229, 2007.

_____. Digital longevity. In: SITTS, M. K. (Ed.). **Handbook for Digital Projects***:* a management tool for preservation and access. Massachusetts: Northeast Document Conservation Center, 2000. p. 155-166.

BLAKE, M. German network of expertise in digital preservation. **Electronic Library**, v. 22, p. 368-369, Apr. 2004.

BOERES, S. A. A. **Política de preservação da informação digital em bibliotecas universitárias brasileiras**. 2004. 167 f. Dissertação (Mestrado em Ciência da Informação e Documentação) - Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2004.

_____; MÁRDERO ARELLANO, M. A. Políticas e estratégias de preservação de documentos digitais. In: CINFORM, 4., 2005, Salvador, BA. **Proceedings**... Salvador, BA: Diálogo Científico, 2005. Disponível em: <http://dici.ibict.br/archive/00000263>. Acesso em: 16 fev. 2007.

BORBINHA, J.; CORREIA, A. M. R. Deposit of scientific and technical gray literature in Portugal: a case study. In: INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS, 2001, Setubal, Portugal. **Proceedings**… Setubal, Portugal: ICEIS Press, 2001. p. 8-19.

_____. Metadata: conceito e sua relevância para as bibliotecas. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 7., 2001, Porto. **Actas**... Porto: [s. n.], 2001.

BRADLEY, K. **APSR** s**ustainability issues discussion paper**. Canberra: National Library of Australia, 2004. Disponível em: <http://www.apsr.edu.au/documents/APSR_Sustainability_Issues_Paper.pdf>. Acesso em: 18 ago. 2006.

BROOKES, B. C. Numerical methods of bibliographical analysis. **Library Trends**, v. 22, n. 1, p. 18-43, July, 1973.

BROWN, A. **Selecting file format media for long term preservation**. UK: The National Archives, 2003.

BRUCE, H. Personal, anticipated information need. **Information Research**, Washington, v. 10, n. 3, Apr. 2005. Disponível em: <http://informationr.net/ir/10-3/paper232.html>. Acesso em: 9 jan. 2006.

BUDAPEST OPEN ACCESS INITIATIVE. [site]. 2002. Disponível em: <http://www.soros.org/openaccess/read.shtml>. Acesso em: 18 jun. 2006.

BULLOCK**,** A. **Preservation of digital information**: issues and current status. Ottawa: National Library of Canada, 1999. Disponível em: <http://www.collectionscanada.ca/9/1/p1-259-e.html>. Acesso em: 28 jul. 2005.

BURGOS, M. B. **Ciência na periferia**: a luz síncroton brasileira. Juiz de Fora: EDUFJF, 1999.

CAFÉ, L. *et al.* Repositórios Institucionais: nova estratégia de publicação científica na rede. In: ENCONTRO NACIONAL DE INFORMAÇÃO EM CIÊNCIAS DA COMUNICAÇÃO, 13., 2003, Belo Horizonte. **Anais**… Belo Horizonte: [s.n.], 2003.

CAMILEON: **Creative archiving at Michigan e Leeds**: emulating the old on the new. United States: JISC and National Science Fundation, 2001. Disponível em: <http://www.si.umich.edu/CAMILEON/>. Acesso em: 19 ago. 2006.

CANDÁS ROMERO, J. El papel de los metadados en la preservación digital. **El Profesional de la Información**, v. 15, n. 2, mar./abr. p. 126-136, 2006.

CAPLAN, P. Building a digital preservation archive: tales from the front. **VINE**, v. 34, n. 1, p. 38-42, 2004.

CARLIN, A. P. Aspects of spatial arrangement in libraries. In : LASK, T. (Ed.). **Constructions sociales de l'espace, les territoires de l'anthropolgie de la communication**. Liège: Lês Éditions de l'Université de Liège, 2003. p. 87-99.

_____. Disciplinary debates and bases of interdisciplinary studies: the place of research ethics in library and information science. **Library and Information Science Research**, v. 25, n. 1, p. 3-18, Spring 2003.

CATALYST. **Technical evaluation of selected open source repository solutions on behalf of CPIT**. Version 1.3 approved. Wellington, New Zealand: [s. n.], 2006.

CEDARS PROJECT TEAM. **Cedars Guide to Preservation Metadata**. The Cedars Project, 2002.

CENTER FOR RESEARCH LIBRARIES. **Trustworthy Repositories Audit and Certification (TRAC)**: criteria and checklist (TRAC). 2007. Disponível em: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>. Acesso em: 14 mar. 2005.

CHAPMAN, S. Counting the costs of digital preservation: is repository storage affordable? **Journal of Digital Information,** v. 4, n. 2, 2003.

CHEPESUIK, R. The future is here: America's libraries go digital. **American Libraries***,* v. 2, n. 1, p. 47-49, 1997.

CHILVERS, A. H. **Managing long-term access to digital data approach**: a metadata approach. Inglaterra: Loughborough University, 2000.

CHUDNOV, D. DSpace durable digital documents. **Serials**, v. 14, n.3, Nov. 2001.

CLOONAN, M. V. W(h)ither preservation? **Library Quarterly**, Chicago, v. 71, n. 2, p. 231-242, Apr. 2001.

COLEMAN, J.; WILLIS, D. **SGML as a framework for digital preservation and access**. Washington: Commission on Preservation and Access, 1997.

COMMISSION ON PRESERVATION AND ACCESS AND THE RESEARCH LIBRARIES GROUP. **Preserving digital information**: report of the task force on archiving of digital Information. Washington, 1996. 64 p.

CONSULTIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). **Producer-archive interface methodology abstract standard**. Washington: Blue Book, 2002. Disponível em: <http://ssdoo.gsfc.nasa.gov/nost/isoas/CCSDS--651.0-R-1-draft.pdf>. Acesso em: 27 jan. 2004.

_____. **Producer-archive interface methodology documents**. Washington: Blue Book, 2003: http://nssdc.gsfc.nasa.gov/nost/isoas/paim.html>. Acesso em: 4 dez. 2006.

_____. **Reference model for an Open Archival Information System (OAIS)**. Washington: Blue Book, 2002. Disponível em: <http://public.ccsds.org/publications/archive/650x0b1.pdf> Acesso em: 10 mar. 2006.

CONSELHO NACIONAL DE ARQUIVOS. **Carta para a preservação do patrimônio arquivístico digital**. Rio de Janeiro: UNESCO, 2004. Disponível em: <http://www.arquivonacional.gov.br/conarq/cam_tec_doc_ele/preservacao/cartapreservacao.asp>. Acesso em: 6 abr. 2006.

CORNEL UNIVERSITY LIBRARY. **Digital preservation management**: implementing short-term strategies for long-term problems. 2004. Disponível em: <http://www.library.cornell.edu/iris/tutorial/dpm/>. Acesso em: 16 nov. 2005.

COSTA, M. F.; COSTA, M. F. B. **Metodologia de Pesquisa**: conceitos e técnicas. Rio de Janeiro: Interciência, 2001.

COUNCIL ON LIBRARY AND INFORMATION RESOURCES (Library of Congress). **Capturing analog sound for digital preservation**: report of a roundtable discussion of best practice for transferring analog discs and tapes. Washington, 2006.

COX, J.; COX, L. **Scholarly publishing practice**: the ALPSP report on academic journal publishers: policies and practices in online publishing. United Kingdom: John Cox Associates, 2003. Disponível em: <http://www.arl.org/sparc/IR/ie.html>. Acesso em: 25 out. 2006.

CROW, R. **The case for institutional repositories**: a SPARC position paper. Washington, D.C.: Scholarly Publishing & Academic Resources Coalition. 2002. DIsponível em: <http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>. Acesso em: 2 set. 2008.

CROWE, R. **The case for institutional repositories**: a SPARC position paper. Washington: The Scholarly Publishing & Academic Resources Coalition, 2002. Disponível em: <http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>. Acesso em: 14 jul. 2008.

CULYBA, R. J.; HEIMER, C. A.; Petty, J. C**.** The ethnographic turn: fact, fashion, or fiction? **Journal Qualitative Sociology**, v. 27, n. 4, p. 365-389, Dec. 2004.

CUNHA, M. B. Desafios na construção de uma biblioteca digital. **Ciência da Informação**, Brasília, v. 28, n. 3, set./dez. 1999. Disponível em: <http://www.ibict.br/cienciadainformacao/rst/viewarticle.php?id=323>. Acesso em: 28 jul. 2004.

DAY, M. E-print services and long-term access to the record of scholarly and scientific research. **Ariadne**, n. 28, June 2001. Disponível em: <http://www.ariadne.ac.uk/issue28/metadata/>. Acesso em: 11 ago. 2006.

_____. The Final Cedars Workshop: a report from Manchester, **RLG DigiNews**, v. 6, n. 2, Apr., 2002. Disponível em: <http://www.rlg.org/preserv/diginews/diginews6-2.html#conference>. Acesso em: 18 jul. 2006.

_____. Metadata for digital preservation: an update. **Ariadne**, n. 22, Dec. 1999. Disponível em: <http://www.ariadne.ac.uk/issue22/metadata/intro.html>. Acesso em: 18 ago. 2006.

_____. Preservation metadata. In: GORMAN, G. E.; DORNER D. G. (Ed.). **Metadata applications and management**. London: Facet Publishing, 2004, p. 253-273.

DEFENSE TECHNICAL INFORMATION CENTER. **Technical metadata for the long-term management of digital materials**: preliminary guidelines. 2002. Disponível em: <http://dvl.dtic.mil/metadata_guidelines/TechMetadata_26Mar02_1400.pdf.>. Acesso em: 27 jan. 2006.

DENZIN, N. K.; LINCOLN, Y. S. (Ed.). **Handbook of qualitative research**. 2. ed. Housand Oaks, CA: Sage, 2005.

DEPARTMENT OF THE PREMIER AND CABINET (DPC). **A security management framework for online services**. [Perth, W. A.]: CAMS and Department of Commerce and Trade, 2000. Disponível em: <http://www.egov.dpc.wa.gov.au/documents/security_management.doc>. Acesso em: 6 dez. 2006.

DESCOMBE, M. **The good research guide**: for small-scale social research projects. 2nd ed. Maidenhead, Gran Bretaña: Open University Press, 2003.

DIAS, J. L. M. **FINEP**: trinta anos de projetos para o Brasil. Rio de Janeiro: FGV, 2002.

DIGITAL PRESERVATION COALITION. **Dpc Forum on Web Archiving**. Disponível em: <http://www.dpconline.org/graphics/events/060612web-archiving.html>. Acesso em: 1 set. 2006.

_____; DIGITAL PRESERVATION EUROPE. Digital Repository Audit Method Based on Risk Assessment (DRAMBORA). Disponível em: <http://www.repositoryaudit.eu/download/>. Acesso em: 20 out. 2007.

DOBRATZ, S.; SCHOGER A.; STRATHMANN, S. The nestor catalogue of criteria for trusted digital repository evaluation and certification. **JoDI**, v. 8, n. 2, 2007. Disponível em: <http://journal.tdl.org/jodi/issue/view/34>. Acesso em: 14 maio 2005.

DOLLAR, C. M. O impacto das tecnologias de informação sobre os princípios e práticas de arquivos: algumas considerações. **Revista do Arquivo Nacional,** v. 7, n. 1/2, p. 3-38, jan./dez. 1994.

DSPACE**. Format Support**. [S. l.]: MIT Libraries & Hewlett-Packard Company, 2002.

DUKART, J. R. Electronic document archival dilemma. **E-doc**, v.16, n. 5, p. 43, 2002.

ERPANET/CODATA INTERNATIONAL ARCHIVING WORKSHOP ON THE SELECTION, APPRAISAL AND RETENTION OF DIGITAL SCIENTIFIC DATA, 2003, Lisbon, Portugal. **Atas**… Lisbon, Portugal: Biblioteca Nacional de Portugal, 2003.

EWING, G. *et al*. Role and experience determine decision support interface requirements in a neonatal intensive care environment. **Journal of Biomedical Informatics**, v. 36, p. 240, 2003.

FALK, H. Digital archive developments. **The Electronic Library**, v. 21, n. 4, p. 375-379, 2003.

FEATHER, J. (Ed.). **Managing preservation for libraries and archives**. Aldershot, Hants: Ashgate, 2004.

FEENEY, M. Towards a national strategy for archiving digital materials. **Alexandria**, v. 11, n. 2, p. 107-121, 1999.

_____. (Ed.). **The digital culture**: maximising the nation's investment: a synthesis of JISC/NPO studies on the preservation of electronic materials. London: British Library Board, National Preservation Office, 1999.

FENTON, E; SCHONFELD, R. Digital preservation library periodicals expenses: variance between non-subscription costs for print and electronic formats on a life cycle basis. In: TASK FORCE MEETING, Portland, OR, 2003. **Proceedings...** Portland, OR: CNI, 2003.

FERREIRA, M. **CRIB**: a service oriented architecture for digital preservation outsourcing. Portalegre, Portugal: University of Minho, 2006.

_____. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Portalegre, Portugal: Universidade do Minho, 2006. Disponível em: <https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. Acesso em: 20 jan. 2007.

_____.; BAPTISTA, A. A.; RAMALHO, J. C. A foundation for automatic digital preservation. **Ariadne**, n. 48, Aug. 2006.

FERREIRA, S.; SOUTO, L. Dos sistemas de informação federados à federação de bibliotecas digitais. **Revista Brasileira de Biblioteconomia e Documentação**, v. 2, n. 1, 2006. Disponível em: <http://www.febab.org.br/rbbd/ojs-2.1.1/index.php/rbbd/article/view/3/20>. Acesso em: 25 jan. 2005.

FINK, F; KOSECOFF, J. **How to conduct surveys**: a step-by-step guide. [S. l.]: SAGE Publications, 1985.

FOOT, M. M. **Building blocks for a preservation policy**. London: National Preservation Office London, 2001. Disponível em: <www.bl.uk/services/preservation/upos.pdf>. Acesso em: 17 mar. 2006.

_____. Towards a preservation policy for European research libraries. **Liber Quarterly**, v. 9, n. 3, p. 323-328, 1999.

FRESKO, M. **Digital Preservation Guidelines**: The State of the Art in Libraries, Museums and Archives, 1998. Disponível em: <ftp://ftp.cordis.lu/pub/ist/docs/digicult/study1.doc> Acesso em: 15 fev. 2008.

FRIEDLANDER, A. The national information infrastructure preservation program: expectations, realities, choices and progress to date. **D-Lib Magazine**, v. 8, n. 3, Apr. 2002. Disponível em: <http://www.dlib.org/delib/september95/04friedlander.html>. Acesso em: 14 abr. 2005.

FUNARI, P. P. A. Os Arquivos no Século XXI: políticas e práticas de acesso às informações. In: CONGRESSO NACIONAL DE ARQUIVOLOGIA, 1., 2004, Brasília,. **Anais...** Brasília: Universidade de Brasília, 2004.

_____. **A diversidade cultural e a gestão, preservação e acesso a documentos digitais**. Campinas**,** SP: Siarq Unicamp, 2005. Fóruns Permanentes da UNICAMP: conhecimento e tecnologia da informação, gestão e preservação de documentos digitais.

GANDELMAN, H. **De Gutemberg à Internet**: direitos autorais na era digital. Rio de Janeiro: Record, 1997.

GEERTZ, C. **Works and lives**: the anthropologist as author. Stanford, CA: Stanford University Press, 1988.

GILLELAND-SWETLAND, A. Setting the stage. In: BACA, M. (Ed.). **Introduction to metadata**: pathways to digital information, 2000. Disponível em: <http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.htm>. Acesso em: 28 jul. 2006.

_____. **Enduring paradigm, new opportunities**: the value of the archival perspective in the digital environment. Washington: Council on Library and Information Resources, 2000. Disponível em: <http://www.clir.org/pubs/reports/pub89/pub89.pdf>. Acesso em: 25 jun. 2006.

_____.; EPPARD, P. B. Preserving the authenticity of contingent digital objects: the InterPARES Project. **D-Lib Magazine**, v. 6, July/Aug. 2000. Disponível em: <www.dlib.org/dlib/july00/eppard/07eppard.html>. Acesso em: 15 ago. 2006.

GLADNEY, H. M. **Critique**: attributes of a trusted digital repository. Saratoga, CA: HMG Consulting, 2001. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.6694>. Acesso em: 12 abr. 2006.

_____. Principles for digital preservation. **Communications of the ACM**, v. 49, p. 111-116, Feb. 2006.

GLASER, B. G. **Theoretical sensitivity**: advances in the methodology of grounded theory. Mill Valley: Sociology Press, 1978.

GORMAN, G. E.; CLAYTON, P. **Qualitative research for the information professional**: a practical handbook. London: Library Association Publishing, 1997.

GRAHAM, P. S. Long-term intellectual preservation. In: ELKINGTON, N. E. (Ed.). **Digital imaging technology for preservation.** Mountain View, California: Research Libraries Group, 1998. p. 41-57.

GRANGER, S. Emulation as a digital preservation strategy. **D-Lib Magazine**, v. 6, n. 10, Oct. 2000. Disponível em: <http://www.dlib.org/dlib/october00/granger/10granger.html>. Acesso em: 5 ago. 2006.

_____. Digital preservation and deep infrastructure. **D-Lib Magazine**, v. 8, n. 2, Feb. 2002. Disponível em: <http://www.dlib.org/dlib/february02/granger/02granger.html>. Acesso em: 28 set. 2006.

GRAY, J. **Cachorros de palha**: reflexões sobre humanos e outros animais. Rio de Janeiro: Record, 2005.

GREENAN, M. **ERPANET OAIS**: Training Seminar Report. Glasgow, United Kingdom: Erpanet, 2003. Disponível em: <http://www.erpanet.org/www/products/copenhagen/ERPANET%20OAIS%20Training%20Seminar%20Report_final.pdf>. Acesso em: 14 jul. 2006.

GREENSTEIN, D.; GEORGE, G. Digital reproduction quality*:* benchmark recommendations. **RLG DigiNews**, v. 5, n. 4, 1998. Disponível em: <http://www.rlg.org/preserv/diginews/diginews5-4.html#featured>. Acesso em: 28 jul. 2006.

_____. Managing digital collections: part II: in search of guidance. **AHDS Publications**, 1998. Disponível em: <http://www.ahds.ac.uk/about/publications/index.htm> Acesso em: 12 jun. 2005.

GREENSTEIN, S. M. **Lock-in and the costs of switching mainframe computer vendors**: what do buyers see? Oxford: Industrial and Corporate Change, 1997.

_____. **Universal service in the digital age**: the commercialization and geography of U.S. internet access. Cambridge, MA: National Bureau of Economic Research, 2008. Disponível em: <http://ideas.repec.org/p/nbr/nberwo/6453.html>. Acesso em: 14 out. 2007.

GWATNEY, M. C. Reality versus the ideal: a case study of three elementary school library media specialists performing their jobs. **Knowledge Quest**, v. 29 n. 5, p. 36-42, May/June 2001.

HAHN, K. The state of the large publisher bundle: findings from an ARL Member Survey. **ARL Bimonthly Report**, n. 245, Apr. 2006. Disponível em: <http://www.arl.org/newsltr/245/bundle.html>. Acesso em: 5 jun. 2007.

HANNA, M. Born digital: live digital? maybe. **Quarterly Newsletter of the Interactive council for Scientific and Technical Information**, v. 38, June 2001. Disponível em: <http://www.icsti/forum/38/>. Acesso em: 21 jan. 2006.

HARNAD, S. **Acesso Livre**: por quê? Belém, PA: SBPC, 2007. Disponível em: <http://users.ecs.soton.ac.uk/harnad/Temp/brazilb.ppt#626>. Acesso em 1 nov. 2008.

_____. Free at last: the future of peer-reviewed journals. **D-Lib Magazine**, v. 5, n. 12, Dec. 1999. Disponível em: <http://www.dlib.org/dlib/december99/12harnad.html>. Acesso em: 19 maio 2005.

_____. **Comunicação pessoal.** Belém, PA. 9 jul. 2007

_____. [Mensagem para lista de discussão]. 2003. Disponível em: <http://users.ecs.soton.ac.uk/harnad/Hypermail/Amsci>. Acesso em: 18 set. 2007.

_____. **Online, continuous, metrics-based research assessment**. Technical Report, ECS, University of Southampton [Unpublished]. Disponível em: <http://eprints.ecs.soton.ac.uk/12130/>. Acesso em: 4 ago. 2006.

_____. The self-archiving initiative: freeing the refereed research literature online. **Nature**, v. 10, Apr. 2001. Disponível em: <http://www.ecs.soton.ac.uk/~harnad/Tp/nature4.htm>. Acesso em: 16 jan. 2006.

HARVARD UNIVERSITY LIBRARY. **Proposal for a study of electronic journal archiving**. 2000. Disponível em: <http://www.diglib.org/preserve/ejp.htm> Acesso em: 18 jun. 2006.

HAYNES, D. **Responsibility for digital archiving and long term access to digital data**: a JISC/NPO. London: Library Information Technology Centre, London, 1997. (Study with the Electronic Libraries (eLib) Programme on the Preservation of Electronic Materials)

HEDSTROM, M. **Digital preservation**: a time bomb for digital libraries. 1996. Disponível em: <www.uky.edu/~kiernsm/DL/hedstrom.html>. Acesso em: 19 jul. 2006.

_____; MONTGOMERY, S. **Digital preservation needs and requirements in RLG members institutions**. Mountain View, CA: RLG, 1998.

HENDLEY, T. **Comparison of methods e costs of digital preservation**: British Library research and innovation report. London: British Library and Innovation Centre, 1998. Disponível em: <http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>. Acesso em: 18 set. 2006.

HILDRETH, C.R. Preserving what we really want to access, the message, not the medium: challenges and opportunities in the digital age. In: HELAL, A. H.; Weiss, J. W. (Ed.). INTERNATIONAL ESSEN SYMPOSIUM, 18., 1995, Essen. **Proceeding**… Essen: Germany University library, 1996. p. 76-95.

HIRTLE, P. Editorial: OAI and OAIS: what's in a name? **D-Lib Magazine**, v. 7, n. 4, Apr. 2001. Disponível em: <http://www.dlib.org/dlib/april01/04editorial.html>. Acesso em: 25 ago. 2007.

HITCHCOCK, S. **Metalist of open access e-print archives**: the genesis of institutional archives and independent services. **ARL Bimonthly Report**, n. 227, p. 4-11, Oct. 2003.

_____; *et. al.* Digital preservation service provider models for institutional repositories. **D-lib Magazine**, v. 13, n. 5/6, May/June 2007. Disponível em: <http://wwwdlib.org/dlib/may07/hitchcock/05hitchcock.html>. Acesso em: 13 fev. 2006.

HODGE, G. M. Best practices in digital archiving: a life cycle approach. **D-Lib Magazine,** v. 6, n. 1, Jan. 2000. Disponível em: <http://www.dlib.org/dlib/january00/01hodge.html>. Acesso em: 18 jun. 2006.

_____; CARROLL, B. C. **Digital electronic archiving**: the state of the art and the state of the practice. [S. l.]: International Council for Scientific and Technical Information Information Policy Committee and CENDI, 1999. Disponível em: <http://www.icsti.org/Dig_Archiving_Report_1999.pdf>. Acesso em: 18 mar. 2005.

_____; FRANGAKIS, E. **Digital preservation and permanent access to scientific information**: the state of the practice. [S. l.]: International Council for Scientific and Technical Information and CENDI, 2004. Disponível em: <http://www.dtic.mil/cendi/publications/04-3dig_preserv.html.>. Acesso em: 14 maio 2006.

IBM. **Digital Information Archiving System**. 1996. Disponível em: <http://www-5.ibm.com/nl/dias/>. Acesso em: 19 fev. 2007.

_____. **IBM/KB long-term preservation study**. 2003. Disponível em: <http://www.5.ibm.com/nl/dias/preservation.html>. Acesso em: 3 out. 2005.

IKPAAHINDLI, L. An overview of bibliometrics: its measurements, laws and their applications. **Libri**, v. 35, n. 2, p. 163-177, June 1985.

INERA Incorporated. **E-journal archive DTD feasibility study**. [Prepared for the] Harvard University Library, Office of Information Systems, E-Journal Archiving Project, 2002. Disponível em: <http://www.diglib.org/preserve/hadtdfs.pdf>. Acesso em: 15 set. 2006.

INSTITUTO BRASILEIRO DE INFORMAÇÃO CIENTÍFICA E TECNOLÓGICA. **Contribuição para políticas de ICT**. Brasília: IBICT, 2001.

_____. **Ordem Interna** nº. 009/02, de 13 de novembro de 2002. Brasília, DF, 2002.

_____. **PresDigi**: sistema brasileiro de preservação de documentos em arquivos digitais. Brasília, DF, 2003. (proposta preliminar para discussão [por] Miguel Ángel Márdero Arellano e Piotr Trzesniak).

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **Resolução nº 204.01**. Caracteriza a Política de Editoração e Preservação da Produção Intelectual do INPE.

JAMES, H. *et al.* **Feasibility and requirements study on preservation of e-prints.** 2003. Report Commissioned by the Joint Information Systems Committee (JISC).

JISC. **Open access briefing paper, joint information systems committee**. 2005. Disponível em: <www.jisc.ac.uk/index.cfm?name5pub_openaccess>. Acesso em: 6 ago. 2007.

JONES, M. Digital preservation activities in the U.K: building the infrastructure. In: WORLD LIBRARY AND INFORMATION CONGRESS: General Conference and Council, 69., 2003, Berlin. **Proceedings**… Berlin: IFLA, 2003.

KACZMAREK, J. *et al.* Using the audit checklist for the certification of a trusted digital repository: as a framework for evaluating repository software applications: a progress report. **D-lib Magazine**, v. 12, n. 12, Dec. 2006. Disponível em: <http://www.dlib.org/dlib/december06/kaczmarek/12kaczmarek. html>. Acesso em: 9 set. 2005.

KUNZE, J. **Towards electronic persistence using ARK identifiers**. Oakland, CA: University of Califórnia, 2003. Disponível em: <http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=kunze>. Aceso em: 27 jan. 2004.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica**. 3. ed. São Paulo: Atlas, 1991.

LAVOIE, B. Implementing metadata in digital preservation systems: the PREMIS activity. **D-Lib Magazine**, v. 10, n. 4, Apr. 2004. Disponível em: <http://www.dlib.org/dlib/april04/lavoie/04lavoie.html>. Acesso em: 28 ago. 2005.

_____. **The incentives to preserve digital materials**: roles, scenarios and economic decision-making. Columbus, OH: OCLC, 2003.

_____. Meeting the challenges of digital preservation: the OAIS reference model. **OCLC-Newsletter**, n. 243, p. 26-30, Jan./Feb. 2000.

_____. **The open archival information system reference model**: introductory guide. Dublin, USA: Office of Research OCLC, 2004. Disponível em: <http://www.dpconline.org/docs/lavoie_OAIS.pdf>. Acesso em: 24 abr. 2006.

_____. PREMIS with a fresh coat of paint: highlights from the revision of the PREMIS data dictionary for preservation metadata. **D-Lib Magazine**, v. 14, n. 5/6,

May/June 2008. Disponível em: <http://www.dlib.org/dlib/may08/lavoie/05lavoie.html.>. Acesso em: 3 fev. 2006.

_____; DEMPSEY, L. Thirteen ways of looking at - digital preservation. **D-Lib Magazine**, v. 10, n. 7/8, jul./ago., 2004. Disponível em: <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>. Acesso em: 28 abr. 2005.

_____; GARTNER, R. **Preservation metadata**. Oxford: DPC Technology watch report, 2005.

LAZORCHAK, W. M. **The ghost in the machine**: traditional archival practice in the design of digital repositories for long-term preservation. Chapel Hill: University of North Carolina, 2005. Disponível em: <http://etd.ils.unc.edu/dspace/bitstream/1901/93/1/butchlazorchak.pdf>. Acesso em: 28 mar. 2006.

LAWRENCE, G. W. *et al*. **Risk management of digital information**: a file format investigation. Washington: Council on Library and Information Resources, 2000.

LEE, K. *et al*. The state of the art and practice in digital preservation. **Journal of Research of the National Institute of Standards and Technology**, v. 107, n. 1, p. 93-106, Jan./Feb. 2002. Disponível em: <http://nvl.nist.gov/pub/nistpubs/jres/107/l/j71lee.pdf>. Acesso em: 16 jul. 2006.

LeFURGY, W. G. Building preservation partnerships: The Library of Congress National digital information infrastructure and preservation program. **Library Trends**, Washington, v. 54, n. 1, p. 163-172, 2005.

_____. Levels of service for digital repositories. **D-Lib Magazine**, v. 8, n. 5, May 2002.

LESK, M. **Practical digital libraries**: books, bytes and bucks. San Francisco, CA: Morgan Kaufmann Publishers, 1997.

LEVACOV, M. Bibliotecas virtuais: (r)evolução? **Ciência da Informação**, Brasília, v. 26, n. 2, p. 125-135, maio/ago. 1997.

LEVY, D. M. Heroic measures: reflections on the possibility and purpose of digital preservation. In: ACM

CONFERENCE ON DIGITAL LIBRARIES, 3., 1998, Pittsburgh, Pennsylvania. **Proceedings**… New York: ACM Press, 1998.

LOCKSS. **Lots of copies keep stuff safe**. 2003 Disponível em: <http://lockss.stanford.edu/index.html>. Acesso em: 28 maio 2005.

_____. **LOCKSS Alliance**. 2007. Disponível em: <http://www.lockss.org/lockss/LOCKSS_Alliance>. Acesso em: 20 out. 2005.

LUPOVICI, C.; MASANÈS, J. **Metadata for the long term preservation of electronic publications**. Den Haag: NEDLIB Consortium, 2000. Disponível em: <http://www.kb.nl/nedlib/results/preservationmetadata.pdf>. Acesso em: 25 set. 2005.

LYMAN, P. **Archiving the world wide web**: in building a national strategy for preservation: issues in digital media archiving. Washington: Council on Library and Information Resources, 2002. Disponível em: <www.clir.org/pubs/reports/pub106/web.html>. Acesso em: 14 mar. 2006.

LYNCH, C. Institutional repositories: essential infrastructure for scholarship in the digital age. **ARL Bimonthly Report**, n. 226, Feb. 2003. Disponível em: <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml> Acesso em 2/nov/2008.

MARCUM, D. B.; FRIEDLANDER, A. Keepers of the crumbling culture: what digital preservation can learn from library history. **D-Lib Magazine***,* v. 9, n. 5, May 2003. Disponível em: <http://www.dlib.org/dlib/may03/friedlander/05friedlander.html>. Acesso em: 7 jul. 2006.

MÁRDERO ARELLANO, M. Á. Preservação de documentos digitais. **Ciência da Informação**, Brasília, v. 33, n. 2, maio/ago 2004.

_____. Preservação digital: uma preocupação dos editores de revistas científicas eletrônicas no Brasil. In: ENCONTRO NACIONAL DE EDITORES CIENTÍFICOS, 9., 2007, Ouro Preto, MG. **Proceedings**... Ouro Preto, MG: [s. n], 2007.

_____. Publicação científica e arquivos abertos. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 13., 2004, Natal. **Anais...** Natal: UFRN, 2004.

_____; ANDRADE, R. S. Preservação digital e os profissionais da informação. **DataGramaZero**, v. 7, n. 5, out. 2006.

MARTIN, J.; COLEMAN, D. Change the metaphor: the archive as an ecosystem. **The Journal of Electronic Publishing**, Michigan, v. 7, n. 3, Abr. 2002. Disponível em: <http://www.press.umich.edu/jep/07-03/martin.html>. Acesso em: 28 nov. 2005.

MAYRING, P. **Introdução à pesquisa social qualitativa**: uma introdução para pensar qualitativamente. 5. ed. Weinheim: Beltz, 2002

MELLOR, P. CAMILEON: Emulation and BBC Domesday. **RLG DigiNews**, v. 7, n. 2, Apr. 2003.

_____; WHEATLEY, P.; SERGEANT, D. M. Migration on request: a practical technique for preservation. In: ECDL: EUROPEAN CONFERENCE ON RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES, 6, 2002, London. **Proceedings**… London: University of Leeds, 2002. Disponível em: <http://www.si.umich.edu/CAMILEON/reports/migreq.pdf>. Acesso em: 14 jul. 2005.

MESSERSCHMITT, D. **Opportunities for Libraries in the NSF Cyberinfrastructure Program**. 2003. Disponível em: <http://www.loc.org/standards/mets/>. Acesso em: 14 out. 2006.

MCGOVERN, N. Y. A digital decade: where have we been and where are we going in digital preservation? **RLG DigiNews**, v. 11, n. 1, Apr. 2007. Disponível em: <http://digitalarchive.oclc.org/da/ViewObject.jsp?objid=0000070519&reqid=8514>. Acesso em: 28 ago. 2005.

MINAYO, M. C. S. **O desafio do conhecimento**. São Paulo: Hucitec; Rio de Janeiro: Abrasco, 1992.

MOEN, W. The Metadata approach to accessing government information. **Government Information Quarterly**, v. 18, n. 3, p. 155-165, 2001.

MOORE, R. W.; SMITH, M. **Assessment of RLG trusted digital repository requirements**. 2005

_____; _____. Automated validation of trusted digital repository assessment criteria. **JoDI**, v. 8, n. 2, 2007. Disponível em: <http://journal.tdl.org/jodi/issue/view/34>. Acesso em: 21 ago. 2006.

MORGAN, C. **Metadata and deposit protocols**. 2000. Disponível em: <http://bic.org.uk/Cliff%20Mrgan>. Acesso em: 28 ago. 2005.

MOREQ2. **Cornell Management Consultants plc**. 2007. Disponível em: <http://www.moreq2.eu/>. Acesso em: 20 jun. 2008.

MORRIS, S. Open access: how are publishers reacting? **Serials Review**, v. 30, n. 4, p. 304-307, Nov. 2004.

MUELLER, E., *et al*. The DiVA Project: development of an electronic publishing system. **D-Lib Magazine,** v. 9, n. 1, Nov. 2003. Disponível em: <http://www.dlib/november03/muller/11muller.html>. Acesso em: 14 set. 2006.

MUIR, A. Digital preservation: awareness, responsibility and rights issues. **Journal of information science**, v. 30, n. 1, p. 73-92, 2004.

MULLEN, A. GILS metadata initiative at the state level. **Government Information Quarterly**, v. 18, n. 3, p. 167-180, Autumn 2001.

MUSEU DE ASTRONOMIA E CIÊNCIAS AFINS. **5º curso de segurança de acervos culturais:** de 7 a 11 de maio de 2007, Coordenação de Documentação em História da Ciência, Rio de Janeiro, 2007. 1 folder.

NATIONAL ARCHIVES OF AUSTRALIA. **Recordkeeping metadata standard for commonwealth agencies**. Version 1.0. 1999. Disponível em: <http://www.naa.gov.au/recordkeeping/control/rkms/rkms_pt1_2.pdf>. Acesso em: 16 jul. 2005.

NATIONAL LIBRARY OF AUSTRALIA. **Digital preservation policy**. 2002. Disponível em: <http://www.nla.gov.au/policy/digpres.html>. Acesso em: 16 nov. 2005.

_____. **Guidelines for the preservation of digital heritage**. Paris: UNESCO, 2003. Disponível em: <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>. Acesso em: 25 nov. 2006.

_____. **Archiving web resources**: issues for cultural heritage institutions. 2004. Disponível em: <http://www.nla.gov.au/webarchiving/program.html>. Acesso em: 27 ago. 2005.

NATIONAL LIBRARY OF NEW ZEALAND. **Initiative**. 2007. Disponível em: <http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#extraction>. Acesso em: 20 fev. 2007.

_____. **Metadata standards framework**: preservation metadata schema. 2003. Disponível em: <http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta>. Acesso em: 28 jul. 2007.

NATIONAL SCIENCE FOUNDATION. **Revolutionizing science and engineering through cyberinfrastructure**: report of the National Science Foundation Blue-Ribbon Advisoty Panel on Cyberinfrastructure. Washington: National Science Goundation, 2005. Disponível em: <http://www.communitytechnology.org/nsf_ci_report/>. Acesso em: 24 fev. 2006.

NESTOR WORKING GROUP ON TRUSTED REPOSITORIES CERTIFICATION. **Catalogue of criteria for trusted digital repositories**. 2006. Disponível em: <http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8enpdf>. Acesso em: 6 nov. 2006.

NIXON, W. The evolution of an institutional e-prints archive at the University of Glasgow. **Ariadne**, n. 32, jul. 2002. Disponível em: <http://www.ariadne.ac.uk/issue32/eprint-archives/>. Acesso em: 14 abr. 2006.

OCLC. **Digital archive**. 2003. Disponível em: <http://www.oclc.org/digitalpreservation/>. Acesso em: 27 jan. 2006.

_____. **OCLC digital preservation resources**. Disponível em: <http://www.oclc.org/digitalpreservation/about/co-op/>. Acesso em: 15 nov. 2003.

OCLC RESEARCH. **Preservation metadata working group II**: implementation strategies. 2003. Disponível em: <http://www.oclc.org/research/pmwg/>. Acesso em: 27 nov. 2004.

OCLC/RLG. **Data dictionary for preservation metadata**: final report of the PREMIS Working Group. Dublin, OH: OCLC, 2005.

_____**. Preservation metadata for digital objects**: a review of the state of the art. 2001. Disponível em: <http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf>. Acesso em: 28 jul. 2006.

_____. **Preservation metadata working group**: a metadata framework to support the preservation of digital objects. Dublin, Ohio: Online Library Computer Center, 2002. Disponível em: <http://www.oclc.org/research/pmwg/pm_framework.pdf.>. Acesso em: 14 ago. 2006.

OCLC/RLG PREMIS WORKING GROUP. **Implementing preservation repositories for digital materials**: current practice and emerging trends in the cultural heritage community. Dublin, Ohio: Online Computer Library Center, 2004. Disponível em: <http://www.oclc.org/research/projects/pmwg/surveyreport.pdf>. Acesso em: 2 ago 2006.

OKUBO, Y. **Bibliometric indicators and analysis of research systems**: methods and examples. Paris: OCDE/GD, 1997.

OLTMANS, E; KOL, N. A comparison between migration and emulation in terms of costs. **RLG DigiNews**, v. 9, n. 2, 2005.

PADI. **Legal deposit**. 2003. Disponível em: <http://www.nla.gov.au/padi/topics/67.html>. Acesso em: 27 jan. 2004.

PEREIRA, J. C. R. **Análise de dados qualitativos**: estratégias metodológicas para as ciências da saúde, humanas e sociais. 3. ed. São Paulo: Universidade de São Paulo, 2001.

PETERS, T. Digital repositories: individual, discipline-based, institutional, consortia, or national? **Journal of Academic Librarianship**, v. 28, n. 6, p. 414-417, Nov. 2002.

PINFIELD, S. Creating institutional e-print repositories? **Serials**, v.15, n.3, p. 261-264, nov. 2002, Disponível em: <http://eprints.nottingham.ac.uk/archive/00000064/00/serials2002.pdf>. Acesso em: 29 de jun. 2006.

_____; GARDNER, M.; MACOLL, J. Setting up an institutional e-print archive. **Ariadne**, n. 31, Mar./Apr. 2002. Disponível em: <http://www.ariadne.ac.uk/issue31/eprint-archives/intro.html>. Acesso em: 29 jun. 2006.

_____; JAMES, H. The digital preservation of e-prints. **D-Lib Magazine**, v. 9, n. 9, Aug. 2003. Disponível em: <http: www.dlib.org/dlib/september03/pinfield/09pinfield.html>. Acesso em: 4 fev. 2006.

PINHEIRO, L. V. R.; LOUREIRO, J. M. M. Traçados e limites da ciência da informação. **Ciência da Informação**, Brasília, v. 24, n. 1, p. 42-53, jan./abril 1995.

PLUMMER, K. **Documents of life**: an Introduction to the problems and literature of a humanistic method. London: George Allen & Unwin, 1983.

POTHEN, P. **Digital curation**: digital archives, libraries and e-science. York, UK: Digital Preservation Coalition and the British National Space Centre, 2001. Disponível em: <http://www.dpconline.org/graphics/events/richtext/digital-seminarrepdg.html>. Acesso em: 4 dez. 2004.

PUBLIC RECORD OFFICE VICTORIA - PROV. *Victorian Electronic Records Strategy*. Melbourne: 1999. Disponível em: <http://www.prov.vic.gov.au/vers/published/final/final.pdf>. Acesso em: 10/dez/2007.

POTTER, M. Researching long term digital preservation approaches in the dutch digital preservation testbed (Testbed Digitale Bewaring). **RLG DigiNews**, v. 6, n. 3, June 2002. Disponível em: <http://www.rlg.org/preserv/diginews/diginews6-3.html#feature2>. Acesso em: 5 fev. 2006.

PRITCHARD, A. Statistical bibliography or bibliometrics? **Journal of Documentation**, London, v. 25, n. 4, p. 348-349, Dec. 1969.

QUENAULT, H. VERS: building a digital record heritage. In: BIENNIAL CONFERENCE AND EXHIBITION OF THE VICTORIAN ASSOCIATION FOR LIBRARY AUTOMATION, 12., 2004, Austrália. Paper presented at VALA 2004: breaking boundaries: integration and interoperability. Disponível em: <http://www.vala.org.au/vala2004/2004pdfs/13Quena.PDF>. Acesso em: 3 nov. 2007.

REICH, V. A. Diffused knoledge immortalizes itself: the LOCKSS Program. **High Energy Physics Libraries Webzine**, n. 7, Jul. 2003. Disponível em: <http://library.istl.org/02-fall/article1.html>. Acesso em: 14 jun. 2006.

_____; ROSENTHAL, D. S. H. LOCKSS: a permanent web publishing and access system. **D-Lib Magazine**, v. 7, n.6, Jun. 2001.

RESEARCH LIBRARIES GROUP. **Attributes of a trusted digital repository for digital materials**: meeting the Needs for research resources. Mountain View, CA: RLG, 2001. Disponível em: <http://www.rlg.org/longterm/attributes01.pdf>. Acesso em: 17 jul. 2006.

_____. **Audit checklist for certifying digital repositories**. 2006. <http://www.rlg.org/en/page.php?Page_ID=20769>. Acesso em: 4 maio 2005.

_____. **Working group on preservation issues of metadata**: final report, research libraries group. Mountain View, CA: RLG, 1998. Disponível em: <www.rlg.org/preserv/presmeta.html>. Acesso em: 12 dez. 2005.

RIBEIRO, D. **O processo civilizatório**. Rio de Janeiro: Civilização Brasileira, 1975.

RICHARDSON, R. J. *et al.* **Pesquisa social**: métodos e técnicas. 3. ed.rev. e ampl. São Paulo: Atlas, 1999.

RLG/OCLC. **Trusted digital repositories**: attributes and responsibilities. Montain View, Canada: RLG-OCLC Report, 2002. Disponível em: <http://www.rlg.org/longterm/repositories.pdf>. Acesso em: 23 jun. 2005.

_____. Open Archival Information System (OAIS). Resources. 2002. Disponível em: <http://www.rlg.org/longterm/oais.html>. Acesso em: 2 dez. 2005.

RLG-NARA. Audit checklist for certifying digital repositories. [S. l.]: RLG-NARA Task Force on Digital Repository Certification, 2006. <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>. Acesso em: 15 nov. 2005.

RLG/NARA TASK FORCE ON DIGITAL REPOSITORY CERTIFICATION. Trustworthy repositories audit and certification. 2007. Disponível em: <http://www.crl.edu/PDF/trac.pdf>. Acesso em: 30 mar. 2007.

RODRIGUES, E. *et. al*. RepositóriUM: criação e desenvolvimento do Repositório Institucional da Universidade do Minho. In: CONGRESSO NACIONAL DE BIBLIOTECÁRIOS, ARQUIVISTAS E DOCUMENTALISTAS, 9., 2004., Estoril, Portugal. **Anais**… Estoril, Portugal: [s. n.], 2004.

RODRIGUES, M. L. T. S. **Preservação digital de longo prazo**: estado da arte e boas práticas em repositórios digitais. 2003. Dissertação (Mestrado em Estudos de Informação e Bibliotecas Digitais) - Departamento de Ciências e Tecnologias da Informação, Lisboa, 2003.

RONDINELLI, R. C. **Gerenciamento arquivísticos de documentos eletrônicos**. Rio de Janeiro: FGV, 2002. 160 p.

ROSENTHAL D. S. H. et. al. Requirements for digital preservation systems: a bottom-up approach. **D-Lib Magazine**, v. 11, n. 11, Nov. 2005. Disponível em:

<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>. Acesso em: 28 jul. 2005.

ROTHENBERG, J. **Avoiding technological quicksand**: finding a viable technical foundation for digital preservation. Washington: Council on Library and Information Resources, 1999. Disponível em: <http:// www.clir.org/pubs/reports/rothenberg/contents.html>. Acesso em: 26 jul. 2006.

_____. **Digital preservation summary**. 2003. Disponível em: <http://www.nationalarchives.gov.uk/documents/rothenberg.pdf>. Acesso em: 2 nov. 2008.

_____. Ensuring the longevity of digital documents. 1995. **Scientific American**, 272 (1), pp. 24-29. Disponível em: <http://www.clir.org/pubs/archives/ensuring.pdf>. Acesso em: 16 nov. 2006.

_____. **An experiment in using emulation to preserve digital publications**. Netherlands: Koninklijke Bibliotheek Den Haag, 2000. Disponível em: <http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf.>. Acesso em: 9 jun. 2005.

_____. Metadata to support data quality and longevity. In: IEEE METADATA CONFERENCE, 1., 1996, Silver Spring, Md. **Proceedings**… Silver Spring, Md.: NOAA Complex, 1996. p. 16-18. <http://www.computer.org/conferences/meta96/rothenberg_Paper/ieee.data-quality. html>. Acesso em: 29 nov. 2005.

RUIZ, J. Á. **Metodologia científica**: guia para eficiência nos estudos. São Paulo: Atlas, 1979.

RUSSELL, K; SERGEANT, D. The Cedars Project: implementing a model for distributed archives. **RLG DigiNews**, v. 3, n. 3, June 1999.

SANNETT, S. Toward developing a framework of cost elements for preserving authentic electronic records into perpetuity. **College & Research Libraries**, v. 63, p. 388-404, Sep. 2002.

_____. The cost to preserve authentic electronic records in perpetuity: comparing costs across cost models and cost frameworks. **RLG DigiNews,** v. 7, n. 4, Aug. 2003.

SANT'ANNA, M. L. Os desafios da preservação de documentos públicos digitais. **Revista IP**, ano 3, n. 2, dez. 2001. Disponível em: <http://www.informaticapublica.mg.gov.br/revista0302/ip0302santanna.pdf>. Acesso em: 18 fev. 2006.

SARACEVIC, T. Information science: origin, evolution and relations. In: VAKKARI, P., CRONIN, B. (Ed.). **Conceptions of library and information science**. Los Angeles: Taylor Grahan, 1992.

SAWYER, D. M. **Framework for digital archiving**: OAIS reference model. 2002. Presentation delivered at the OCLC Steering by Standards Teleconference on the OAIS Imperative: Enduring Record or Digital Dust? Columbus, Ohio, Apr. 19, 2002.

SAYÃO, L. F. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In. MARCONDES, Carlos H. *et al*. (Org.). **Bibliotecas digitais**: saberes e práticas. Salvador: EDUFBA; Brasília: IBICT, 2005. p. 115-146.

SEARLE, S.; THOMPSON, D. Preservation metadata. Pragmatic First Steps at the National Library of New Zealand. **D-Lib Magazine**, v. 9, n. 4, Abr. 2003. Disponível em: <http://www.dlib.org/dlib/april03/thompson/04thompson.html>. Acesso em: 24 jan. 2006.

SELLTIZ, C. *et al.* **Métodos de pesquisa nas relações sociais**. São Paulo: EDUSP, 1975.

SILVERMAN, D. **Doing qualitative research**: a practical handbook. London: Sage Publications, 2000.

SIMPSON, D. **Contracting out for digital preservation services**: Information leaflet and checklist. [S. l.]: Digital Preservation Coalition, 2004.

SINCLAIR, K. **The VERS standards**. 2003. Disponível em: <http://vers.imagineering.com.au/erecord_library/library.htm#voapstandard>. Acesso em: 2 abr. 2006.

SMITH, A. Long term archiving of digital documents in physics. In: INTERNATIONAL UNION OF PURE AND APPLIED PHYSICS, 2001, Lyon, France: American Physical Society, 2001. Report of the Meeting sponsored by the Working Group on Communication in Physics. Disponível em: <http://publish.aps.org/IUPAP/Itaddp_report.html>. Acesso em: 13 jul. 2005.

SMITH, K. Institutional repositories and e-journal archiving: what are we learning? **The Journal of Electronic Publishing**, v. 11, n. 1, Winter 2008. Disponível em: <http://hdl.handle.net/2027/spo.3336451.0011.107>. Acesso em: 16 maio 2005.

SMITH, M. Practical preservation activities. In: WORKSHOP ON LONG-TERM CURATION WHITIN DIGITAL REPOSITORIES, 2005. **Resumos**… Cambridge, UK: MIT Libraries, 2005.

STAPLES, T; WAYLAND, R; PAYETTE, S. The Fedora Project: an open source digital repository management system. **D-Lib Magazine**, v. 9, n. 4, Apr. 2003. Disponível em: <http://www.dlib.org/dlib/april03/staples/04staples.html>. Acesso em: 14 nov. 2005.

STEENBAKKERS, J. F. Preserving electronic publications. **Information Services & Use**, v. 22. p. 89-96, Jan. 2002.

_____. **Setting up a deposit system for electronic publications**: the NEDLIB guidelines. NEDLIB Report Series 5, 2000. Disponível em: <http://www.kb.nl/coop/nedlib/results/NEDLIBguidelines.pdf>. Acesso em: 14 set. 2005.

STEPHENS, D. O; WALLACE, R. C. **Electronic records retention**: new strategies for data life cycle management. Lenexa, KS: ARMA International, 2003.

STINSON, D; AMELI, F; ZAINO, N. **Lifetime of KODAK Writable CD and Photo CD Media**. Rochester, NY: Eastman Kodak Company, Digital & Applied Imaging, 1995.

SUBER, P. **How should we define open access**? 2003. SPARC Open Access Newsletter. Disponível em: <https://mx2.arl.org/Lists/SPARC-OANews/Message/96.html> 2003. Acesso em: 18 de jun. de 2006.

SUTTON, S. Conceptual design and deployment of a metadata framework for educational resources on the Internet. **Journal of the American Society for Information Science**, v. 50, n. 13, p. 1182-1192, 1999.

SWADE, D. The problems of software conservation**. History and Computing**, v. 4, n. 2, 1992. Disponível em: <http://www.cs.man.ac.uk/CCS/simulate/sim_home.htm>. Acesso em: 14 ago. 2006.

TANSLEY, R., *et al.* The Dspace Institutional digital repository system: current functionality. In: CONFERENCE ON DIGITAL LIBRARIES, 2003, Los Alamitos, CA. **Proceeding**… Los Alamitos, CA: IEEE. p. 89-97.

TASK FORCE ON THE ARCHIVING OF DIGITAL INFORMATION. **Preserving digital information**: report of the task force on archiving of digital information. Washington: Commission on Preservation and Access, 1996. Disponível em: <http://www.rlg.org/ArchTF/>. Acesso em: 28 jul. 2004.

TEPER, T. H.; KRAMER, B. Long-term retention of electronic theses and dissertations. **College & Research Libraries**, v. 63, n. 1 p. 61-72, 2002.

TERRIO, R. Electronic metaphors and paper realities. **Progressive Librarian**, v. 21, p. 28-37, 2002.

THIBODEAU, K. **Overview of technological approaches to digital preservation and challenges in coming year**: what does it mean to preserve digital objects? Washington: Council on Library and Information Resources, 2002. Disponível em: <http://www.clir.org/pubs/reports/pub107/thibodeau.html>. Acesso em: 27 jan. 2004.

THOMAS, D. **Digital preservation at the National Archives**. London: The National Archives, [2003?]. Disponível em: <http://www.nationalarchives.gov.uk/preservation/digitalarchive/pdf/dpattna.pdf>. Acesso em: 7 ago. 2004.

THOMAZ, K. **A preservação de documentos eletrônicos de caráter arquivístico**: novos desafios, velhos problemas. 2004. 388 p. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

_____. Repositórios digitais confiáveis e certificação. **Arquivistica.net**, Rio de Janeiro, v. 3, n. 1, jan./jun.2007. Disponível em: <http://www.arquivistica.net/ojs/viewarticle.php?id=118>. Acesso em: 5 mar. 2008.

_____. **TECH-IN**. [mensagem pessoal]. Mensagem recebida por <miguel@ibict.br. em 1 nov. 2008.

_____; SOARES, A. J. A preservação digital e o modelo de referência Open Archival Information System (OAIS). **Revista de Ciência da Informação**, Brasília, v. 5, n. 1, fev. 2004. Disponível em: <http://www.dgz.org.br/fev04/Art_01.htm>. Acesso em: 7 jun. 2005.

UNESCO. **Carta para la preservación del patrimonio digital**. [Paris?]: UNESCO, 2003. Disponível em: <http://www.arquivonacional.gov.br/conarq/cam_tec_doc_ele/preservacao/cartapreservacao.asp>. Acesso em: 23 jul. 2004.

_____. The legal deposit of electronic publications, 1998. Disponível em: <http://www.unesco.org/webworld/memory/legaldep.htm>. Acesso em: 18 nov. 2006.

_____. **Recomendações sobre Software Livre para Repositório e Sistema de Preservação**. Paris: UNESCO, 2007.

VAN BOGART, J. W. **Magnetic tape storage and handling**: a guide for libraries and archives. Washington: Commission on Preservation and Access, 1995.

VAN DE SOMPEL, H.; LAGOZE, C. The Santa Fe Convention of the Open Archives Initiative. **D-lib Magazine**, v. 6, n. 2, Feb. 2000. Disponível em: <http://www.dlib.org>. Acesso em: 14 fev. 2006.

VAN NUYS, C. The Paradigma Project. **RLG DigiNews**, v. 7, n. 2, Apr. 2003. Disponível em: http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file3600.html#feature2. Acesso em: 31 out. 2008.

WAIBEL, G. Produce, publish and preserve: a holistic approach to digital assets management. **Journal of the Museum Computer Network,** Spectra, v. 26, n. 7, 2000. Disponível em: <http://www.bampfa.berkeley.edu/moac/imaging/index.html>. Acesso em: 7 ago. 2006.

WALKER, F. L.; THOMA, G. R. A web-based paradigm for file migration. In: ARCHIVING CONFERENCE, 2004, San Antonio, Texas, USA. **Proceedings**… San Antonio, Texas, USA: IS&T´s, 2004.

WATERS, D.; GARRETT, J. **Preserving digital information**: report of the task force on archiving of digital information commissioned. Washington, DC: Commission on Preservation and Access. 1996.

WEBB, C. **The role of preservation and the library of the future**. [S. l.]: National Library of Australia, 2000. Disponível em: <http://www.nla.gov.au/nla/staffpaper/cwebb9.html>. Acesso em: 17 nov. 2006.

WHEATLEY, P. **Institutional repositories in the context of digital preservation**. Leeds: University of Leeds, 2004.

_____. **A way forward for developments in the digital preservation functions of dspace**: options, issues and recommendations. 2003. Disponível em: <http://dspace.org/news/articles/DpAndDSpace.pdf>. Acesso em: 14 jul. 2007.

_____. Migration: a camileon discussion paper. **Ariadne**, 2001. Disponível em: <http://www.ariadne.ac.uk/issue29/camileon/>. Acesso em: 1 ago. 2006.

WIGGINS, Richard. Digital preservation: paradox and promise. **Library Journal**, p. 12-15, spring 2001.

WIPO COPYRIGHT TREATY. In: DIPLOMATIC CONFERENCE ON CERTAIN COPYRIGHT AND NEIGHBORING RIGHTS QUESTIONS, 10., 1996, Geneva. **Anais…** Geneva: International Bureau of WIPO, 1996. Disponível em: <http://www.wipo.int/eng/diplconf/distrib/94dc.htm>. Acesso em: 10 mar. 2005.

WOODS, A. When preserving the past: think of the future. **Multimedia Information and Technology**, v. 28, n. 3, p. 71-72, 1998.

ZENG, M. Metadata elements for object description and representation: a case report from a digitized historical fashion collection project. **Journal of the American Society for Information Science**, v. 50, p. 1193-1208, 1999.

APÊNDICE A – Questionário

Levantamento de dados

Instituição:
Endereço:
Site:
Responsável pelo(a) Biblioteca/Centro de Informação/Arquivo

Número de servidores Biblioteca/Centro de Informação/Arquivo:			
Técnico (Bibliotecários)		**Auxiliares:**	

Situação administrativa da Biblioteca/Centro de Informação/Arquivo:	
( ) Central	( ) Setorial
( ) especializada	( ) abrangência geral

Informações quanto ao tratamento dado aos seguintes tipos de materiais:

TIPO DE MATERIAL	SISTEMA ONLINE? QUAL SOFTWARE UTILIZA?	SETOR RESPONSÁVEL	UTILIZA FERRAMENTA INDEXAÇÃO? QUAL?	FORMA DE ACESSO AO CONTEÚDO	
				BIBLIOGRÁFICO	TEXTO COMPLETO
Monografias (livros/ anais/ teses/dissertações)					
Analíticas (capítulos de livros)					
Periódicos					
Artigos revistas					
Comunicações em eventos					
Multimeios (filmes, CDs, DVDs, fotografias etc.)					
TCC/Trabalhos experimentais					
Produção docente					
Outros					

Obs.:

(1) sistema on-line => mencionar se os diversos tipos de materiais estão recebendo algum tipo de tratamento e se este tratamento já está automatizado. Se sim, qual o software utilizado.

(2) Setor responsável => existência de algum banco de dados interessante na área e que não necessariamente está na Biblioteca.

(3) Ferramenta de indexação => informar qual o vocabulário e/ou tesauro está sendo utilizado, e se há o desenvolvimento pela unidade de algum vocabulário e/ou tesauro específico para o tratamento e controle da indexação.

Classificação, armazenamento e conservação da produção cientifica da unidade		
	Igual a do acervo geral	Específico para a coleção
Classificação		
Forma de armazenamento		
Metodologia para conservação (em qualquer suporte)		

Caso utilize um tratamento especifico para a coleção, informe como é feito:

Classificação	
Forma de armazenamento	
Metodologia para conservação (em qualquer suporte)	

Descrição do tratamento dado aos materiais digitais:		
	Sim	**Não**
Aplicação de metodologia de digitalização		
Aplicação de estratégia de preservação:		
Participação em iniciativa nacional/internacional de preservação digital:		

APÊNDICE B – Roteiros de entrevistas

Roteiros das entrevistas com representantes das unidades de pesquisa do MCT

Primeiro Roteiro de Entrevista
(Aplicado em eventos científicos)

Primeiro grupo: formação profissional e características socioespaciais das tarefas realizadas pelo entrevistado:

f) profissão
g) titulação máxima
h) tempo de serviço
i) cargo ocupado atualmente
j) tempo no cargo atual

Segundo grupo: descrição das atividades desenvolvidas pelos respondentes que envolvem algum tipo de tratamento técnico nos arquivos digitais da instituição:

g) Quais atividades científicas são realizadas pela sua instituição?
h) Que tipo de produção científica já está em formato digital?
i) Quem são os usuários dessa produção científica?
j) Quais atividades de preservação de documentos digitais estão sendo realizadas pela sua instituição?

Terceiro grupo: opiniões sobre o material a ser preservado e seus responsáveis:

k) Qual produção científica deve ser preservada pela sua instituição?
l) Quem são os responsáveis por essas atividades de preservação?

Segundo Roteiro de Entrevista
(Aplicado nas sedes das UPs)

1. Conferindo a identificação individual
(Nome, faixa etária, profissão, titulação, cargo, tempo de trabalho na instituição, tempo no cargo)

2. Conferindo a identificação da unidade
(Nome, localização, histórico)

3. Conferindo as informações institucionais
(Atividades científicas, linhas de pesquisa, tipo de produção científica, publicações)

4. Conferindo as informações institucionais do setor da Biblioteca/Centro de Informação e do Setor de Informática
(Dados na descrição oficial, tipo de usuários, site)

5. Atualização dos dados coletados no questionário enviado em 2004

5.1 Funcionários
5.2 Tipo de biblioteca
5.3 Tipo de material/ferramentas de software
5.4 Tipo de material/ setor responsável
5.5 Tipo de material/Ferramenta de indexação
5.6 Tipo de material/Tipo de registro para acesso

5.7 Tratamento da produção científica (classificação, armazenamento e conservação)
5.8 Tratamento da produção científica em formato digital (digitalização, estratégia de preservação, participação em iniciativa)

6. **Descrição do registro da produção científica da instituição** (gerenciamento do registro da produção)

7. **Descrição do ciclo de vida da produção científica em formato digital na instituição** (tipo de documentos, tratamento)

8. **Funcionários do setor de informática que atendem à biblioteca**
   8.1 Tipo de atendimento
   8.2 Tipo de *software*/infra-estrutura
   8.3 Setor responsável
   8.4 Gerenciamento de sistemas
   5.6 Segurança
   5.7 Recuperação de desastres
   5.8 Suporte à preservação digital (Formatos de identificação, validação, extração de metadados, normalização de entrada de dados; replicação de conteúdo; migração de versões; emulação).

9. **Descrição do tratamento dos acervos digitais pelo setor de suporte técnico da instituição** (atividades de armazenamento, segurança e preservação dos dados)

10. **Visão pessoal sobre a preservação digital da produção técnico-científica na unidade** (Opinião e plano de preservação digital para a instituição)

APÊNDICE C – Lista de projetos e iniciativas de preservação digital

1. **ArchiSig**
   URL: www.archisig.de/english/index.html

2. **Archiving the Avant Garde**
   URL: http://www.bampfa.berkeley.edu/about/avantgarde

3. **ARELDA Project**
   URL: http://www.bar.admin.ch/themen/00532/00536/index.html?lang=fr

4. **ARTISTE**
   URL: http://www.it-innovation.soton.ac.uk/projects/artiste/

5. **Assessment of UK Data Archive and The National Archives compliance with OAIS/METS**
   URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/assetmanagement/oais.aspx

6. **Auditing and Certification of Digital Archives**
   URL: http://digitalarchive.oclc.org/

7. **Automatic metadata extraction**
   URL: http://www.springerlink.com/content/j75r8rw691664273/

8. **BBC Domesday Rescue Project**
   URL:http://old.cs.ncl.ac.uk/research/events/colloquia/abstract.php?id=139

9. **Biblioteca Universalis**
   URL:http://www1.bncf.firenze.sbn.it/progetti/Bib_Uni/prototipo/pag13_ns.htm

10. **British Broadcasting Corporation (BBC)**
    URL: http://en.wikipedia.org/wiki/BBC

11. **British Library DOM (Digital Object Management)**
    URL: http://www.bl.uk/aboutus/stratpolprog/digi/dom/index.html

12. **Cairo (Complex Archive Ingest for Repository Objects)**
    URL: http://cairo.paradigm.ac.uk/

13. **Camileon**
    URL: http://www.si.umich.edu/CAMILEON/

14. **Canadian National Digital Inventory**
    URL: http://www.library.uiuc.edu/spx/inventory/projects.html

15. **CEDARS - CURL Exemplars in Digital Archives**
    URL: http://www.leeds.ac.uk/cedars/

16. **CENSA (Collaborative Electronic Notebook Systems Associations)**
    URL: http://www.censa.org/

17. **CHIN: Creation and Managing Digital Content**
    URL: http://www.chin.gc.ca/English/About_Chin/chin.html

18. **Clinical Data Interchange Standards Consortium (CDISC)**
    URL: http://www.cdisc.org/

19. **CMS Metadata Interoperability Project: Ensuring Metadata Interoperability Across Scottish Content Management Systems and Digital Repositories**
    URL: http://cms.cdlr.strath.ac.uk/

20. **Conservation Online (CoOL)**
    URL: http://palimpsest.stanford.edu/

21. **Council for the Central Laboratory of the Research Councils (CCLRC)**
    URL:http://www.hero.ac.uk/uk/reference_and_subject_resources/resources/opacs_in_britain_and_ireland_obi/obi_council_for_the_central_la4756.cfm

22. **DAAT: Digital Asset Assessment Tool**
    URL: ahds.ac.uk/about/projects/daat/

23. **DELOS/NSF study on the digital preservation research agenda**
    URL: http://www.dpconline.org/graphics/whatsnew/issue12.html

24. **Digicult**
    URL: http://www.digicult.info/pages/index.php

25. **Digital Academic Repositories (DARE), 2003-2006**
    URL: http://digitalarchive.oclc.org/

26. **Digital Curation Centre**
    URL: http://digitalarchive.oclc.org/

27. **Digital Image Archive of Medieval Music (DIAMM)**
    URL: http://www.diamm.ac.uk/index.html

28. **Digital Images Archiving Study**
    URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/imagesarchiving

29. **Digital Information Archiving System (DIAS) Dutch National Library**
    URL: http://www-05.ibm.com/nl/dias/preservation2.html

30. **Digital Preservation**
    URL: http://www.digitalpreservation.gov/

31. **Digital Preservation Cluster, 2004-2006**
    URL: http://digitalarchive.oclc.org/

32. **Digital Preservation Coalition**
    URL: http://www.dpconline.org/graphics/index.html

33. **Digital Preservation Training Programme**
    URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/assetmanagement/dptp.aspx

34. **DiVA Electronic Publishing Centre, Uppsala University Library**
    URL: DiVA Electronic Publishing Centre, Uppsala University Library

35. **Dlib (D-Lib Forum)**
URL: http://www.dlib.org/

36. **Domain UK**
URL: http://bibnum.bnf.fr/ecdl/2002/uk/uk.html

37. **DOMEA: Dokumentenmanagement und elektronische Archivierung**
URL: http://www.uni-koblenz.de/fvi/projekte/domea.html

38. **DOMS (National Library of Scotland)**
URL: http://en.wikipedia.org/wiki/National_Library_of_Scotland

39. **Dorothy Dunnett Disks Project**
URL: http://www.dpconline.org/graphics/join/projects.html

40. **DSpace MIT**
URL: http://dspace.mit.edu/

41. **DSpace@ Cambridge**
URL: http://www.lib.cam.ac.uk/dspace/

42. **Dublin Core Metadata Initiative**
URL: http://dublincore.org/

43. **Earth Resources Observation Systems (EROS) Data Center**
URL: http://edc.usgs.gov/

44. **ECHO DEPository Project, 2004-2007**
URL: http://www.ndiipp.uiuc.edu/

45. **E-Journal Archiving Study**
URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/ejournalarchiving.aspx

46. **E-Learning Project**
URL: http://www.browsemystuff.com/browse/elearning/elearning-project/

47. **ERA: Electronic Records Archive Program**
URL: http://www.diglib.org/preserve/ERA2004.htm

48. **EROS: Electronic Records in Office Systems**
URL: http://www.nationalarchives.gov.uk/recordsmanagement/

49. **ERPANET Project**
URL: http://www.erpanet.org/

50. **E-Science Data Curation Audit**
URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/escience.aspx

51. **European Museums Information Institute**
URL: http://www.emii.org/map/

52. **EVA project**
URL: http://www.evaproject.it/

53. **Feasibility Study On E-Prints**
URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/eprints.aspx

54. **FEDORA**
URL: http://fedoraproject.org/

55. **Fedora™ (Flexible Extensible Digital Object Repository Architecture) Cornell University and the University of Virginia Library**
URL: http://blog.eukhost.com/webhosting/flexible-extensible-digital-object-repository-architecture/

56. **File integrity and identification**
URL: http://www.freepatentsonline.com/7043495.html

57. **HEDS - Higher Education Digitisation Service**
URL: http://heds.herts.ac.uk/

58. **Hybrid Archives Project**
URL: http://www.webarchive.org.uk/tep/15939.html

59. **ICA work-book on ERM**
URL: http://www.dpconline.org/graphics/join/projects.html

60. **Incentives for Data Producers to Create Archive-Ready Data Sets,2005-2006**
URL: http://digitalarchive.oclc.org/

61. **International Internet Preservation Consortium (IIPC)**
URL: http://www.netpreserve.org/about/index.php

62. **International Research on permanent Records in Electronic Systems**
URL:http://www.interpares.org/

63. **InterPARES**
URL: http://www.interpares.org/

64. **IRI Scotland Project**
URL:
http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/iriscotland.aspx

65. **JSTOR**
URL: http://www.jstor.org/

66. **MALIBU - Managing the hybrid library for the benefit of users**
URL: http://cat.inist.fr/?aModele=afficheN&cpsidt=2000614

67. **mandate: Managing Digital Assets in Tertiary Education**
URL: http://www.jisc.ac.uk/publications/publications/mandatefinalreport.aspx

68. **Metadata Generation Research (MGR)**
URL: ils.unc.edu/mrc/mgr.html

69. **Metadata Encoding & Transmission Standard (METS)**
URL: http://www.loc.gov/standards/mets/

70. **Metadata Watch (Forum for Metadata Schema Implementors)**

URL: http://www.schemas-forum.org/metadata-watch/

71. **METAe (The Metadata Engine Project)**
URL: http://meta-e.aib.uni-linz.ac.at/

72. **MetaNet**
URL: http://www.epros.ed.ac.uk/metanet/

73. **METS Awareness Training**
URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/assetmanagement/mets.aspx

74. **Models for Action Project**
URL:
http://www.ctg.albany.edu/publications/journals/models_for_action?chapter=1&section=1

75. **MoPark Metadata Options Appraisal**
URL: http://eprints.rclis.org/archive/00002278/

76. **Moving Pictures and Sound Archiving Study**
URL: http://64.233.169.104/search?q=cache:JTo9N4775yIJ:www.jisc-collections.ac.uk/uploaded_documents/MPS%2520vision.doc+Moving+Pictures+and+Sound+Archiving+Study&hl=pt-BR&ct=clnk&cd=7&gl=br

77. **National Archives (UK) Digital Archive**
URL: http://www.nationalarchives.gov.uk/preservation/digital.htm

78. **NDIIPP - National Digital Information Infrastructure and Preservation Program**
URL: http://www.lib.ncsu.edu/ndiipp/

79. **NSW Recordkeeping Metadata Standard**
URL:http://www.records.nsw.gov.au/recordkeeping/nsw_recordkeeping_metadata_standard_4614.asp

80. **OCLC Digital Archive**
URL: http://www.oclc.org/digitalarchive/

81. **OCLC Preservation Policy**
URL: http://www.nla.gov.au/padi/topics/172.html

82. **OCLC Preservation Resources**
URL: http://www.oclc.org/preservation/default.htm

83. **OCLC: Digital Preservation Education and Planning**
URL: http://www.oclc.org/preservation/education/default.htm

84. **OCLC: Incentives to Preserve Digital Materials**
URL: http://jobfunctions.bnet.com/abstract.aspx?docid=161155

85. **OCLC: Registry of Digital Masters**
URL: http://www.oclc.org/digitalregistry/

86. **Oxford Digital Library**
URL: http://www.odl.ox.ac.uk/

87. **PADI (Preservation Access to Digital Information)**
URL: http://www.nla.gov.au/padi/

88. **PANDORA - National Library of Australia**
URL: http://www.nla.gov.au/policy/pandje97.html

89. **PANDORA Archive (Preserving and Accessing Networked Documentary Resources of Australia)**
URL: http://pandora.nla.gov.au/

90. **Persistent Identifiers**
URL: http://www.nla.gov.au/initiatives/persistence.html

91. **Personal Archives Accessible in Digital Media (paradigm)**
URL: http://www.paradigm.ac.uk/

92. **PREMIS**
URL: http://www.oclc.org/research/projects/pmwg/

93. **PRESERV (PREServation Eprint SERVices)**
URL: http://preserv.eprints.org/

94. **Preservating Digital Reformatting Program**
URL: http://www.loc.gov/preserv/prd/presdig/presintro.html

95. **Preservation Metadata Working Group**
URL: http://www.oclc.org/research/projects/pmwg/wg1.htm

96. **PRESTO Preservation Technologies**
URL: http://www.it-innovation.soton.ac.uk/projects/prestospace/

97. **Profiles in Science, National Library of Medicine**
URL: http://www.encyclopedia.com/doc/1G1-83583418.html

98. **Project Open Book**
URL: http://openbookproject.net/

99. **PRONOM**
URL: http://www.nationalarchives.gov.uk/pronom/

100. **PubMed Central**
URL: http://www.pubmedcentral.nih.gov/

101. **Records lifecycle**
URL: http://www.city.ac.uk/ic/rm/recordslifecycle.html

102. **Representation and Rendering Project (file formats)**
URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/fileformat.aspx

103. **Risk analysis of corporate-wide file types Rosetta Stone**
URL: http://www.rosettastone.com/

104. **Rosetta Stone**
URL: http://www.rosettastone.com/

105. **Runeberg**
URL: http://runeberg.org/

106. **Safeguarding European Photographic Images for Access (SEPIA)**
URL: http://www.knaw.nl/ecpa/sepia/

107. **Scottish Party Political Press Releases Preservation and Access pilot**
URL: http://www.dpconline.org/graphics/join/projects.html

108. **Server replication for back-ups**
URL: http://www.onlamp.com/pub/a/onlamp/2005/06/16/MySQLian.html

109. **SHERPA DP**
URL: *ahds.ac.uk/about/projects/**sherpa-dp**/index.html*

110. **Sound Archive acetate disks**
URL: http://www.fonoteca.ch/yellow/SoundCarriers/acetate_en.htm

111. **STARGATE**
URL: http://www.imdb.com/title/tt0111282/ http://stargate.mgm.com/

112. **Strategic Partnerships With Industry – Research & Training – SPIRT – Australian RecordKeeping Metadata Schema**
URL: http://www.sims.monash.edu.au/research/rcrg/research/spirt/

113. **UKDA**
URL: http://www.data-archive.ac.uk/

114. **UKWAC (UK Web Archiving Consortium)**
URL: http://www.webarchive.org.uk/

115. **ULCC**
URL: http://www.ulcc.ac.uk/

116. **Universal Virtual Com-puter (UVC)**
URL: http://www.nla.gov.au/padi/topics/492.html

117. **VERS: Victorian Electronic Records Strategy The Victorian Electronic Records System (VERS)**
URL: http://www.prov.vic.gov.au/vers/standard/advice_12/7-2.htm

118. **Voluntary Deposit of Electronic Publications (VDEP)**
URL: http://www.dpconline.org/graphics/join/life.html

119. **Web Archiving: UK Central Government Web Archive**
URL: http://www.nationalarchives.gov.uk/nowhere/

120. **Web-archiving feasibility study**
URL:http://www.jisc.ac.uk/whatwedo/programmes/preservation/webarchiving.aspx

APÊNDICE D – Centros de informação e bibliotecas das unidades de pesquisa do Ministério da Ciência e Tecnologia

Nome:	**Agência Espacial Brasileira - AEB**
Biblioteca:	Diretoria Planejamento Orçamento e Administração - DPOA
Endereço:	SPO - Área 5 - Quadra 3 Bloco A, 1º andar CEP 70610-200 – Brasília DF
URL:	http://www.aeb.gov.br
Descrição:	**Missão** Foi criada em 10 de fevereiro de 1994, pela Lei Nº 8.854. Atendendo a uma antiga reivindicação da comunidade científica e tecnológica brasileira, a AEB nasceu com o objetivo de promover o desenvolvimento das atividades espaciais brasileiras de forma descentralizada. Autarquia vinculada ao Ministério da Ciência e Tecnologia **Unidade de documentação** Não tem unidade de documentação. **Publicações on-line** Permite acesso on-line aos seus relatórios de gestão e outras publicações.

Nome:	**Centro Brasileiro de Pesquisa Física - CBPF**
Biblioteca:	Coordenação de Documentação e Informação Científica – CDI Biblioteca do CBPF
Endereço:	Rua Dr. Xavier Sigaud, 150 Ed. César Lattes, Urca 22290.180 - Rio de Janeiro - RJ
URL:	http://www.biblioteca.cbpf.br
Descrição:	**Missão** Realizar pesquisa básica em Física e desenvolver suas aplicações, atuando como instituto nacional de Física do Ministério da Ciência e Tecnologia e pólo de investigação científica e de formação, treinamento e aperfeiçoamento de pessoal científico. Em 1976 recebeu recursos financeiros do governo federal para a manutenção das linhas de pesquisa e implantação dos laboratórios de cálculo numérico e de instrumentação. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** A Biblioteca do Centro Brasileiro de Pesquisas Físicas, fundada em 1949, é uma das bibliotecas nacionais de referência na área de Física, tendo atualmente um acervo de aproximadamente 21.000 livros e 850 títulos de periódicos (240 correntes), em Física e ciências afins. Parcialmente destruída em 1959 por um incêndio, foi reconstruída graças à solidariedade da comunidade nacional e internacional. Constam de seu acervo periódicos de grande importância histórica, como *The Philosophical Magazine* (1800), *Comptes Rendus des Séances de L'Académie des Sciences* (1835), *Philosophical Transactions of the Royal Society of London* (1875), *Proceedings of the Royal Society of London* (1877), entre outros.

	Outras informações A Área de Informação Científica (AIC) tem por finalidade promover o intercâmbio da documentação e informação no campo da Física, e desenvolver trabalhos científicos e técnicos nesta área de informação. Para atingir esse objetivo a AIC coleta, organiza, atualiza e divulga serviços e produtos necessários à obtenção de informações. Dentre os seus serviços e produtos destaca-se a **Memória da Produção Técnico-Científica do CBPF** cujo objetivo é registrar e disseminar todo trabalho produzido ou em andamento, de forma escrita ou oral, dos pesquisadores, técnicos e alunos de pós-graduação do CBPF, isoladamente ou em colaboração com pesquisadores brasileiros e estrangeiros de outras instituições. Cada volume da Memória cobre o período de um ano. A AIC também disponibiliza os *preprints* recebidos de instituições nacionais e estrangeiras através de exposição organizada por instituição Disponibiliza acesso on-line aos seus relatórios de gestão e outras publicações em sua página na *web*. **Publicações on-line** Permite acesso on-line aos seus relatórios de gestão e outras publicações.	
Equipe:	Alberto Passos Guimarães Filho – Coordenador Justina de Fátima Bacellar Couto – Coordenadora Substituta	

Nome:	**Centro de Excelência em Tecnologia Eletrônica Avançada - CEITEC**	
Biblioteca:		
Endereço:	Estrada João de Oliveira Remião, 777, Porto Alegre - RS – Brasil - Cep: 91550-000	
URL:	http://www.ceitecmicrossistemas.org.br/portal/home.php	
Descrição:	**Missão** Desenvolver soluções inovadoras em eletrônica e tecnologias da informação, com alto padrão de qualidade, atendendo às necessidades do mercado. Empresa vinculada ao Ministério da Ciência e Tecnologia **Unidade de documentação** Não possui unidade de documentação. **Publicações on-line** Permite acesso on-line a artigos e trabalhos apresentados em seminários, worshop etc.	

Nome:	**Centro de Tecnologia da Informação Renato Archer – CTI (antigo CENPRA)**
Biblioteca:	Biblioteca Sede
Endereço:	Rodovia D. Pedro I, km 143,6 – Amarais CEP 13082-120 Campinas - SP
URL:	http://www.cenpra.gov.br
Descrição:	**Missão** Gerar, aplicar e disseminar conhecimentos em Tecnologia da Informação, em articulação com os agentes socioeconômicos, promovendo inovações que atendam às necessidades da sociedade. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Possui unidade de documentação, mas não disponibiliza base de dados para acesso on-line. Disponibiliza artigos e trabalhos em sua páginas na *web*. **Publicações on-line** Permite acesso on-line a artigos e outros trabalhos, de autoria dos seus colaboradores, publicados em revista especializada.
Equipe:	1 Bibliotecária responsável

Unidade:	**Centro de Tecnologia Mineral - CETEM**
Biblioteca:	Serviço de Informação - SEDIN
Endereço:	Avenida Ipê, no. 900 - Ilha da Cidade Universitária 21941-590 - Rio de Janeiro - RJ
URL:	http://www.cetem.gov.br/biblioteca.htm
Descrição:	**Missão** Desenvolver tecnologia para o processamento do uso sustentável dos recursos minerais brasileiros. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** A biblioteca destina-se a dar apoio a todas as atividades de pesquisa do CETEM. É uma biblioteca de referência, especializada na área de tecnologia mineral, processos mínero-metalúrgicos e tecnologia ambiental. A biblioteca conta com um acervo de 9178 livros, mantém 97 assinaturas de periódicos científicos especializados e reúne toda a produção técnico-científica da organização, constituída de 5459 relatórios técnicos. Também fazem parte de seu acervo às teses dos colaboradores e os 221 volumes editados através das Séries Qualidades e Produtividade, Estudos e Documentos, Tecnologia Ambiental, Tecnologia Mineral, Rochas e Minerais Industriais, Jornadas de Iniciação Científica e avulsos. **Publicações on-line** Permite acesso on-line a publicações editadas pelo CETEM.

Equipe:	Coordenação da Biblioteca - Jackson de Figueiredo Neto Responsável da Biblioteca - Ana Maria Sá de Oliveira Processamento Técnico - Sônia Maria Mamede Lourenço	

Nome:	**Centro de Excelência em Tecnologia Eletrônica Avançada - CETENE**	
Biblioteca:		
Endereço:	Av. Professor Luís Freire, 1 - Cidade Universitária – CEP. 50740-540 - Recife - PE	
URL:	http://www.inpe.br	
Descrição:	**Missão** Desenvolver, introduzir e aperfeiçoar inovações tecnológicas que tenham caráter estratégico para o desenvolvimento econômico e social da região Nordeste, promovendo cooperações baseadas em redes de conhecimento e nos agentes da economia nordestina. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Não possui unidade de documentação. **Publicações on-line** Permite acesso on-line a fotos, relatório, vídeos e artigos através da Biblioteca Virtual.	

Nome:	**Centro de Gestão e Estudos Estratégicos - CGEE**	
Biblioteca:	Núcleo Centro de informações	
Endereço:	SCN Quadra 2, Bloco A, Ed. Corporate Center Financial Center, 11º andar, Salas 1102 e 1103 CEP 70712-900 - Brasília-DF	
URL:	http://www.cgee.org.br	
Descrição:	**Missão** Promover e realizar estudos e pesquisas prospectivas de alto nível na área de ciência e tecnologia e suas relações com setores produtivos, bem como realizar atividades de avaliação de estratégicas e de impactos econômicos e sociais das políticas, programas e projetos científicos e tecnológicos, além de difundir informações, experiências e projetos à sociedade. Organização Social, vinculada ao Ministério da Ciência e Tecnologia. **Unidade de documentação** Não possui unidade de documentação. **Publicações on-line** Permite acesso on-line a publicações editadas pelo CGEE.	
Equipe:	1 Cientista da Informação 1 Analista 1 Bibliotecária 2 Técnicos de informação	

Nome:	**Comissão Nacional de Energia Nuclear - CNEN**
Biblioteca:	Biblioteca Sede Unidades subordinadas à Comissão Nacional de Energia Nuclear em 2007: CRCN-NE; CRCN-CO; CDTN ; IEN ; IPEN; IRD Até 2006 eram 14 unidades vinculadas ao CNEN.
Endereço:	Rua General Severiano, 90 – Botafogo CEP 22294-900 - Rio de Janeiro - RJ
URL:	http://www.cnen.gov.br
Descrição:	**Missão** Garantir o uso seguro e pacífico a energia nuclear, desenvolver e disponibilizar tecnologia nuclear e correlata, visando ao bem-estar da população. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Encontra-se em fase de implantação a Rede de Bibliotecas da CNEN, cujo objetivo é proporcionar um funcionamento integrado, com otimização dos processos técnicos e administrativos de gerenciamento das coleções, ampliando a abrangência de consulta e serviços de informações bibliográficas especializadas prestadas à comunidade científica. A Rede é composta por sete bibliotecas, sendo coordenada pelo Centro de Informações Nucleares - CIN. • Centro de Desenvolvimento da Tecnologia Nuclear – CDTN Belo Horizonte - MG biblio@cdtn.br • Centro Regional de Ciências Nucleares do Centro-Oeste - CRCN-CO Goiânia - GO elaine@cnen.gov.br • Centro Regional de Ciências Nucleares do Nordeste – CRCN-NE Recife - PE bibliotecacrcn@cnen.gov.br • Instituto de Engenharia Nuclear – IEN Rio de Janeiro - RJ biblioteca@ien.gov.br • Instituto de Pesquisas Energéticas e Nucleares – IPEN São Paulo - SP bibli@ipen.br • Instituto de Radioproteção e Dosimetria - IRD Rio de Janeiro - RJ bibli@ird.gov.br • Sede Rio de Janeiro - RJ btc@cnen.gov.br] Disponibiliza **Outras informações** **ANAIS - CATÁLOGO COLETIVO DE ANAIS DE EVENTOS** Base de dados cooperativa, coordenada pelo CIN/CNEN e aberta a participação de instituições que desejem compartilhar seus acervos. Esta base de dados inclui dados bibliográficos de todo tipo de anais de eventos como conferências, simpósios, congressos, reuniões e outros realizados no Brasil ou no exterior e que se encontram disponíveis nos acervos de bibliotecas brasileiras e de algumas latino-americanas. As informações contidas nos registros são fornecidas pelas bibliotecas cooperantes e padronizadas pelo CIN. **Período**: Ilimitado - inclui informações a partir de qualquer data. **Atualização**: mensal **CATÁLOGO COLETIVO DE NORMAS TÉCNICAS** Base de dados cooperativa, coordenada pelo CIN/CNEN e aberta a participação de instituições que desejem compartilhar seus acervos. Arrola informações sobre normas técnicas

	nacionais e estrangeiras existentes nos acervos de bibliotecas brasileiras. As informações contidas nos registros são fornecidas pelas bibliotecas cooperantes e padronizadas pelo CIN. **Período**: Ilimitado: inclui informações a partir de qualquer data. **Atualização**: mensal **Base INIS - Energia Nuclear** Sistema cooperativo com a participação de mais de 120 países e organizações internacionais. O principal produto do sistema é à base de dados bibliográficos INIS. Esta base tem a mais completa cobertura sobre aplicações pacíficas da ciência e tecnologia nucleares. Atualmente, a base contem mais de 2,9 milhões de registros sendo a fonte mundial mais completa sobre literatura técnico-científica na área nuclear. Através do CIN, Centro de Informações Nucleares, serviços de informação técnico-científica com a base de dados INIS estão disponíveis gratuitamente no Brasil. **Base Energy - Tecnologias de Energia** A participação da Comissão Nacional de Energia Nuclear (CNEN) no Energy Technology Data Exchange (ETDE) tornou disponível no Brasil a base de dados ENERGY, formada por este sistema e abrangendo todas as tecnologias de energia. A CNEN, através do CIN - Centro de Informações Nucleares incorpora a literatura brasileira à base de dados ENERGY, divulgando, no país e no exterior a produção científica brasileira. O acesso à base de dados via Internet, é gratuito no Brasil para os usuários cadastrados. Os demais países que participam do ETDE, incluindo sua literatura nacional na base Energy, são Alemanha, Canadá, Coréia, Dinamarca, Espanha, Estados Unidos, Finlândia, França, Holanda, Inglaterra, México, Noruega, Portugal, Suécia e Suíça. **LivRe! - Revistas na Web** Relaciona periódicos eletrônicos de acesso livre na Internet. **PERIÓDICOS - SEDE** Contém os títulos das revistas e os fascículos existentes na biblioteca da Sede. Também estão incluídos nesta base as séries da IAEA como Safety Series e Technical Report Series e a coleção de relatórios anuais de instituições nacionais e estrangeiras. **Atualização**: mensal **RELATÓRIOS** Contém informações sobre os relatórios técnico-científicos e as normas técnicas pertencentes ao acervo da Biblioteca da CNEN/CIN. **Período**: Ilimitado - inclui informações a partir de qualquer data. **Atualização**: mensal	
Equipe:	Maria Emilia Frade de Mello	

Nome:	**Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq**	
Biblioteca:	Biblioteca Lygia Portocarrero Velloso	
Endereço:	SEPN 507, Bloco "B", Ed. Sede CNPq CEP 70.740-901 - Brasília - DF	
URL:	http://www.cnpq.br/	
Descrição:	**Missão**	

	Promover e fomentar o desenvolvimento científico e tecnológico do País e contribuir na formulação das políticas nacionais de ciência e tecnologia. Fundação vinculada ao Ministério da Ciência e Tecnologia **Unidade de documentação** Inicialmente, o acervo do CPO constitui-se a partir do acervo do Centro de Documentação - CDO - criado em 1977 e extinto em 1980, e, sobretudo da biblioteca trazida da antiga SUP - Superintendência de Planejamento, que funcionava no Rio de Janeiro. O acervo também era formado de publicações do Conselho, mesmo as anteriores a criação do CPO, e de doações pessoais. (Perspicillum, 1987). Após a extinção arbitraria do CPO, seu acervo foi transferido para o IBICT e teve parte de seus periódicos incorporados a Biblioteca deste. Posteriormente o acervo foi transferido para a Biblioteca do Ministério da Ciência e Tecnologia - MCT, para finalmente retornar ao CNPq em abril de 2004. Na sua configuração atual a Biblioteca Lygia Portocarrero Velloso possui um recorte temático com obras especializadas em políticas e história da ciência e tecnologia, com cerca de 30.000 volumes, como por exemplo, livros nacionais e estrangeiros, periódicos especializados, publicações de técnicos do CNPq, relatórios, obras de referência, entre outros. Em 2001 foi criado o Centro de Memória Institucional do CNPq, vinculado à vice-presidência. **Publicações on-line** Permite acesso on-line a publicações editadas pelo CNPq	
Equipe:	Roberto Muniz Barretto de Carvalho - Chefe do Serviço de Documentação e Acervo – SEDOC Maria Aliette Peixoto Wanderley – Responsável pela Biblioteca Patrícia Simas de Andrade - Bibliotecária	

Nome:	**Financiadora de Estudos e Projetos - FINEP**	
Biblioteca:	Biblioteca – Serviço de Informação	
Endereço:	Praia do Flamengo, nº 200 CEP 20210-030 Rio de Janeiro – RJ	
URL:	http://www.finep.gov.br/	
Descrição:	**Missão** Promover e financiar a inovação e pesquisa científica e tecnológica em empresas, universidades, institutos tecnológicos, centros de pesquisa e outras instituições públicas ou privadas, mobilizando recursos financeiros e integrando instrumentos para o desenvolvimento econômico e social do País. Empresa pública vinculada ao Ministério da Ciência e Tecnologia **Unidade de documentação** A Biblioteca foi criada em 1973 e tem como objetivo assessorar e apoiar as atividades desenvolvidas pela FINEP na busca, acesso e disseminação do seu acervo, contribuindo para geração de novos conhecimentos. Desde 1997, a Biblioteca	

	utiliza o Sistema SYSBIBLI, que permite a apresentação do catálogo on line disponível na internet. A coleção está orientada para o desenvolvimento de temas relacionados à Ciência e Tecnologia e o acervo de aproximadamente 20.000 documentos é composto de livros, periódicos, artigos e material multimídia. **Outras informações** **Biblioteca Virtual de Inovação Tecnológica** É resultado de uma ação conjunta envolvendo a FINEP e o Programa Prossiga/CNPq. Reúnem de forma estruturada e seletiva, sites brasileiros e estrangeiros contendo informações relevantes sobre inovação tecnológica, em seus múltiplos aspectos. **Portal Capital de Risco Brasil** Tem como objetivo a construção de um ambiente institucional que favoreça o desenvolvimento do capital de risco, de forma a estimular o fortalecimento das empresas nascentes e emergentes de base tecnológica brasileiras, contribuindo, em última instância, para o desenvolvimento tecnológico nacional, bem como para a geração de emprego e renda. **Publicações on-line** Permite acesso on-line a publicações editadas pelo FINEP	
Equipe:	2 Bibliotecárias	

Nome:	**Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT**	
Biblioteca:	Biblioteca	
Endereço:	SAS Quadra 5, Lote 6, Bloco H 70070.912 - Brasília - DF	
URL:	http://www.ibict.br	
Descrição:	**Missão** Promover a competência e o desenvolvimento de recursos e infra-estrutura de informação em Ciência e Tecnologia para a produção, a socialização e a integração do conhecimento científico-tecnológico. Unidade de pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Criada no mesmo ano da inauguração do Instituto – 1954. Atuar como biblioteca especializada em ciência da informação e biblioteconomia, tendo em seu acervo monografias, periódicos, anais de eventos, normas técnicas, materiais alternativos, obras de referência, teses e dissertações em ciência da informação, memória técnica. Em 23 de fevereiro de 2005, a Biblioteca do Ibict automatizou seus produtos e serviços com a aquisição do software PHL (Personal Home Library), incorporando a base Literatura em Ciência da Informação (LICI). **Publicações on-line** Permite acesso on-line as revistas: Ciência da Informação; Inclusão Social e Liinc	

Equipe:	Priscilla Mara Bermudes Araújo - Chefe do Setor da Biblioteca Suzelayne Eustáquio de Azevedo - Atendimento Walter Alves Nascimento Junior - Atendimento	

Nome:	**Instituto de Desenvolvimento Sustentável Mamirauá - IDSM**	
Biblioteca:	Biblioteca Sede	
Endereço:	UFPA Campus do Guamá Departamento de Antropologia (CFCH) Caixa Postal: 531 66073.250 - Belém - PA	
URL:	http://www.mamiraua.org.br/	
Descrição:	**Missão** Promover pesquisa científica para a conservação da biodiversidade por meio de manejo participativo e sustentável dos recursos naturais na Amazônia. Organização Social vinculada ao Ministério da Ciência e Tecnologia **Unidade de documentação** Não possuía biblioteca até 2007 **Publicações on-line** Permite acesso on-line a artigos e outros trabalhos, de autoria dos seus colaboradores, publicados em revista especializada.	
Equipe:	Josivaldo Ferreira Modesto 1 Bibliotecária 1 Técnico auxiliar	

Nome:	**Associação Instituto de Matemática Pura e Aplicada - IMPA**	
Biblioteca:	Biblioteca Sede	
Endereço:	Estrada Dona Castorina, 110 Ed. Lélio Gama, Jardim Botânico 22460.320 - Rio de Janeiro - RJ	
URL:	http://www.impa.br/	
Descrição:	**Missão** Realizar pesquisas em ciências matemáticas e afins, formar pesquisadores, disseminar o conhecimento matemático em todos os seus níveis e integrá-lo com outras áreas da Ciência, Cultura, Educação e do setor produtivo. Organização Social vinculada ao Ministério da Ciência e Tecnologia **Unidade de documentação** Especializada em Matemática Pura e Aplicada, possui um notável acervo de cerca de 30.000 volumes em livros clássicos e modernos, além de cerca de 30.000 volumes em periódicos, correspondendo a 670 títulos dos mais importantes em Matemática e áreas científicas mais afins. É considerada de excelente nível em padrão internacional. **Publicações on-line** Permite acesso on-line a publicações editadas pelo FINEP	

Equipe:	Maria da Graça Araújo Penna Duarte 1 Bibliotecárias 2 Técnicos auxiliares 1 Analista	

Nome:	**Indústrias Nucleares do Brasil - INB**	
Biblioteca:		
Endereço:	Rua Mena Barreto, 161 – Botafogo - 22271-100 - Rio de Janeiro - RJ - Brasil	
URL:	http://www.inb.gov.br/	
Descrição:	**Missão** Exploração do urânio, desde a mineração e o beneficiamento primário até a produção e montagem dos elementos combustíveis que acionam os reatores de usinas nucleares. O conjunto dessas atividades constitui o Ciclo do Combustível Nuclear. Atua também na área de tratamento físico dos minerais pesados com a prospecção e pesquisa, lavra, industrialização e comercialização das areias monazíticas e obtenção de terras-raras. Empresa de economia mista, vinculada à Comissão Nacional de Energia Nuclear - (CNEN) e subordinada ao Ministério da Ciência e Tecnologia. **Unidade de documentação** O acervo da biblioteca é constituído basicamente de normas técnicas nacionais e estrangeiras para atender às necessidades das duas fábricas de combustível nuclear. **Publicações on-line** Não possui publicações para acesso on-line.	
Equipe:	Gilda Menezes	

Nome:	**Instituto Nacional de Pesquisa da Amazônia - INPA**	
Biblioteca:	Biblioteca	
Endereço:	Avenida André Araújo, 2936, Petrópolis Manaus, Amazonas CEP 69083-000 – Manaus - AM	
URL:	http://www.biblioteca.inpa.gov.br	
Descrição:	**Missão** Gerar e disseminar conhecimentos e tecnologias e capacitar recursos humanos para o desenvolvimento da Amazônia. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Criada em julho de 1954, possui um acervo com mais de 199 mil títulos voltados às Ciências Puras e Aplicadas com ênfase às Ciências Biológicas e reúne uma das maiores bibliografias nacionais sobre a Amazônia. A biblioteca constituiu-se a partir do acervo particular do eminente botânico João Barbosa Rodrigues (1842-1909), fundador do Museu Botânico do Amazonas e diretor do Museu Nacional do Rio de Janeiro.	

	A Divisão de Publicações Seriadas reúne no acervo 7.472 mil títulos de publicações periódicas, técnica científica. Dentre esses títulos encontra-se a coleção completa do periódico Acta Amazônica, que divulga os trabalhos dos pesquisadores do INPA e de outras instituições. Possui uma coleção de Obras Raras e Especiais que se destaca por sua importância científica e institucional. O acervo de aproximadamente 3 mil exemplares reúne raridades dos séculos XVII, XVIII e XIX. Encontram-se obras clássicas de Wallace, Bates e Darwi. **Publicações on-line** Permite acesso on-line a publicações editadas pelo INPA, em destaque a revista Acta Amazonica	
Equipe:	1 Bibliotecária responsável	

Nome:	**Instituto Nacional de Pesquisas Espaciais - INPE**	
Biblioteca:	Serviço de Informação e Documentação - SID	
Endereço:	Av. dos Astronautas,1.758 12227.010 - São José dos Campos - SP	
URL:	http://www.inpe.br/biblioteca/	
Descrição:	**Missão** Promover e executar estudos, pesquisas científicas, desenvolvimento tecnológico e capacitação de recursos humanos nos campos da Ciência Espacial e da Atmosfera, das Aplicações Espaciais, da Meteorologia e da Engenharia e Tecnologia Espacial, bem como em domínios correlatos. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Criada em 1965, com a missão específica de promover informação especializada para a comunidade técnico-científica do instituto, o acervo físico dispõe atualmente de mais de 73.000 volumes entre livros, teses, relatórios e mapas, e mais de 160.000 fascículos de periódicos. Possui ainda uma Mapoteca composta por aproximadamente 4.800 mapas, entre cartas topográficas, imagens de radar e de satélite, cartas náuticas, mapas políticos, hidrográficos, geológicos, Projeto RADAM, e outras publicações como atlas, enciclopédias, estudo de águas subterrâneas e aptidão agrícola dos solos. Possui acervo digital, oferecendo acesso a mais de 6000 documentos eletrônicos, como por exemplo: e-prints, livros on-line, teses e dissertações **Outras informações** **Biblioteca Digital** Atualmente o acervo oferece acesso a mais de 6000 documentos eletrônicos como, por exemplo: e-prints, livros on-line, teses e dissertações. Além disso foi desenvolvida uma nova interface de busca para a Biblioteca Digital, adaptada ao novo padrão do portal, com opções de busca rápida e busca combinada. Desde 2003, a Biblioteca Digital tornou-se um provedor de	

	dados, conforme o padrão internacional OAI-PMH ("Open Archives Initiative Protocol for Metadata Harvesting"), fazendo com que os trabalhos armazenados no acervo estejam acessíveis a partir dos provedores de serviços OAI espalhados pelo mundo. Assim, os trabalhos do INPE constam, por exemplo, na base de dados da Universidade de Michigan - OAIster (com mais de 9 milhões de referências) e na base de dados da Biblioteca Digital de Teses e Dissertações (BDTD) do IBICT. A Biblioteca Digital também tem fornecido ao INPE os Indicadores da sua produção técnico-científica para os anos de 2006, 2005, 2004, 2003 e 2002.	
Equipe:	Marciana Leite Ribeiro – Chefe de Serviço de Informação e Documentação Técnicos bibliotecários, auxiliares e analistas - 10	

Nome:	**Instituto Nacional do Semi-Árido - INSA**	
Biblioteca:		
Endereço:	Av. Floriano Peixoto, nº 715 - Centro / Campina Grande – PB - CEP 58.100-001 / Fone (83) 2101-6400	
URL:	http://www.insa.gov.br/	
Descrição:	**Missão** Promover a realização de estudos e pesquisas científicas e tecnológicas sobre a região do semi-árido brasileiro, em estreita cooperação com outras entidades locais, bem como sua divulgação e difusão, visando à integração dos pólos socioeconômicos e ecossistemas estratégicos e o desenvolvimento sustentável da região. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Não tem unidade de documentação. **Publicações on-line** Possui publicações para acesso on-line.	

Nome:	**Instituto Nacional de Tecnologia - INT**	
Biblioteca:	Divisão de Informação e Prospecção Tecnológica	
Endereço:	Av. Venezuela, 82 -Saúde, 8º andar 20081.310 – Rio de Janeiro - RJ	
URL:	http://www.int.gov.br/sol_tec_info_main.html#1	
Descrição:	**Missão** Desenvolver e transferir tecnologias e executar serviços técnicos para o desenvolvimento sustentável do País, norteado pelo avanço do conhecimento e em consonância com as políticas e estratégias nacionais de Ciência, Tecnologia e Inovação. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** A Biblioteca do INT tem por objetivo oferecer produtos e	

	serviços que atendam às expectativas do corpo funcional da Instituição e da comunidade de C&T em geral, além de atender empresas, empreendedores, pesquisadores e estudantes. O acervo é composto de cerca de 20.000 livros, periódicos, teses, cadastros, manuais, guias e folhetos, além de aproximadamente 12.000 normas técnicas, nas áreas de química industrial, orgânica, inorgânica, tecnologia de materiais metálicos, cerâmicos e poliméricos, corrosão e proteção, desenho industrial, gestão da produção, gestão do conhecimento e informação tecnológica, além de obras de referência em engenharia, ciências básicas e administração. **Outras informações** Disponibiliza, em versão online, a Biblioteca Virtual do Gás Natural, desenvolvida no âmbito do Núcleo de Tecnologias de Uso do Gás Natural. A BVGN reúne, de forma estruturada, informações atualizadas sobre sites no Brasil e no mundo relacionado ao tema. **Publicações on-line** Não possui publicações para acesso on-line.	
Equipe:	Vera Lúcia Maia Lellis - Divisão de Informação e Prospecção Tecnológica	

Nome:	**Laboratório Nacional de Astrofísica - LNA**	
Biblioteca:	Biblioteca Sede	
Endereço:	Rua Estados Unidos,154 - Bairro das Nações 37500.000 - Itajubá - MG	
URL:	http://www.lna.br	
Descrição:	**Missão** Planejar, desenvolver, prover, operar e coordenar os meios e a infra-estrutura para fomentar, de forma cooperada, a Astronomia Observacional brasileira. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Possui biblioteca especializada na área de astrofísica e áreas afins com aproximadamente 1800 livros, proceedings de simpósios e colóquios. Destaca-se a coleção de periódicos em astronomia e astrofísica, que representa uma das mais completas no país. Possuiu ainda, periódicos na área de ciências computacionais, óptica, eletro-eletrônica, difusão científica, tanto quanto preprints e boletins de outros observatórios e associações de amadores. Não possui base de dados online **Publicações on-line** Não possui publicações para acesso on-line	
Equipe:	1 Técnico auxiliar	

Nome:	**Laboratório Nacional de Computação Científica - LNCC**
Biblioteca:	Serviço de Biblioteca - SBI
Endereço:	Av. Getúlio Vargas, 333 – Quitandinha 25651.070 - Petrópolis - RJ
URL:	http://www.lncc.br
Descrição:	**Missão** Realizar pesquisa e desenvolvimento em computação científica, em especial a criação e aplicação de modelos e métodos matemáticos e computacionais na solução de problemas científicos e tecnológicos; desenvolver e gerenciar ambiente computacional de alto desempenho que atenda às necessidades do País; formar recursos humanos, promovendo transferência de tecnologia e inovação. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Com um acervo especializado nas áreas de Matemática Aplicada e Computacional, Mecânica dos Sólidos e Fluidos, Teoria dos Sistemas e Controle, Métodos e Análise Numérica em Ciências e Engenharia, a Área de Biblioteca e Documentação tem como objetivo fornecer o apoio bibliográfico necessário ao desenvolvimento das atividades desenvolvidas no LNCC bem como pela disseminação deste acervo para a comunidade técnico-científica. A Biblioteca do LNCC possui um acervo especializado nas áreas de interesse do Laboratório, tais como: Computação Científica, Matemática Aplicada, Mecânica Aplicada, Teoria dos Sistemas e Controle, Ciência da Computação, Estatística e Biomatemática. Conta com cerca de 10.000 livros. **Publicações on-line** Possui publicações para acesso on-line e participa da BDTD
Equipe:	Maria Cristina Albuquerque de Almeida – Chefe Total de servidores - 5

Nome:	**Laboratório Nacional de Luz Síncrotron / ABTLus - LNLS**
Biblioteca:	Departamento Científico
Endereço:	Rua Giuseppe Máximo Scolfaro, 10.000 - 13083.970 - Campinas - SP
URL:	http://www.lnls.br/
Descrição:	**Missão** Atuar como laboratório nacional, aberto, multiusuário, multi e interdisciplinar, capaz de criar e prover soluções integradas para problemas científicos e tecnológicos complexos nas áreas de materiais avançados, nanotecnologia, biotecnologia, aceleredores e tecnologias relacionadas e instrumentação científica. Organização Social vinculado ao Ministério da Ciência e Tecnologia **Unidade de documentação** Biblioteca do LNLS tem como objetivo dá suporte às atividades técnico-científicas, localizando textos, obras completas,

	periódicos e outros materiais documentais. As atividades da Biblioteca tiveram início em 1987, logo que o LNLS começou a ser implantado e sempre foi vinculada ao Departamento Científico. Além de atender ao quadro de funcionários do LNLS, a Biblioteca também atende pesquisadores de inúmeras outras instituições brasileiras e do exterior, em especial os que utilizam a infra-estrutura de pesquisa e desenvolvimento existente no LNLS. Mantém intercâmbio com outras bibliotecas e integra os sistemas COMUT e o British Library (BL) que permite solicitar e receber cópias de artigos de periódicos, teses e capítulos de livros. **LNCC**: foi à primeira instituição a conectar o país a uma rede internacional acadêmica de computadores. **Publicações on-line** Possui publicações para acesso on-line	
Equipe:	Margarida M. S. Abreu de Lima	

Nome:	**Museu de Astronomia e Ciências Afins - MAST**	
Biblioteca:	Serviço de Biblioteca	
Endereço:	Rua General Bruce, 586, São Cristóvão 20921.030 - Rio de Janeiro - RJ	
URL:	http://www.mast.br/	
Descrição:	**Missão** Ampliar o acesso da sociedade ao conhecimento científico e tecnológico por meio da pesquisa, preservação de acervos e divulgação da História da Ciência e da Tecnologia no Brasil. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** Biblioteca especializada em história da ciência, educação em ciências e divulgação científica. Possui tanto obras do interesse de pesquisadores, como direcionadas para estudantes dos níveis fundamentais e médios, e também para o público em geral. Seu acervo compõe-se de aproximadamente 7000 volumes, incluindo clássicos da ciência, livros, teses de pós-graduação e um amplo sistema de referência sobre fontes de consulta no país e no exterior. O acervo se encontra totalmente informatizado. O acervo de periódicos é um dos mais completos do país nas áreas de história da ciência e divulgação científica, além de possuir assinaturas nas áreas de museologia, documentação e ensino em ciências. A coleção também está disponível para consulta no Catálogo Coletivo Nacional, e a biblioteca do MAST participa do programa de empréstimo entre bibliotecas. A biblioteca do MAST publica mensalmente os Sumários Correntes, que reúne os índices das revistas científicas recebidas durante o mês. **Publicações on-line** Não possui publicações para acesso on-line	
Equipe:	Lúcia Alves da Silva	

	1 Bibliotecária 1 Arquivista 1 Técnico auxiliar	

Nome:	**Museu Paraense Emílio Goeldi - MPEG**	
Biblioteca:	Biblioteca Domingos Soares Ferreira Penna	
Endereço:	Av. Perimetral, 1901 - Terra Firme CEP: 66077-530 - Belém - PA	
URL:	http://www.museu-goeldi.br/	
Descrição:	**Missão** Realizar pesquisas, promover a inovação científica, formar recursos humanos, conservar acervos e comunicar conhecimentos nas áreas de ciências naturais e humanas relacionadas à Amazônia. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** O acervo da Biblioteca Domingos Soares Ferreira Penna é especializado em antropologia, arqueologia, botânica, ciências da terra, ecologia, lingüística, zoologia e em assuntos amazônicos, reunindo documentos adquiridos ao longo dos anos, através de compra, doação e permuta. Seu acervo é composto por livros, periódicos, folhetos, separatas, mapas, CD's, fotografias, filmes, fitas e microfilmes. A Coleção Especial, com edições desde o século XVI, apresenta um valioso acervo de livros raros, antigos e valiosos, de grande importância para a Amazônia. **Publicações on-line** Possui publicações para acesso on-line	
Equipe:	Doralice Romeiro - Coordenação de Informação e Documentação Fátima Guedes - Bibliotecária	

Nome:	**Nuclebrás Equipamentos Pesados - NUCLEP**	
Biblioteca:		
Endereço:	Rod. Rio Santos, Km 18 – BR 101 Itaguaí - RJ	
URL:	http://www.nuclep.br	
Descrição:	**Missão** Fundada em 1975 Empresa Estatal vinculada ao Ministério da Ciência e Tecnologia **Unidade de documentação** Não tem unidade de documentação. **Publicações on-line** Não possui publicações para acesso on-line	

Nome:	**Observatório Nacional - ON**
Biblioteca:	Divisão de Documentação - DID
Endereço:	Rua General José Cristino, 77 – São Cristóvão 20921.400 – Rio de Janeiro - RJ
URL:	http://www.on.br/
Descrição:	**Missão** Realizar pesquisa e desenvolvimento em Astronomia, Geofísica e Metrologia em Tempo e Freqüência, formar pesquisadores em seus cursos de pós-graduação, capacitar profissionais, coordenar projetos e atividades nacionais nestas áreas e gerar, manter e disseminar a Hora Legal Brasileira. Unidade de Pesquisa do Ministério da Ciência e Tecnologia **Unidade de documentação** É considerada uma das melhores bibliotecas especializadas do país, com um acervo de valor inestimável nas áreas de Astronomia, Geofísica, Metrologia e ciências correlatas. A seção de obras raras inclui livros, fotografias e negativos em vidro. Grande parte desse material remonta ao século XIX, encontrando-se, no entanto, livros do século anterior. Desde 2004 disponibiliza na página oficial do ON a Biblioteca Digital de Obras Raras e Especiais do Observatório Nacional, que consiste na reprodução em meio eletrônico do acervo raro, visando a sua preservação e disseminação. **Publicações on-line** Possui publicações para acesso on-line
Equipe:	1 Bibliotecária

Nome:	**Associação Rede Nacional de Ensino e Pesquisa - RNP**
Biblioteca:	Diretoria de Gestão da Informação - DGI
Endereço:	Rua Lauro Muller, 116 sala 3.902 CEP 22290-160 Rio de Janeiro - RJ
URL:	http://www.rnp.br/
Descrição:	**Missão** Promover o uso inovador de redes avançadas no Brasil. Organização Social vinculada ao Ministério da Ciência e Tecnologia. **Unidade de documentação** Não tem unidade de documentação. Conta com um pequeno acervo dentro da Biblioteca da Embrapa em Campinas **Publicações on-line** Possui publicações para acesso on-line
Equipe:	1 Cientista da Informação

APÊNDICE E – Produção bibliográfica de servidores do Ministério da Ciência e Tecnologia sobre preservação digital

Banon, G. J. F. Hiperdocumentos versus URLib. 2002. Disponível em: <dpi.inpe.br/banon/2002/10.10.08.39>. Acesso em: 25 fev. 2004.

_____. Uniform repositories for a digital library (URLib). 1998. Disponível em: <dpi.inpe.br/banon/1997/08.14.12.32>. Acesso em: 25 fev. 2004.

_____; Banon, L. C. Creating persistent citation links. 2001. Disponível em: <iconet.com.br/banon/2001/07.27.00.56>. Acesso em: 25 fev. 2004.

_____; RIBEIRO, M. L.; BANON, L. C. Contribuição ao livre acesso da literatura científica em ciência espacial: implementação da política de auto-arquivamento na biblioteca digital do INPE. In: SEMINÁRIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS BRASIL, 1., 2007, São Paulo. **Proceedings**... São Paulo: INPE, 2007.

_____; _____; _____. Preservação digital da memória técnico-científica do INPE. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2., 2004, Campinas. **Anais eletrônicos**... Campinas: INPE, 2004. Disponível em: <http://libdigi.unicamp.br/document/?view=8264>. Acesso em: 19 set. 2005.

_____; _____; SILVA, M. D. **Preservação, enriquecimento e disseminação das Memórias Técnico-Científica e Documental do INPE**. São José dos Campos: INPE, 2003. Disponível em: <http://urlib.net/sid.inpe.br/marciana/2003/10.21.16.25>. Acesso em: 15 fev. 2004.

BARBEDO, S. A. D.; RIBEIRO, M. L. Estudo de padronização de metadados para preservação da memória documental da biblioteca digital do INPE. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 3., 2005, São Paulo. **Anais**... São Paulo: INPE, 2005. Disponível em: <http://ePrint.sid.inpe.br:80/rep-/sid.inpe.br/ePrint@80/2005/12.02.12.25>. Acesso em: 16 fev. 2006.

BERGAMINI, E. W. et al. CCSDS Standards: a reference model for an Open Archival Information System (OAIS). In: SIMPÓSIO SPACEOPS, 98., 1998, Tokio. **Anais**... Tókio: [s.n.], 1998. Disponível em: <http://iris.sid.inpe.br:1905/col/sid.inpe.br/banon/2001/04.03.15.36.19/doc/mirror.cgi:IISarau na Biblioteca>. Acesso em: 12 jan. 2006.

BOERES, S. A. A.; MÁRDERO ARELLANO, M. A. Políticas e estratégias de preservação de documentos digitais. In: CINFORM, 4., 2005, Salvador, BA.

Proceedings... Salvador, BA: Diálogo Científico, 2005. Disponível em: <http://dici.ibict.br/archive/00000263>. Acesso em: 16 fev. 2007.

MÁRDERO ARELLANO, M. Á. **As coleções de obras raras na biblioteca digital**. 93 f. 1998. Dissertação (Mestrado em Ciência da Informação)- Departamento de Ciência da Informação e Documentação, Universidade de Brasília. Brasília: [s.n], 1998.

_____. Digital preservation of scientific information in Brazil: an initial approach of existing models. In: INTERNATIONAL CONFERENCE ON ELETRONIC PUBLISHING, 8., 2004, Brasília. **Proceedings**… Brasília: Universidade de Brasília, 2004.

_____ Preservação de documentos digitais. **Ciência da Informação**, v. 33, n. 2, p. 15-27, 2004.

_____; ANDRADE, R. S. Preservação digital e os profissionais da informação. **DataGramaZero**, v. 7, n. 5, out. 2006.

_____. Preservação digital: uma preocupação dos editores de revistas científicas eletrônicas no Brasil. In: ENCONTRO NACIONAL DE EDITORES CIENTÍFICOS, 9., 2007, Ouro Preto, MG. **Proceedings**... Ouro Preto, MG: [s. n], 2007.

RIBEIRO, M. L. **Aspectos da memória científica do Instituto Nacional de Pesquisas Espaciais (INPE)**. São José dos Campos: INPE, 2007. Disponível em: <http://urlib.net/sid.inpe.br/mtc-m17@80/2007/07.10.18.17>. Acesso em: 21 out. 2008.

_____. Reflexões sobre o resguardo da memória científica do INPE. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 14., 2006, Salvador, BA. **Anais**... Salvador, BA: INPE, 2006. Disponível em: <http://urlib.net/sid.inpe.br/mtc-m16@80/2006/10.31.16.16>. Acesso em: 21 out. 2008.

SAYÃO, L. F. O modelo de referência ISO/OAIS: Open Archival Information System. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2., 2004, Campinas. **Anais**... Campinas: Unicamp, 2004.

_____. Padrões para bibliotecas digitais abertas e interoperáveis. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 4, n.1, p. 101-130, jan./jun. 2008.

_____. Preservação digital no contexto das bibliotecas digitais: uma breve introdução. In. MARCONDES, Carlos H. et al. (Org.). **Bibliotecas digitais**: saberes e práticas. Salvador: EDUFBA; Brasília: IBICT, 2005. p. 115-149.

_____. **Preservação Digital**: uma brevíssima introdução. Rio de Janeiro: [s. n.], 2003. Disponível em: <http://www.arquivonacional.gov.br/conarq/cam_tec_doc_ele/preservacao/preservacao.asp>. Acesso em: 03 dez. 2003.

TOLMASQUIM, A. T. Preservação de objetos tridimensionais para a história da ciência e tecnologia. In: SEMANA NACIONAL DO LIVRO E DA BIBLIOTECA, 2., 2005, São José dos Campos. **Atas**... São José dos Campos: INPE, 2005.

APÊNDICE F - Glossário de termos técnicos

Acessibilidade de longo prazo: acessibilidade do sistema de preservação digital que permite medir as facilidades de uso que ele proporciona aos usuários, a dimensão das tarefas realizadas e o grau de conhecimento que eles têm das funções do sistema; depende da sua capacidade de ser interoperável; usa de ferramentas de *softwares* livres, possibilitando a alteração de qualquer parte do sistema sem nenhuma conseqüência nas suas funcionalidades. Como um critério pode ser medido por meio de indicadores de desempenho que proporcionam os dados sobre a eficiência do sistema no atendimento de usuários e suas expectativas de tempo e tipo de resposta.

Acesso Livre: disponibilização livre e pública na Internet, de forma a permitir a qualquer usuário a leitura, *download*, cópia, impressão, distribuição, busca ou *link* com o conteúdo completo de artigos, bem como a indexação ou o uso para qualquer outro propósito legal.

Análise Etnográfica de Conteúdo: técnica de análise qualitativa de dados, usada para o estudo da origem e uso de conceitos formulados nas mídias. Diferentemente da análise de conteúdo tradicional, a AEC interage com as informações sem ficar restrita a amostras; tratar o processo de amostragem como uma comparação constante, cujas descobertas são usadas para delinear categorias específicas e narrativas descritivas.

Arqueologia digital: uso de material obsoleto para pesquisa da história dos inícios da arquitetura computacional, aplicativos periféricos, sistemas operacionais, linguagens de programação, ferramentas e sistemas de gestão eletrônica de documentos.

Arquivos Abertos (*Open archives*): arquivos eletrônicos conectados por meio do protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), permitindo a interoperabilidade entre bibliotecas e repositórios digitais.

Atributo: característica inerente ou qualidade associada a um elemento já mencionado.

Atualização (*refreshment*): migração de dados digitais que procura substituir uma mídia com uma cópia que seja suficientemente exata, para que a infra-estrutura de *hardware* e *software* continue a funcionar como antes.

Auditoria: permite que os desenvolvedores dos repositórios analisem e respondam de forma sistemática às carências e acertos na criação dos sistemas.

Autenticidade: comprovação de autoria do documento por meio de mecanismos de verificação como o *layout*, tipologia de fontes, vocabulários controlados da época e assinatura digital.

Autodepósito: que possibilita ao próprio autor a publicação de seu trabalho científico diretamente na Rede, sem necessidade de intermediários.

Biblioteca digital: biblioteca com textos completos em formatos digitais, armazenados e disponíveis via Internet.

***Bit (binary bit)*:** dígito binário, menor unidade de informação armazenada em um computador. Um *bit* tem um valor único, 0 ou 1.

***Bitstream*:** é a camada mais elementar para dotar de significado um objeto digital, emprega a leitura bit por bit em vez do método convencional de 16 bits por vez.

***Bundles*:** agrupamento ou pacotes de *bitstreams*. A possibilidade de um documento incluir vários *bundles* significa que ele pode pertencer a uma ou mais coleções e uma coleção a uma ou mais comunidades. Essa função dos *bundles* pode ser compreendida na seqüência de duas funcionalidades já disponíveis na versão 1.3 do DSpace.

***Byte*:** grupo de dígitos binários que são armazenados e operados como se fossem uma unidade.

Certificação: envolve mais do que a aplicação de critérios resultantes de normas e requisitos básicos. Ela deve prover ferramentas para o planejamento da auto-avaliação e formas de auditorias internas e externas. O contexto de ambas é geopolítico, mas o processo não varia significativamente.

Confiabilidade: garantir o armazenamento e acesso da informação por longo prazo

Cópia de segurança (*backup*): cópia de dados de um dispositivo de armazenamento a outro para que possam ser restaurados em caso da perda dos dados originais.

Critérios de preservação digital: conjunto de atividades que determinam a manutenção de coleções digitais; a maioria das propostas de critérios que vem surgindo menciona a influência do modelo de referência OAIS/SAAI, elas definem as funções dos repositórios, os processos e procedimentos, a comunidade alvo a usabilidade da informação e as tecnologias e a infra-estrutura técnica. Usados em procedimentos de auditoria e modelos de certificação.

Digitalização: transformação de informação analógica para formato digital, por meio de dispositivo adequado e seguindo requisito de qualidade no seu processo de produtividade.

Documento digital: documento codificado em linguagem binária digital ou convertido para formato digital, legível por máquinas, independente do suporte físico e de sua mensagem original, sendo que a digitalização pode chegar a produzir mudanças no se registro.

Documento eletrônico: um registro que foi criado (elaborado ou recebido) de forma eletrônica.

DOI® (*Digital Object Identifier*): sistema de identificador de objetos digitais on-line para gerenciar a propriedade intelectual e o uso comercial dos objetos materiais digitais.

Dublin Core: esquema de descrição de metadados para descoberta de recursos *on-line*.

Emulação: criação de um novo *software* que imita o funcionamento do antigo *hardware* e/ou *software*, com o objetivo de reproduzir seu comportamento.

Encapsulamento: preservação da informação necessária para o desenvolvimento de conversores, emuladores e visualizadores de objetos digitais.

Espelhamento ou redundância (*Mirroring or Redundancy*): manter uma cópia exata de um arquivo em diversos lugares; muito usado na cópia de segurança de grandes bases de dados.

Estratégia de preservação digital: tipo de abordagem que permite o acesso contínuo à informação em formato digital. Esquemas técnicos e conceituais que permitem identificar as partes integrantes do processo de preservação digital, assim como sua padronização em modelos de referência.

Formato: estrutura do arquivo que define o modo como é armazenado ou apresentado na tela ou impressão; organização seqüencial de dados em termos de seus componentes, também chamado de extensão. Formatos abertos de arquivo são aqueles em que os usuários têm acesso aos detalhes técnicos e ao significado de sua estrutura de *bits*.

***Handle System*®:** um exemplo de identificador persistente, desenvolvido pela *Corporation for National Research Initiatives*.

***Harvesting*:** busca automática de endereços *web*, com o resgate das referências para outros documentos, *links* e informações relevantes.

Identificador: um caractere ou conjunto de caracteres designados para identificar inequivocamente um documento. Alguns exemplos de identificadores são o URN (*Uniform Resource Name*), o DOI®, o ISBM (*International Standard Book Number*) e o SICI (*Serial Item and Contribution Identifier*).

Identificador persistente: um único identificador permanentemente associado a um objeto digital, quando gerenciado, sempre proverá acesso permanente a um objeto independentemente de qualquer mudança de local.

Informação digital: dados organizados e com significados, codificados em dígitos binários, *bits*, produzidos e armazenados em computador.

Interoperabilidade: interação de documentos digitais entre diferentes sistemas tecnológicos; configuração de todos os arquivos, padronizando-os de forma global.

Metadado: informações estruturadas e codificadas que descrevem e permitem gerenciar, compreender, preservar e acessar os documentos digitais ao longo do tempo.

Metadados administrativos: tipo de metadados que permite o gerenciamento de uma coleção digital.

Metadados de preservação: atributos especiais de dados ou documentos, geralmente descritivos; são aqueles relacionados com o conteúdo do documento, seu contexto e estrutura no momento da sua criação, assim como das mudanças acontecidas em todo seu ciclo de vida.

Migração: conjunto de atividades para copiar, converter ou transferir a informação digital de uma tecnologia antiga ou obsoleta para uma tecnologia nova; transferência periódica de dados de uma geração tecnológica para outra.

Modelo de Referência: estrutura para o entendimento de relações significativas entre entidades de um ambiente e para o desenvolvimento de padrões consistente ou especificações que lhe dão suporte. Um modelo de referência esta baseado em um número pequeno de conceitos unificados.

OAIS (*Open Archival Information System)*: esquema conceitual que disciplina e orienta um sistema para a preservação e manutenção de acesso à informação digital por longo prazo; contém o termo *archival* para distingüi-lo de outros usos do termo “arquivo”. Arquivo no OAIS consiste em uma organização de pessoas e sistemas que aceitam a responsabilidade de preservar informação e torna-la disponível.

Objeto digital: tipos de arquivos encontrados em meio digital, compostos de conjuntos de seqüências de *bits* sobre conteúdos informacionais. metadados e identificadores.

Pacote de Arquivamento de Informação (PAI): informação de conteúdo e sua descrição de preservação necessária para preservar-la dentro de um pacote de informação por longo prazo.

Pacote de Disseminação de Informação (PDI): pacote de informação que contém partes de um PAI que são distribuídas a um Cliente/Usuário quando solicitado.

Pacote de Submissão de Informação (PSI): pacote de informação identificado pelo produtor no acordo de submissão na estrutura conceitual do OAIS.

Preservação digital de longo prazo: o ato de manter informação, de forma correta e independente por longo prazo. Informação armazenada independentemente que possui documentação suficiente para permitir que

seja interpretada e usada por uma comunidade sem necessidade de recursos especiais obsoletos. Requer procedimentos específicos e técnicas apropriadas para cada tipo de formato e mídia. Com ela pretende-se garantir a inalterabilidade dos registros digitais

***Preservation Master File*:** o objeto derivado do original que será preservado, gerenciado e disseminado por longo prazo.

Protocolo: padrões que controlam a transferência de informação entre computadores. Desenvolvidos para melhorar a confiabilidade e a velocidade de dados transferidos.

Repositório digital: sistema de informação *on-line* encarregado de gerar e armazenar objetos digitais para seus subseqüentes acessos e recuperações.

Repositório digital confiável: resultado da associação de repositórios institucionais serviços de preservação e arquivos por área de conhecimento; tem como missão prover acesso de longo prazo confiável a recursos digitais gerenciados para suas comunidades alvo, hoje e no futuro.

Repositório Institucional: repositório digital que incluem não só artigos submetidos ao processo tradicional de revisão dos pares (*pos-print*), como também versões eletrônicas preliminares de documentos científicos (*pré-prints*), teses, manuais, materiais de ensino ou qualquer outro tipo de material que os autores ou sua instituição desejem que se torne disponível para o público. Eles devem estar respaldados por política de gerenciamento que os reconheçam como vitais para a instituição.

SGML (*Standard Generalized Markup Language*): Linguagem de Marcação Generalizada Padrão que é a base para a criação de todas as linguagens de marcação, definido pela norma ISO 8879:1986.

Software: programa de computador. São instruções que o computador é capaz de entender e executar.

***Software* livre:** programa de computador em que se tem liberdade de executar, copiar, distribuir, modificar e aperfeiçoar, de acordo com interesses pessoais ou comerciais. Apresenta um código-fonte aberto.

URL (*Uniform Resource Locator*): padronização da localização ou dos detalhes de endereçamento dos recursos da Internet.

XML (*Extensible Markup Language*): linguagem para descrição de documentos eletrônicos; estrutura sintática para descrição de dados entre aplicações de informática, em formato de texto; é um formato padrão para publicação, armazenamento e transferência de documentos por via eletrônica, independente dos fabricantes de *software,* facilmente interpretável por múltiplas aplicações e, em certa forma, auto-explicativo.