

DIGITAL PRESERVATION OF SCIENTIFIC INFORMATION IN BRAZIL: AN INITIAL APPROACH OF EXISTING MODELS

MIGUEL ÁNGEL MÁRDERO ARELLANO

Department of Information Science at the University of Brasilia – CID-UNB
C. Postal 04561, 70919-970 Brasilia-DF, Brazil
miguel@ibict.br

The problem addressed in this study is to investigate how digital science and technical information produced and used by brazilian government research institutes has been preserved and what kind of digital preservation strategies can be proposed. It presents the results of an exploratory study from a Doctoral dissertation research being developed at the Department of Information Science at the University of Brasilia, Brazil.

Keywords: digital preservation; digital preservation strategies; digital electronic archiving; information science; scientific information.

This proposal aims to present the actual conditions relating to the preservation of science and technical information from government research institutions¹ maintained by the Brazilian Science and Technology Ministry. The unit of analysis was the 22 research institutions from the government sector. This group of participants was selected because they correspond to the government scientific and technical community that is responsible for the archiving and access to the main official scientific and technical projects. A questionnaire was mailed and the assessment of the conditions was constrained both to those centers and the Ministry's documentation. Also, part of the basic unit included in this study were 21 digital preservation systems² analysed by CENDI (U.S. Federal Information Managers Group) and ICSTI (International Council for Scientific and Technical Information) in their February 2004 report.³ The preservation strategies⁴ used in formulating those representative systems were considered in the analyses. The strategies that were detected more frequently will be studied in a follow up part of the research in order to gather more detailed information. This information will be used in the formulation of a group of digital preservation models. It is expected that these models might be usefull in the digital preservation of science and technical information from brazilian government institutions.

The conceptual definition of digital preservation presented by Hedstrom (1997/1998)⁵ and The Task Force on Archiving of Digital Information (1996)⁶, as well as the definition of digital preservation strategies introduced by Bullock (1999)⁷ were used in constraining the aspects that should be operationalized in future studies.

Discussions are underway as to how to implement digital preservation strategies in several parts of the world⁸. Operational digital preservation systems, specifically in science and technology, are being developed by a variety of stakeholders and partnertships. The needed of effective and sustainable preservation guidelines for government scientific and technical information should be addressed by government agencies. These agencies should specially determine how to develop technical and social solutions for preserving scientific and technical data.

Chart 1 shows a summary of the activities related to digital preservation detected on 22 Brazilian government scientific and technological institutions.

CHART 1 – SUMMARY CHART OF DIGITAL PRESERVATION ACTIVITIES DETECTED ON THE 22 BRAZILIAN GOVERNMENT SCIENTIFIC AND TECHNOLOGICAL INSTITUTIONS.

ACTIVITY INSTITUTION	Deals with digital material	Digitized from paper	Implements digital preservation strategy	Partnerships on digital preservation
IBICT	Yes, for electronic journal, theses and dissertations	Yes		Yes, Open Archive Initiative
INPA	Yes, for electronic journal			
MPEG				
INPE	Yes, for many document types			Yes, Open Archive Initiative
CRPF	Yes, for many document types	Yes		
CINEN	Yes, for electronic journal and conference information	Yes		
LNCC				
LNLS	Yes, for electronic journal and technical reports			
MAST				
ON				
CENPRA	Yes, for electronic journal, theses and dissertations	Yes		
FINEP				
CNPq				
MAMI				
LNA				
RNP	Yes, for technical reports			
AEB				
CETEM				
INT				
INPA				
CGEE				
XINGÓ				

The data revealed both that science and technical information of Brazilian government institutions lacks standards-related digital preservation activities and the scientific data is being stored in several information center's databases that are proprietary in nature, without a strategic plan that ensure the longevity of the digital information. Current practices are still modeled by traditional preservation activities for paper material. Issues related to digital preservation are not current focus for most of those institutions. The document type that is at the front of digital preservation implementations is electronic journal. Digital technical reports and other gray literature on institutional archives are starting to be collected and also with not directly addressing preservation and long-term access.

Chart 2 shows the preservation strategies adopted by operational systems selected by CENDI/ICSTI. The 21 systems or projects selected on the CENDI/ICSTI report (2004) manage a wide range of scientific resources and organizations. The table provides the group of digital preservation strategies that were detected on those projects interested in supporting digital preservation in science. Most data archives have archiving policies and standards related to activities underway in metadata creation. Migration is the most used preservation strategy in those systems. Partnerships is still important in the digital preservation community as always have been. Emulation strategy was not mentioned as a main approach to archiving and transformation for those

organizations.

CHART 2 – PRESERVATION STRATEGIES ADOPTED BY 21 OPERATIONAL SYSTEMS SELECTED BY CENDI/ICSTI (HODGE AND FRANGAKIS, 2004)

STRATEGY SYSTEM	Archiving Policy	Migration	Open Standards	Metadata Creation	Partnerships	Institutional Repository
AIP	X	X		X		
ALA		X		X	X	X
DIAS		X	X		X	
DVA	X	X		X		X
Dspace	X	X	X	X	X	X
END					X	
EROS	X	X		X		
FEDORA™		X	X	X	X	X
HUR	X	X		X		
ISTOR	X	X		X	X	
LSDA		X		X		
LOCKSS			X		X	
NASA		X		X		X
NDM	X	X				
OCLC						
PANDORA	X	X	X	X	X	
NLM	X	X				X
PLbMed	X	X		X		
ALEXA	X			X	X	
GPO		X		X	X	
VERS	X	X		X		X
Total	12	17	5	15	10	7

The results of this study indicates that brazilian government agencies that deal with scientific and technical information are not ready in providing permanent access and adequate rendering of the digital objects that they use and produce. Models to be proposed specifically in science and technology might be based on the existence of most frequently found digital preservation strategies used on the creation of deposit systems for digital objects. They might also consider the standardization of the storage and long-term preservation processes using digital information repositories. These models should be taken by brazilian agencies as the basis for establishing organizations specialized in maintaining long-term access to digital documents.

NOTES AND REFERENCES

¹ IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia

IMPA – Associação Instituto Nacional de Matemática Pura e Aplicada

MPEG – Museu Paraense Emílio Goeldi

INPE – Instituto Nacional de Pesquisas Espaciais

CBPF – Centro Brasileiro de Pesquisas Físicas

CNEN – Comissão Nacional de Energia Nuclear

LNCC – Laboratório Nacional de Computação Científica

LNLS – Laboratório Nacional de Luz Síncrotron / ABTLus

MAST – Museu de Astronomia e Ciências Afins ON – Observatório Nacional

CENPRA - Centro de Pesquisas Renato Archer FINEP – Financiadora de Estudos e Projetos

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

MAMI – Instituto de Desenvolvimento Sustentável Mamirauá

LNA – Laboratório Nacional de Astrofísica

RNP – Associação Rede Nacional de Ensino e Pesquisa

AEB – Agência Espacial Brasileira

CETEM – Centro de Tecnologia Mineral

INT – Instituto Nacional de Tecnologia

INPA - Instituto Nacional de Pesquisas da Amazônia

CGEE – Centro de Gestão e Estudos Estratégicos

XINGÓ - Instituto de desenvolvimento Científico e Tecnológico Xingó

² AIP – American Institute of Physics

AIA – Aerospace Industries Association/Boeing Co.

DIAS – Digital Information Archiving Systems Dutch National Library

DiVA – Electronic Publishing Centre, Uppsala University Library

Dspace - at MIT

ESD - Elsevier Science Direct – also part of the Dutch National Library

EROS – Earth Resources Observation Systems Data Center

FedoraTM – Flexible Extensible Digital Object Repository Architecture, Cornell

University and the University of Virginia Library

IUCR – International Union of Crystallography

JSTOR – Electronic-Archiving Initiative

LSDA – Life Science Data Archive

LOCKSS – Lots of Copies Keep Stuff Safe

NASA Goddard Space Flight Center Library

NMM – National Motor Museum

OCLC – OCLC's Digital Archive

PANDORA – National Library of Australia

NLM – Profiles in Science, National Library of Medicine PubMed Central – National Library of Medicine

ALEXA - The Internet Archive GPO – U.S. Government Printing Office

VERS – Victorian Electronic Records Strategy – Australia

³ HODGE, Gail; FRANGAKIS, Evelyn. Digital Preservation and Permanent Access to Scientific Information: the State of the Practice. Final Report. February 2004. Sponsored by CENDI/ ICSTI. Revised March 2004.

⁴ Archiving Policy

Migration Strategy

Open Standards

Metadata Creation

Partnerships

Institutional Repository

⁵ HEDSTROM, Margaret. Digital preservation; a time bomb for digital libraries. *Computer and the Humanities*, v.31, n.3, 1997/1998, p.189-202. URL: <<http://www.uky.edu/~kiernan/DL/hedstrom.html>>

⁶ TASK FORCE ON THE ARCHIVING OF DIGITAL INFORMATION. *Preserving digital information*; report of the Task Force on Archiving of Digital Information; commissioned by the Commission on Preservation - CPA and Access and the Research Libraries Group - RLG. Washington, D.C.: Commission on Preservation and Access, 1996. URL: <<http://www.rlg.org/ArchTF>>

⁷ BULLOCK, Alison. *Preservation of digital information*; issues and current status. April 22, 1999. Last updated on February 27, 2001. URL: <<http://www.nlc-bnc.ca/publications/1/p1-259-e.html>>

⁸ BEAGRIE, Neil. National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity. CLIR, Council on Library and Information Resources. 2003. URL:<<http://www.clir.org/pubs/reports/pub116/pub116.pdf>>