The Long and Short of IT: The International Development Research Centre as a Case Study for a Long-term Digital Preservation Strategy

by

Lori Podolsky Nordland

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS
(Joint Masters Program)

Department of History (Archival Studies)

University of Manitoba
and
University of Winnipeg

Winnipeg

The Long and Short of IT: The International Development Research Centre as a Case Study for a Long-term Digital Preservation Strategy

by

Lori Podolsky Nordland

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF ARTS
(Joint Masters Program)

Department of History (Archival Studies)

University of Manitoba
and
University of Winnipeg

Winnipeg

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

The Long and Short of IT: The International Development Research Centre as a Case Study for a Long-term Digital Preservation Strategy

BY

Lori Podolsky Nordland

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

MASTER OF ARTS

Abstract

This thesis is a contribution to the study of the challenges facing archivists and records managers working on the long-term management and preservation of digital records. This thesis discusses the International Development Research Centre (IDRC), a Canadian government Crown agency, as a case study. In 2004 IDRC's Research Information Management Service (RIMS) Division was given the responsibility for developing a digital preservation program for the centre's final reports and related documentation. To facilitate this work, it hired a student intern to research recommendations for a digital preservation strategy. My research as the centre's intern led to the following recommendations for IDRC:

- Choose file formats that are ubiquitous, non-proprietary (when possible), viable, and lossless;
- Implement a strategy of conversion and migration of file formats and media as they become obsolete;
- Capture metadata to support the preservation of and access to digital objects; and
- Comply with the Open Archival Information System (OAIS) reference model.

Much academic study by archivists on digital preservation focuses on the concepts relating to digital records and records management. This thesis offers a practical institutional example of one effort to develop an actual archival program.

Acknowledgements

I would like to thank the International Development Research Centre for providing me with the opportunity to gain experience and knowledge in digital preservation as a centre intern in 2004, and for permission to use the research and final report as the foundation of my thesis. As well, I would like to acknowledge Marjorie Whalen, Diane Ceponkus, Manon Gendron, Kyle Brentnell, and Sam Malenfont for their support, encouragement, and patience in answering my questions and explaining to me the functions and structure of the RIMS Division and IDRC. Without your help, the research and writing of this thesis would not have been possible.

I would also like to thank Dr. Tom Nesmith, my advisor, and Dr. Terry Cook for their guidance and encouragement throughout my graduate studies. Your kind words and insightful discussions have been instrumental in my academic and professional career. To my committee members, Dr. Gerald Friesen, Dr. Len Kuffert, and Mr. Gerry Berkowski, thank you for your contributions to my thesis.

Special mention must be made to Kara Quann who first suggested that I meet with Tom and Terry, and encouraged me to apply to the Archival Studies program. I would also like to mention my classmates – the excellent eight – who have made this a wonderful journey of friendship.

I would also like to acknowledge the United Kingdom National Archives for granting me permission to include portions of their publication, *Digital Preservation Guidance Note 1: Selecting File Formats for long-term preservation,* in my thesis.

Finally, I must thank my parents and my sister for their support throughout my education. And, to my husband, Jon, and my son, Willem, who have remained by my side and travelled with me across the country (multiple times) during the last five years. It is to them that I owe the greatest gratitude and debt for without their sacrifices and unwavering support, none of this would have been possible.

I am grateful to all these people for their contributions to my thesis, archival studies, and career.

Dedicated to:

Jon and Willem
Mom, Dad and Lisa

Thank you for all your support, encouragement and love,
And, yes Dad, the thesis is coming along fine.

Table of Contents

List of Tables and Charts

List of Figures

Introduction

Between June 2004 and January 2005, I worked as a Centre Intern on a digital archiving project in the Research Information Management Service (RIMS) Division of the International Research Development Centre (IDRC), a Canadian Crown corporation. As a graduate student, I had an understanding of archival theory as well as archival work experience. My duties included writing a final report that outlines my recommendations for the long-term management and preservation of digital objects (or digital records) to be held indefinitely by IDRC itself, but does not deal with digital material that the centre will transfer to Library and Archives Canada. I was also fortunate in that the Director of the RIMS Division, Marjorie Whalen, both encouraged and granted permission for me to use the final report as the foundation for this Masters of Arts thesis in Archival Studies. This thesis presents the recommendations with more in-depth analysis than the report. The thesis is divided into four chapters, with each chapter incorporating parts of the report.

The first chapter outlines some of the more common digital preservation challenges facing archivists, and introduces the Open Archival Information System (OAIS) reference model. Beginning with an overview of digital preservation, including a brief literature review of digital preservation projects from the 1980s to the present, the chapter moves into a description of the OAIS reference model. This model has been gaining acceptance within the archival community as a blueprint or roadmap for developing a digital archives, which institutions are using as a benchmark for their preservation functions. The benefits of the OAIS include, but are not limited to, a common definition of terms (or ontology) to foster communication between disciplines (such as information sciences and technology, archives, and information management), an impartiality towards numerous digital preservation strategies (the model does not subscribe to one methodology over another), and the recommendations of the model were (and continue to be) developed in an open forum. Subsequently, OAIS is a recognized standard as per the International Organization of Standardization (ISO). The chapter ends with a description of a trusted digital repository that is a primary component of the OAIS reference model.

Chapter two focuses on the history and organizational structure of IDRC, and the functions of RIMS Division today as the project lead for the centre's digital preservation strategy. In addition, the chapter delineates a number of aspects of RIMS Division policy as they relate to digital preservation. These include appraisal, costs, disaster and emergency recovery plan, and the preservation of e-mail and websites. This chapter provides the organizational background in which RIMS Division sought to address its digital preservation issues through my internship.

Chapter two then outlines the IDRC stakeholders involved in digital preservation when I undertook my internship. The purpose of the internship was twofold. First, RIMS Division could dedicate resources for a short-term project to develop a policy and procedures framework for a digital preservation strategy. Second, the internship was intended to reflect the centre's goal of promoting knowledge through research. As a result, my internship focused on research into existing digital preservation challenges, and provided me with the opportunity to gain a deeper understanding of these issues. Chapters two and three contain my research findings and recommendations. These chapters form the bulk of my report to IDRC.

Chapter three moves from the organizational context of the digital preservation strategy outlined in chapter two to the technical recommendations. This chapter presents preservation strategies, file formats, technical architecture (or infrastructure), and metadata schemas in greater detail. The three most common strategies – emulation, encapsulation, and migration – are compared. Additionally, successful case studies in which these strategies have been implemented are included in this chapter. The section on *Recommended File Formats* is adopted from the National Archives (UK) *Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation,* and outlines the ten criteria for choosing archival formats for digital preservation. This section also expands upon the National Archives (UK) *Guidance Note* by including a short analysis and summary of the specifications of some of the more common file formats being used for the long-term management of digital objects. In the *Digital Preservation Management Architecture* section, recommendations for proper media and storage of digital objects are developed with a view to improve current practices in RIMS Division. The last part of the technical recommendations examines the role of metadata in the digital preservation strategy. Metadata schemas, such as Dublin Core Metadata Initiative, OAIS information attributes, Preservation Metadata: Implementation Strategies (PREMIS), among others, are presented in a manner similar to a literature review or environmental scan. The chapter concludes with a *Final Report Post-Evaluation* that reflects upon my knowledge and experience, and how this formed my recommendations to the RIMS Division.

In the concluding chapter, the importance of research and collaboration as it relates to the development of the recommendations for IDRC's digital preservation strategy is explored. This chapter draws upon Cornell University's *Digital Preservation Management Workshop* and Canada Institution of Scientific and Technical Information Digital Archives Research Framework consultation as RIMS Division moves forward in implementing a digital preservation strategy. The chapter and thesis concludes with a brief assessment regarding the collaboration of information management and information technology professionals who are working in partnership to address the challenges of the management and preservation of digital objects.

CHAPTER ONE

Digital Preservation Challenges

This chapter presents some of the more common challenges related to digital preservation, and briefly explains how these challenges cause concerns for archivists who are responsible for the preservation of digital records or objects. This includes a brief introduction to key projects and initiatives, coalitions, and research groups, which is followed by a more in-depth look at the Open Archival Information System (OAIS) reference model and the attributes of a trusted digital repository. The OAIS reference model is the recognized and official standard for archival institutions managing and preserving digital objects. The OAIS environment also provides the framework for the digital preservation strategy that was presented to the International Development Research Centre, and adheres to the organizational culture of IDRC, as outlined in chapter two.

Digital Preservation Overview

Recording media ara always changing. How we maintain that information is evolving even faster. Cave paintings and cuneiform writing on stone tablets have been around for thousands of years – and we can still view them. Paper records have been around for centuries, and can still be read, even though they deteriorate faster than their stone or clay counterparts. On the other hand, digital records have been in existence for a few decades, in which the lifespan of digital information may be as long (or as short) as eighteen months.[1] Unlike cave drawings, stone tablets, and paper records, digital information strongly relies on its technological environment to be accessible. In other words, we can only access a digital object in its perceptible form when the object is printed to paper, displayed on a monitor, or played on an MP3 player.[2] We cannot read its digital representation or the bits and bytes that make up the digital object. Consequently, the challenges related to preserving digital information are numerous. Foremost among them is the challenge of technological obsolescence. Rapid technological changes require frequent investments to overcome technological obsolescence in order to render (or read) digital information intelligible.[3] Furthermore, environmental decay plays a significant role in the loss of digital material. Deterioration of magnetic and optical media, and exposure to high temperatures and humidity, as well as contaminants, can cause failures in reading devices and hardware components.[4] When deterioration of storage media or hardware failure occurs, the digital information can be permanently lost, even if technological obsolescence is not a factor. Other challenges may include, but are not limited to, the scale or quantity of data being captured for

---

[1] Gail Fineberg, "Capturing the Web: Staff Briefed on National Digital Preservation Plan," *Library of Congress Information Bulletin*, vol. 62, no. 4 (April 2003), http://www.loc.gov/loc/lcib/0304/digital.html (accessed 18 October 2006).

[2] David M Levy, *Scrolling Forward: Making Sense of Documents in the Digital Age* (Arcade, New York, 2003). Simply, a digital object is made of two parts: the digital representation or the bits and bytes (e.g.., a Word document), and its perceptible forms produced from the digital representation. In regards to digital preservation, archivists and information professionals' concerns lie with the digital representation since the digital representations and their perceptible forms rely on a complex technological environment of hardware and software. The nature of the perceptible form means that it is created on the fly, or just-in-time, to maintain a visible and useful presence; consequently, the perceptible form and its digital representation cannot be separated, which impacts our decisions for the long-term management and preservation of digital objects. (For more information see pages 152 to 155.)

[3] Maggie Jones and Neil Beagrie, *Preservation Management of Digital Materials: A Handbook* (British Library; London, 2001), p. 24.

[4] Margaret Hedstrom and Sheon Montgomery, *Digital Preservation Needs and Requirements in RLG Member Institutions*, (Mountain View, CA; 1988), http://www.rlg.org/preserv/digpres.pdf, p. 1 (accessed 08 December 2004; updated 02 January 2007); also cited in the Digital Preservation Coalition, *Preservation Issues* website, http://www.dpconline.org/graphics/digpres/presissues.html (accessed 08 December 2004).

preservation, authenticity and context of the creation of the record, costs related to preservation management, and legal issues regarding intellectual property rights, access, and security. These challenges, therefore, can work independently of each other, or in tandem with each other.

Examples of digital data loss are numerous, and the consequences of this loss can be, and have been, unfortunate. The American National Aeronautics & Space Administration (NASA), for instance, lost track of Martian experiments done in 1975. The original data is stored on technology so old and obsolete that even the programmers have passed away. Consequently, important data is lost as current computers are unable to render the information into a readable format. Fortunately, an abridged printed or hard copy report is available, and some information was recovered.[5] Another news article, entitled "The End of History," examines the larger consequences of data loss. Charles Arthur writes, "In centuries to come, what will scholars be able to learn of the great events and discoveries of our time? As paper records are replaced by unstored e-mails and obsolete hardware, we may be entering a new digital dark age."[6] Arthur notes several examples of this digital dark age, and not all are related to the "calamity" caused by technological obsolescence. The Ivar Aasen Centre of Language and Culture, a Norwegian literary museum, was unable to access its computerized catalogue system following the death of the man who administered the system, and was the only person who knew the access password.[7] Another classic example is the BBC Doomsday Project,

---

[5] Kevin Krolicki, "NASA Data Point to Mars 'Bugs,' Scientist Says," *The Long Now Foundation Press*, (updated 27 July 2001), http://www.longnow.org/10klibrary/darkarticles/ArtNASAMars.htm (accessed 02 December 2004).
[6] Charles Arthur, "The End of History," *The Independent*, London, UK, (30 June 2003).
[7] Cornell University, *Digital Preservation Management Workshop*, http://www.library.cornell.edu/iris/tutorial/dpm/terminology/preservation.html (accessed 08 July 2004).

a multimedia affair to celebrate the 900^{th} anniversary of the original Doomsday Book, and to provide a social snapshot of Britain in the mid-1980s.[8] Within a few years, the software and hardware technology used had become obsolete, and the Doomsday project was in danger of being lost. Fortunately, digital preservation efforts were able to "rescue" the project, and render the multimedia program. Terry Cook, in a 1995 article for MIT's *The Technology Review,* identifies these problems and issues much earlier than much of the popular press.[9] Not only are corporations, government institutions, non-profit organizations, and private individuals at risk of losing data due to technological obsolescence, but the risk is further compounded due to the harsh environmental conditions in which tapes and disks are sometimes stored, as well as inadequate policies regarding electronic records management. Consequently, data loss is subject not just to outdated technology, but also to an organization's attitude towards proper records management and policies. A technical expert states that there are "the three things in life that are certain – death, taxes, and that your hard drive will fail," but best practices in digital preservation management will at least minimize the calamity of the failing hard drive![10]

The ability to capture digital material exceeds the ability to manage and preserve the information over a long period. This creates a twofold effect in that excessive or duplicated information is continually being saved and that information itself is kept on media and software that is no longer supported or accessible at the current time. Without a proper appraisal policy, unneeded information is retained at the expense of preserving

---

[8] Creative Archiving at Michigan & Leeds: Emulating the Old on the New (CAMiLEON), http://www.si.umich.edu/CAMILEON/ (accessed 23 August 2006).
[9] Terry Cook, "It's 10 O'Clock: Do You Know Where Your Data Are?," *The Technology Review* (January 1995).
[10] Arthur, "The End of History".

information of enduring value, while at the same time increasing the human, technical, and financial resources required to preserve and render the information.[11] As the costs of storage decrease, and the amount of space increases from gigabytes, terabytes, and petabytes to exabytes, many organizations tend to preserve more and more data than is necessary. The downside of this strategy occurs as the amount of data becomes harder to manage, organize, describe, and index for successful retrieval and dissemination. RIMS Division has addressed this challenge by implementing an appropriate appraisal policy to ensure that information of research value is retained in its library and archives collections and databases. Hence, the recommended digital preservation architecture will be in line with the development of the technical and managerial challenges in accessioning, managing, and providing access to digital materials.

RIMS Division is moving towards the long-term management of digital objects in establishing an appraisal policy and implementing best practices in records management, but the Division does not yet have a fully implemented digital preservation strategy. It lacked the technical understanding of how to manage and preserve digital objects, and addressed this challenge by hiring an archival intern for eight months. Most of the internship focused on an environmental scan of research and development in digital preservation. Previous and current work in digital preservation includes research not only from the field of archives, but also academic institutions, museums, libraries, other cultural heritage organizations, and science and technology. As a result, digital preservation coalitions have formed to research and develop appropriate methodologies.

---

[11] According to one recent posting on the ERECS list serve, the costs ratio for the storage and the management of digital objects is as high as 1:7, in which every dollar spent on storage requires seven dollars allocated to management and human resources. (Peter Van Garderen to Heather Lyons, "Re: Long-term preservation of electronic records" posted 17 October 2006).

Coalitions include the Digital Preservation Coalition (DPC), the Joint Information Systems Committee (JISC), Creative Archiving at Michigan & Leeds: Emulating the Old on the New (CAMiLEON), the International Research on Permanent Authentic Records in Electronic Systems (InterPARES), and the Consultative Committee for Space Data Systems (CCSDS), to name a few. In making their research, white papers, and other findings available on the Internet, these groups and coalitions have increased collaboration on research into long-term preservation management and have facilitated local, national, and international projects that document the successes and failures in preservation and access to digital material.

One of the earliest projects, the University of Pittsburgh's Functional Requirements for Evidence in Recordkeeping Project (1993-1996), was tasked with developing a Reference Model for Business Acceptable Communications.[12] The Pittsburgh Project, as it was more commonly known, outlined thirteen recordkeeping functional requirements, emphasized the concept of metadata for recordkeeping, used production rules to formally express the functional requirements, and promoted compliance with laws, regulations, and standards through literary warrants. The functional requirements ensure that reliable records are captured, maintained and accessible. They were used to test whether an information system is compliant with or

---

[12] For more information on the University of Pittsburgh's Functional Requirements for Evidence in Recordkeeping Project see http://www.archimuse.com/papers/nhprc/. See also Rick Barry, "Making a Difference: Comments on Electronic Records Management R&D Projects at Ohio State University, Indiana University and City of Philadelphia: SAA Annual Meeting August 29, 1996, Session 6," www.mybestdocs.com/barry-r-saa1996-er-rd.html; David Bearman, *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*, (Archives & Museum Informatics, Pittsburgh, 1994), http://www.archimuse.com/publishing/electronic_evidence/ElectronicEvidence.Intro.pdf#search=%22%22Electronic%20evidence%3A%20strategies%20for%20managing%20records%20in%20contemporary%20organizations%22%22; and Wendy Duff, "Ensuring the Preservation of Reliable Evidence: A Research Project Funded by the NHPRC", *Archivaria* 42 (Fall 1996), pp.28-45 (all websites accessed 11 October 2006.)

meets these recordkeeping standards. Both the functional requirements and production rules delineated the type of metadata required to accompany each record. The metadata also provided detailed information on terms and conditions of use, data structure, provenance, content, and on the use of the record after its creation. Finally the literary warrant provided authority to the functional requirements. The Pittsburgh Project was one of the first approaches that embraced the functional approach to records and preservation management as developed by David Bearman. At the time, Bearman's work and the Project represented advanced archival thinking, and inspired other projects being undertaken at the University of Indiana,[13] the World Bank, and the City of Philadelphia.

Two initiatives out of the United Kingdom are the DPC and JISC that work to foster and support research on the challenges of securing and preserving digital resources.[14] Both organizations have produced and disseminated information on current research and practice in digital preservation. This information, available on the DPC and JISC websites, is provided by groups of professionals and experts in the three major areas of research – migration, emulation, and encapsulation. The DPC and JISC, however, emphasize different aspects of digital preservation research – the DPC fosters action and the JISC supports advanced education through strategic guidance, advice, and opportunities.

The CAMiLEON project was a joint initiative between the UK and the United States, with funding from the JISC and the National Science Federation (US).[15] Headed

---

[13] Information on the University of Indiana's *Sound Directions: Digital Preservation and Access for Global Audio Heritage* can be found at http://www.dlib.indiana.edu/projects/sounddirections/ (accessed 11 October 2006).
[14] Digital Preservation Coalition (DPC), http://www.dpconline.org/graphics/index.html; Joint Information Systems Committee (JISC), http://www.jisc.org.uk/ (all websites accessed 21 October 2006).
[15] CAMiLEON.

by Chris Rusbridge and Margaret Hedstrom, respectively for the UK and US teams, this project analysed and implemented an emulation strategy for recovering the digital information gathered by the BBC Doomsday project in the 1980s. CAMiLEON's foremost objective explored options for long-term digital preservation, including the viability of emulation as a strategy. As a result of its study, CAMiLEON successfully demonstrated the effectiveness of emulation to enable future computer (software and hardware technology) systems to operate and render readable information in obsolete digital technology, as well as re-create that material's "look and feel".

At Australia's Monash University, the Records Continuum Research Group (RCRG) is developing a continuum model for information management and digital preservation.[16] Led by Sue McKemmish, this group focuses on three major research areas – Metadata Control, Continuum Modelling and Research, and Recordkeeping and Legal Systems. The group studies the specifications and requirements for metadata in creation, capture, organization, and archiving of records, information continuum management and core archival functions in conjunction with the Information Enterprise Group, and the socio-legal mandates and compliance issues relating to the records management continuum. Part of the RCRG's research has led to the development of the Australian Recordkeeping Metadata Schema, a standardized set of structured recordkeeping metadata elements and a mapping framework, which evolved out the Strategic Partnerships with Industry - Research & Training (SPIRT) Project with other organizations such as the National Archives of Australia. As with the Pittsburgh Project, the Recordkeeping Metadata Project aligns itself to the functional approach in which

[16] Monash University Records Continuum Research Group (RCRG), http://www.sims.monash.edu.au/research/rcrg/index.html (accessed 19 October 2006).

records are evidence of transactions, yet differs as it takes a continuum perspective in which the archivist takes an "active" role in managing and preserving, as opposed to the archivist's traditional role "as the keeper or guardian of the record".

Closer to home, the United States' National Archives and Records Administration (NARA) has been working on an Electronic Records Archives (ERA) initiative.[17] Launched in 1998, ERA encompasses a comprehensive, systematic, and dynamic approach to the preservation of digital objects, which is based upon the Open Archival Information Systems (OAIS) reference model and standard. This approach is based upon NARA's collection-based persistent digital archives in which NARA will "authentically preserve and provide access to any kind of electronic record, free from dependence on any specific hardware or software".[18] Now moving into its second stage, NARA signed a landmark agreement with the San Diego Supercomputer Center, which provides a state-of-the-art technical infrastructure for managing and administering digital objects transferred to NARA.[19] This innovative collaboration aligns with NARA's goals to develop and maintain the technical capability to capture, manage, preserve, and make accessible digital objects. NARA and the San Diego Supercomputer Center aim to have a functional subset of the system operational in 2007.

The University of British Columbia headed a research project entitled *The Preservation of the Integrity of Electronic Records* (1994-1997), otherwise known as the

---

[17] National Archives and Records Administration (NARA), Electronic Records Archives (ERA), http://www.archives.gov/era/ (accessed 20 October 2006).
[18] Ibid.
[19] NARA, *The National Archives and the San Diego Supercomputer Center Sign Landmark Agreement to Preserve Critical Data* Press Release (28 June 2006), http://www.archives.gov/press/press-releases/2006/nr06-119.html (accessed 21 October 2006).

UBC Project.[20] The findings of the UBC Project led to an international digital preservation coalition known as the International Research on Permanent Authentic Records in Electronic Systems (InterPARES), headed by Luciana Duranti and Terry Eastwood, in collaboration with other academics, national archival institutions and private industry representatives from around the world.[21] The overall goal of InterPARES is to develop a theoretical and methodological strategy for the long-term preservation of authentic records. As such the project has two phases: InterPARES 1 (1999-2001) focused on the preservation of the authenticity of records, including an exploration of the issues related to the long-term digital preservation of audio files; whereas InterPARES 2 (2002-2006) continued with research into the preservation of audio material, while delving into the issues of reliability and accuracy from the perspective of the lifecycle of the record. InterPARES 1 was composed of four domains: Authenticity Domain, Appraisal Domain, Preservation Domain, and Strategy Domain. The Preservation Domain is of particular interest for this thesis. This domain's main findings are that: only the means to reproduce digital objects can be preserved; an account of the document's custodial history must include any transformations or changes that occur to both the physical and perceptible form of the digital object; archival policies must accompany the technical solution; and the digital preservation solutions themselves

---

[20] University of British Columbia, School of Archival, Library and Information Studies (SLAIS), http://www.interpares.org/UBCProject/index.htm (accessed 21 October 2006). One of the outcomes of this project was the collaboration with the United States Military in the development of "Design Criteria Standard for Electronic Records Management Software Applications" recordkeeping standard (DoD 5015.2). Information on this standard can be accessed at http://www.dtic.mil/whs/directives/corres/pdf/50152std_061902/p50152s.pdf (accessed 21 October 2006).

[21] International Research on Permanent Authentic Records in Electronic Systems (InterPARES), http://www.interpares.org/ (accessed 21 October 2006).

are inherently dynamic.[22] InterPARES 2 focused on the concepts, principles and criteria, and methods in the creation and maintenance of authentic digital objects, including the preservation of the object in the context of artistic, scientific, and government activities.[23] The guiding methodological principles were unique to InterPARES 2, in that InterPARES 2 adopted an open inquiry approach in which each working group identified its own perspectives, research design, and methodology each believed was most appropriate for its task. This allowed for greater diversity within the overall objectives of the project. In developing standards, methods, and specifications for the authenticity of digital objects, InterPARES also adopted one of the newly recognized international standards, the OAIS reference model, as part of the project's approach to the long-term management and preservation of digital objects.

Collaborative research between NASA, the Research Libraries Group (RLG), and the Online Computer Library Center (OCLC) led to the development of the ISO long-term digital preservation standard known as the Open Archival Information System (OAIS) reference model (ISO 14721).[24] This model is described as both a blueprint and an architecture for managing digital material. Furthermore, OAIS has developed a well-defined set of terms and definitions for increased communication and collaboration among researchers. This model, therefore, allows for greater standardization in the use ofterms and concepts (such as "archive," "long term," "designated community") among information managers and information management vendors, and promotes an awareness

---

[22] InterPARES 1, Preservation Domain, http://www.interpares.org/ip1/ip1_ptf.cfm (accessed 21 October 2006).
[23] InterPARES 2, Project Summary, http://www.interpares.org/ip2/ip2_index.cfm (accessed 21 October 2006).
[24] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)* (January 2002), http://public.ccsds.org/publications/archive/650x0b1.pdf (accessed 08 July 2004).

of the requirements needed for digital preservation. The model also articulates the need for digital preservation to begin with the object's or document's active phase of its lifecycle when there is concern about the impact of changing technologies on the accessibility and management of the object. For these reasons, the model has been widely accepted in North America, superseding almost any other approach to digital preservation. Consequently, OAIS has had a major impact on archival methodology in North America and Europe. Most strategies being developed today strive to be OAIS compliant.

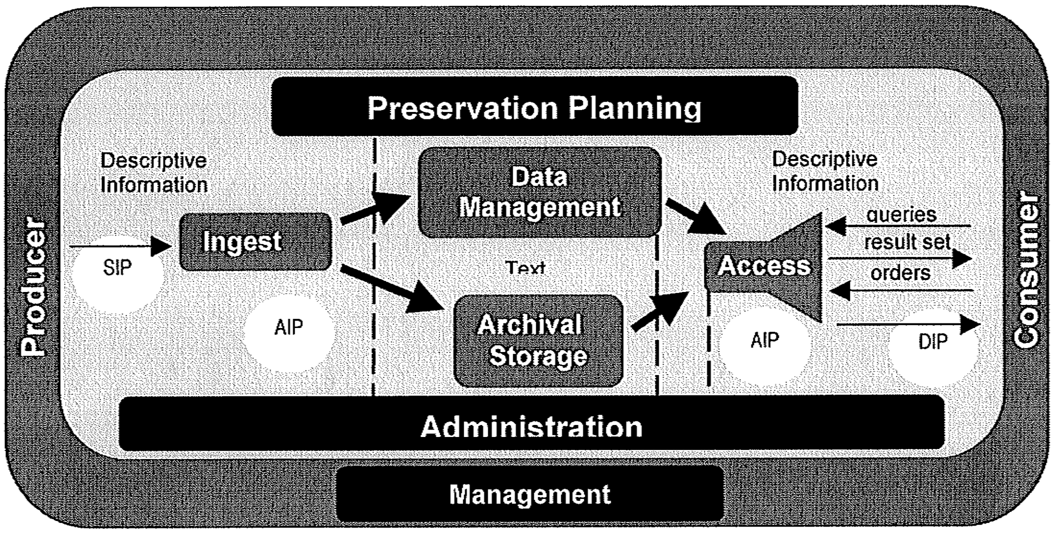
Open Archival Information System (OAIS) Reference Model[25]

The OAIS model allows for the ingestion, or submission, of a submission information package (SIP), the storage of the archival information package (AIP), and the dissemination of the dissemination information package (DIP). Once the submission information package has been accepted, it is then transferred to the archival storage repository. Once placed in the archival storage repository, the submission information package becomes the archival information package. When a request is made for an archival information package, the archival information package is retrieved, and disseminated as the dissemination information package.

---

[25] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*.

Figure 1: Open Archival Information System (OAIS) reference model[26]



Each stage (from ingest to storage to dissemination) of the information package is guided by the requirements for long-term preservation and access. In order to accept the submission information package, the producer must meet the proper metadata, format, and documentation standards. Preservation metadata[27] (which includes descriptive, technical, structural, and rights metadata) is part of the trusted digital repository to ensure authenticity and reliability of the content information contained within the archival information package. Access to the archival information package depends upon the existence, description, location, and availability of information stored in the OAIS archive, as well as on allowing users to request and receive information products. When

[26] Cornell University, *Digital Preservation Management Workshop*, http://www.library.cornell.edu/iris/tutorial/dpm/terminology/oais.html (accessed 08 July 2004).
[27] The OAIS model provides only a general framework regarding preservation metadata. Other pilot studies have expanded upon the OAIS model to identify the specific metadata for digital preservation.

a user requests a particular archival information package, and the package is located according to its unique identifier, the OAIS archive sends the requested dissemination information package to the user. If the dissemination information packages cannot be opened or read by the user, then access to the archival information package has failed. Long-term preservation and access are achieved when the archival information package has remained accessible to the designated user community over a long period of time during which time the hardware and/or software used to read the archival information package may have become obsolete.

The OAIS "is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community." An OAIS model is distinguished from other information repositories in that an OAIS is committed to long-term preservation of information through open communication relating to the development of recommendations and standards for accomplishing its primary commitment. Long-term preservation focuses on the organization's ability to maintain accessibility to information irrespective of technological or designated community users changes. Thus, the term "Open" in OAIS is used to imply that all recommendations are developed in open forums, and it does not imply that access to the archive is unrestricted.[28] However, the OAIS model is concerned with the preservation of materials that are both digital and physical. This reference model:

---

[28] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS),* p.2-1.

- Provides a framework for the understanding and increased awareness of archival concepts needed for Long Term digital information preservation and access;
- Provides the concepts needed by non-archival organizations to be effective participants in the preservation process;
- Provides a framework, including terminology and concepts, for describing and comparing architectures and operations of existing and future archives;
- Provides a framework for describing and comparing different long term preservation strategies and techniques;
- Provides a basis for comparing the data models of digital information preserved by archives and for discussing how data models and the underlying information may change over time;
- Provides a foundation that may be expanded by other efforts to cover long-term preservation of information that is NOT in digital form (e.g. physical media and physical samples);
- Expands consensus on the elements and processes for long-term digital information preservation and access, and promotes a larger market which vendors can support;
- Guides the identification and production of OAIS-related standards.[29]

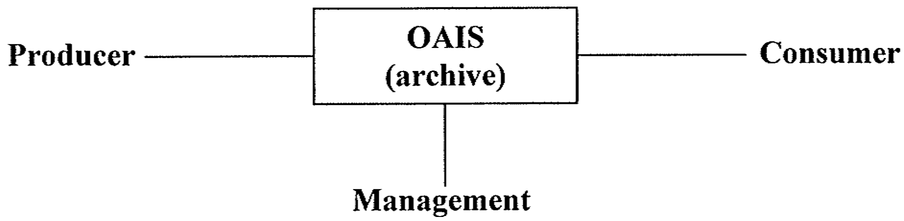
The OAIS model not only helps to preserve records over the long term, but also to make them available to a designated user community. As such, the strategy focuses on

[29] Ibid., p.1-1.

the information deemed for long-term preservation, even if the OAIS repository is not permanent. This model may also be applied to organizations that are responsible for processing and distributing information in response to programmatic needs, as is the case with RIMS Division. Once captured and retained as part of the RIMS Division's Library Archives collection, the digital objects and material are distributed via the Internet, the Intranet (ECHOnet), and the Library research databases (BIBLIO and IMAGES). RIMS Division's goal is to create an open archival repository that is accessible to a designated community of staff, project recipients, and researchers working on program initiatives. Combined with a trusted digital repository, the designated user community will then have access to authentic and reliable research material.

A simple OAIS environment consists of a producer, a consumer, management, and the OAIS archive (which also includes the record). The producer, consumer, and management reside outside the OAIS archive, and are described, respectively, as the person or client system that provides the information to be preserved, the person or client system that requests desired information (such as the designated community), and those who set the overall OAIS policy, including any additional responsibilities associated with the operation of the OAIS archive. The management is not responsible for the daily operation of the OAIS, as that falls within the administration function of the model.

Figure 2: The OAIS Archive[30]



The producer is essentially the creator of the digital object that is contained in the OAIS archive. When the producer submits a digital object for long-term preservation, other information will accompany the content information of the digital object. This other information consists of the preservation description information, packaging information, and descriptive information. Each of these components includes various metadata relating to the content information provenance, reference (unique identifiers), and fixity to protect the archival information package (or object) from unauthorized alterations. Aside from the metadata, volume/file structure and name/directory information are also part of the packaging information. The content information, along with the preservation description information, forms the Archival Information Package.

The Consumer is a member of the Designated Community who may request archival information packages located in the OAIS archive. The consumer submits a query to the OAIS archive that returns a result set that matches the criteria of the query. The consumer chooses the desired record from the result set, and the requested archival

---

[30] Ibid., p.2-2.

information package is processed as the dissemination information package sent to the consumer. The requests may be ad hoc or event-based. The ad hoc request for dissemination information packages involves a broad general search that is refined through subsequent searches until the desired dissemination information package is found. Under these circumstances, the consumer generally does not have prior knowledge of the OAIS archives holdings. Event-based requests differ as the consumer automatically receives dissemination information packages based upon a triggering event. This may be a periodic event, such as the submission of a monthly report, or a unique event, such as the ingestion of a specific archival information package. When the consumer receives the dissemination information package, only the content information is presented in most cases. The other packaging information may or may not be available to the consumer.

Management oversees the policy and procedures, and the technological and resources frameworks that carry out the responsibilities of the OAIS archive. The minimum responsibilities include acceptance of the appropriate submission information packages (referred to as the negotiation of submission information packages), obtain control of the information for long-term preservation, identify the designated community and consumers, ensure that the content information will be understandable for the designated community, prepare and follow the policies and procedures to guard against all reasonable contingencies, and ensure long-term accessibility to the content information. An OAIS archive progresses through five stages (Acknowledge, Act, Consolidate, Institutionalize, and Externalize) as the management achieves these framework responsibilities.[31] As illustrated in Figure 1, the management role "encases" or supports the OAIS reference model since the management provides the foundation for the trusted digital repository, and the OAIS framework responsibilities. In the reference model, the role of management is outlined as follows:

> Management provides the OAIS with its charter and scope. The charter may be developed by the archive, but it is important that Management formally endorse archive activities. The scope determines the breadth of both the Producer and Consumer groups served by the archive.[32]

Some examples of typical interaction between the OAIS and Management include:

- Management is often the primary source of funding for an OAIS and may provide guidelines for resource utilization (personnel, equipment, facilities).
- Management will generally conduct some regular review process to evaluate OAIS performance and progress toward long-term goals.
- Management determines, or at least endorses, pricing policies, as applicable, for OAIS services.
- Management participates in conflict resolution involving Producers, Consumers and OAIS internal administration.

Applying the OAIS archives to IDRC, the following roles may be described as such:

---

[31] The five stages were part of an exercise taught at the Cornell University *Digital Preservation Management Workshop* that I attended in July 2004. These stages were also part of a presentation, entitled *The Three-Legged Stool: Institutional Response to Digital Preservation* (March 2005), given by Anne R. Kenney and Nancy Y. McGovern. This presentation can be accessed at http://www.library.cornell.edu/iris/dpo/docs/Cuba-arknym_final.ppt (accessed 14 March 2006).

[32] Consultative Committee for Space Data Systems, p.2-8.

- The producer would include the author(s) of the research documentation and material submitted for long-term digital preservation.
- The consumer and designated community are those requesting research documentation and material from the Library Archives. This may be limited to IDRC employees, or open to the general public.
- RIMS Division would be responsible for the role of management.

Trusted Digital Repository

The trusted digital repository "structurally maps the organizational context for a digital archive handling the size, range, and type of digital materials that cultural repositories are or will be responsible for in the future."[33] To be considered trustworthy, a digital repository must be able to maintain the integrity of its research documentation and material for both the potential stakeholders (for example, its depositors, and funding agencies/persons) and its designated user community. As such, a trusted digital repository must be sustainable, and

- Identify essential organizational, curatorial, and operational responsibilities.
- Address high-level agenda recommendations such as certification, requisite tools, cooperative models, comprehensive archival system design and development, intellectual property rights, preservation strategies, and metadata.
- Map the corporate trusted digital repository to the OAIS model to ensure OAIS compliance.

---

[33] Cornell University, *Digital Preservation Management Workshop*, http://www.library.cornell.edu/iris/tutorial/dpm/foundation/index.html (accessed 08 July 2004).

Seven key attributes have been identified that are required for a trusted digital repository. These are: [34]

0. OAIS compliance

1. Administrative responsibility

2. Organizational viability

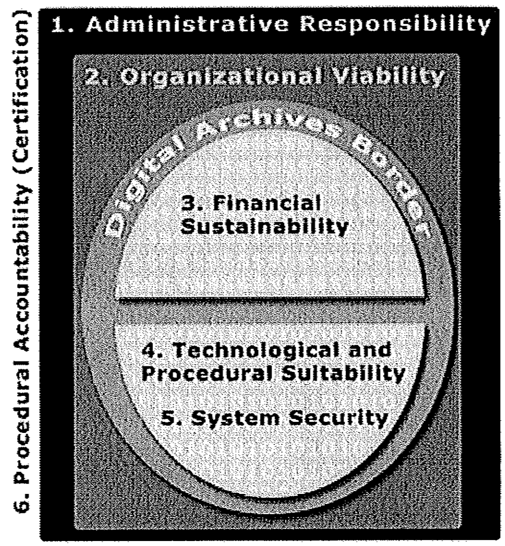
3. Financial sustainability

4. Technological and procedural suitability

5. System security

6. Procedural accountability

Figure 3: Trusted Digital Repository Model[35]



The trusted digital repository model illustrates how these seven attributes are related to each other (see chapter three). Some are nested within a "Digital Archives Border" while others will be external to the digital archive itself, or even the parent institution. In conjunction with the OAIS model, the management makes explicit its commitment to the development of a trusted digital repository that complies with prevailing standards, policies, and practices. In so doing, the mission statement, legal status, and operations are demonstrated to sufficiently support such a repository. In other words, the management provides comprehensive documentation of all relevant policies,

---

[34] Ibid.
[35] Ibid.

procedures, and practices. As part of this commitment, a sound financial base is also established to sustain the management's long-term preservation plan and mandate. Overall, these attributes work together to retain the trusted digital repository's status as being OAIS compliant.

Long-term digital preservation management requires organizational infrastructure, technological infrastructure, and a resources framework in order to build a trusted digital repository. The organizational infrastructure is expressed in a comprehensive policy framework. The technological infrastructure entails a flexible preservation plan that can incorporate changing technological platforms over time. A resources framework focuses on the staffing, technological, operational, and other costs associated with maintaining the organizational and technological infrastructure of a trusted digital repository. These are the core requirements necessary for the long-term preservation management of digital information.

These are just a few of the current projects that focus on long-term digital preservation management and access to digital materials. Most projects, although applicable to a broader scope of digital archival methodologies, are generally more focused on an individual objective such as the CAMiLEON BBC Doomsday project. In addition, these projects have a substantial budget and a team of archival, library, and information technology experts to develop and test a digital preservation strategy. Much of this research is highly theoretical, and not implemented at a more practical level, particularly with a small to mid-size institution such as the International Development Research Centre (IDRC) in mind. In using IDRC as a case study, this thesis will add to the current field of research by demonstrating that a mid-size institution can successfully develop and implement a digital preservation management strategy that incorporates both archival theory and practice. The next two chapters present my recommendations to IDRC for a digital preservation strategy, and the last chapter concludes with a brief summary about the themes of research and collaboration of information management and information technology professionals who are working together in learning about digital preservation strategies.

CHAPTER TWO

Long-term Management and Preservation of Digital Information: The International Development Research Centre as a Case Study

The first part of this chapter focuses on the organizational structure of IDRC, and the role RIMS Division plays as the leader for the centre's digital preservation strategy. The chapter also contains a summary of my eight-month internship that supported RIMS Division's mandate to manage and preserve the corporate memory of the centre and its digital objects. The second half of the chapter discusses RIMS Division policies as they relate to digital preservation issues. These include appraisal, costs, disaster and emergency recovery plan, and the preservation of e-mail and websites. Taken together, chapters two and three form the bulk of my final internship report.

The Creation of the International Development Research Centre

In 1967, at a banquet address given to the Canadian Political Science Association at Carleton University, Prime Minister Lester B. Pearson spoke of Canada's role in international efforts to address the challenges facing developing communities and countries through scientific and technological advances.[1] In this address, the concept of an international development research centre was first proposed. Pearson saw this centre as a place "for a more scientific, a more rational and a more systematic approach to the

[1] International Development Research Centre, RIMS Division Library Archives, "Notes For The Prime Minister's Remarks To The Canadian Political Science Association Banquet Carleton University, Ottawa, June 8, 1967," http://idrinfo.idrc.ca/archive/corpdocs/010474/Pearso5.pdf (accessed 28 February 2005).

practice of politics and to the work of government."[2] In his vision, he urged social and political scientists to work together to make greater practical use of their academic knowledge and training. In Pearson's words, the challenge "for international development is to find new instruments for concentrating more attention and resources on applying the latest technology to the solution of man's economic and social problems on a global basis. One idea for a new Canadian initiative in meeting this challenge that should be considered is for the establishment of a centre of International Development [....] to act as an internationally recognized focal point for concentrating attention and interest on this vital challenge to all of humanity."[3]

Taking on this vital challenge of international development, according to Pearson, meant an approach that was based as much on academic knowledge as it was on practical experience. Maurice Strong, responsible for the Canadian government program for external aid, developed and fostered this idea of a politically independent organization, and hired A.F.W. Plumptre to conduct a feasibility study for the establishment of an international development research centre.[4] Several institutions and individuals in Canada and around the world indicated that such a centre was needed and that Canada possessed the means to promote and sustain the type of research that was to be the core of the International Development Research Centre (IDRC) of Canada. As part of the creation of IDRC, Plumptre outlined several objectives in his study – appraisal of past and present development activities, provision of a comprehensive databank on research

---

[2] Ibid., p.4.
[3] Ibid., p.6.
[4] International Development Research Centre, RIMS Division Library Archives, "A.F.W. Plumptre, Feasibility Study on the establishment of IDRC, December 22, 1967," http://idrinfo.idrc.ca/archive/corpdocs/117258/01e.pdf, and "Memorandum Regarding the Proposed International Development Research Centre, January 24, 1968," http://idrinfo.idrc.ca/archive/corpdocs/117258/letters1.pdf (all websites accessed 15 October 2005).

development, harnessing modern information media and communication for improving the quality of life in developing communities and countries, and support and encouragement of Canadian research in the fields of humanities and social sciences in developing countries. Plumptre also indicated in his study that first class leadership, not just in management, but also in the areas of research, education, and communications were essential.

In the September 1968 Speech from the Throne, the Governor General briefly mentioned Canada's opportunity to provide foreign aid to developing countries. He said that the natural wealth Canada was "blessed" with "must serve to remind us that we have a substantial responsibility and a great opportunity in matters of foreign aid. My Government intends to press forward its plans and programs for international cooperation and development which include the establishment in Canada of an International Development Research Centre."[5] This statement was followed by a Memorandum to the Cabinet on the IDRC dated 9 October 1968. The memorandum outlined the proposed mission of IDRC. On an administrative level, IDRC would be a Crown corporation, which would report to Parliament through the Minister of Foreign Affairs; yet the centre would operate autonomously from the Canadian government. Hence, the centre "would be established as an independent Canadian entity" for initiating, conducting, and supporting studies, projects, and programmes addressing the needs of developing communities, assisting in the development of scientific and technological capabilities of developing countries, and providing facilities to the Canadian government for effectively

[5] International Development Research Centre, RIMS Division Library Archives, "House of Commons Debates (First Session – Twenty Eighth Parliament 17 Elizabeth II, Volume 1, 1968) Ottawa 1969," http://idrinfo.idrc.ca/archive/corpdocs/069849/03aE.pdf (accessed 15 October 2005).

dealing with the socioeconomic problems of developing countries. Cabinet approved the IDRC initiative on 15 April 1969.

Over the next few months, a proposed strategy for the establishment of the IDRC and a draft of *An Act to establish the International Development Research Centre of Canada* were presented to Prime Minister Trudeau.[6] This strategy explored the possible scope of IDRC's research. The report supported the establishment of the centre as a "think-tank" or an institute of advanced study and the formation of a research council to support development research in Canadian universities. Specialized research institutions in developing countries would be identified and receive support from IDRC. Once the centre's core staff had been established and projects identified, advisory groups and a "home-base" for Canadian scientists working abroad were to be created. The advisory groups would work in consultation with senior management to identify the priority areas of research and establish an informal network of communications with other centres and institutions in both the developing countries and internationally. The proposed act followed the recommendations found in the strategy to establish the centre. The draft legislation also reflected a recommendation that stressed the importance of information management. The proposal for the centre stated that a "special role [was] envisaged for the Centre in the field of assimilation, storage and retrieval of information and data in

---

[6] International Development Research Centre, RIMS Division Library Archives, "Working Paper on a Strategy for the Establishment of the International Development Research Centre of Canada, 26 May 1969," http://idrinfo.idrc.ca/archive/corpdocs/117258/letters3.pdf, and International Development Research Centre, RIMS Division Library Archives, "Archiv No. 341.1 IDRC I581, International Development Research Centre of Canada: Memorandum to Cabinet Supporting Proposed International Development Research Centre and Draft of an Act to Establish the IDRC of Canada, 30 June 1969," http://idrinfo.idrc.ca/archive/corpdocs/069849/02e.pdf (all websites accessed 15 October 2005).

various areas of the development field."[7] This special role was expressed in Section 4(2) of the proposed act as the power to,

> (a) establish, maintain and operate information and data centres and facilitate research and other activities related to its purpose;
> (e) publish or otherwise disseminate scientific, technical or other information[8]

Since 13 May 1970, when the IDRC act received Royal Assent, the centre has been widely recognized as a leading institution in development research. From the beginning IDRC has emphasized the need for "social innovation" to guide its programming. To do so, IDRC created four program areas – agriculture, food and nutrition sciences; information sciences; population and health sciences; and social sciences and human resources. These program areas, which managed the research activities of the centre, focused on profitable farm and fishing productivity; a cooperative system of information sharing and exchange; health concerns arising from rapid population growth; and identification of major research issues in urban-rural relations, human resources development, science policy, and education, respectively. The centre's program areas remained fairly stable until the 1980s when IDRC obtained a new diplomatic role in international aid.

IDRC's diplomatic role was formalized into a new research area in the 1990s, when democracy and development was added as a new program area.[9] Environment also became another important area of research as IDRC expanded its mandate to include

---

[7] International Development Research Centre, RIMS Division Library Archives, "Archiv No. 341.1 IDRC I581, International Development Research Centre of Canada: Memorandum to Cabinet Supporting Proposed International Development Research Centre and Draft of an Act to Establish the IDRC of Canada, 30 June 1969," pp.2 and 5.
[8] Ibid., p.9.
[9] International Development Research Centre, *The International Development Research Centre: A Brief History*, http://www.idrc.ca/uploads/user-S/11394255461History_rev_e.pdf (accessed 16 January 2007), p.10.

research associated with the politics of oil and nuclear power in developing countries.[10] These changes led to a new three-year program framework (1993-1996), in which program areas were required to encompass some aspect of sustainable and equitable development. These changes from the centre's initial objectives meant that IDRC had become a more integral *expression* of Canadian foreign policy. The 1990s was also marked by a decline in funding from the Canadian government, as it attacked massive budget deficits. To make up the loss of revenue, IDRC formed partnerships with international secretariats to share costs on research projects that IDRC could not support on its own.[11] In this manner, IDRC continued to fund its program areas, and maintain its reputation as a leading institution in research for developing countries.

As IDRC entered the new millennium, the centre developed a new five-year strategic plan (2000-2005), which promoted the theme of "Empowerment through Knowledge."[12] The plan continued to uphold the key principles that guide IDRC's work; however, a new emphasis was placed on understanding the relationship between research, knowledge, and gender issues as they relate to development. In terms of projects, IDRC began to fund research on social and economic equity, environment and natural resource management, and information and communication technologies for development. These research areas carried on into IDRC's current five-year strategic plan (2005-2010).[13] As well, increased public awareness of international research and financial support from the

---

[10] At the 1992 United Nations Conference on Environment and Development, then Prime Minister Brian Mulroney designated IDRC as the Canada's lead organization in a global multi-government environmental action plan.

[11] International secretariats were conglomerates of various agencies and institutions that shared a similar interest in funding an IDRC research project.

[12] International Development Research Centre, *Corporate Strategy and Framework Program 2000-2005*, http://idrinfo.idrc.ca/archive/corpdocs/116539/CSPF_2000-2005.htm (accessed 19 January 2007).

[13] International Development Research Centre, *Corporate Strategy and Framework Program 2005-2010*, http://www.idrc.ca/uploads/user-S/11250758901CSPF_2005_e.pdf (accessed 24 April 2006).

federal government gave IDRC the ability to increase and secure funding for its existing program areas while funding new projects such as gender equity, as mentioned above.

Today, IDRC funds more than 460 active research projects around the world. In funding applied research, researchers from and working in developing countries are able to take the lead in producing and applying knowledge to improve the health, prosperity, and social equality of their communities. Consequently, IDRC is one of the world's leading institutions in the generation and application of new and applied knowledge, and has a strong international reputation for providing much needed support and resources in the application of research knowledge and the creation of innovation that benefits the developing communities.

International Development Research Centre: Functional Structure

The IDRC act outlines the objectives, powers, and mandate of the centre. The act grants IDRC the authority to initiate research projects, assist other organizations, individuals or governments (including non-Canadian governments) in international development research, and disseminate information and data through regional centres, to name a few. The act is also noteworthy (and unusual at the time of its writing) because it provided for international membership on the Board of Governors. Thus, "the perspective and experience of developing countries [was] represented."[14] Furthermore, IDRC's governance ensured that IDRC could not become an instrument of the Canadian government and its foreign policy. In fact, "the *Act* was explicit in making sure IDRC

---

[14] International Development Research Centre, *The International Development Research Centre: A Brief History*, p.4.

was not an 'agent of Her Majesty.'"[15] Another unique feature is the provision for the creation of a library and research centre. The library and research centre is also identified as one of the centre's core functions, which provides technical and information support to four hundred employees, research recipients, and external researchers located in the centre's Ottawa headquarters and the six regional offices (Singapore, Dakar, Cairo, Montevideo, New Delhi, and Nairobi). Under the RIMS Division, the centre's library and archives are responsible for the corporate information and project research outputs, and for providing research material and support to staff, project recipients, and researchers.

IDRC's organizational structure is unlike larger foreign aid organizations such as the Canadian International Development Agency (CIDA).[16] Unlike CIDA, which operates entirely within the federal government, IDRC is an independent Crown corporation that works at arm's length from the Ministry of Foreign Affairs. Consequently, IDRC's primary function is to provide technical support through research and development, and funding to applied research in developing countries.

IDRC governance is distinct from European and other North American international research centres in that its Board of Governors consists of twenty-one Canadian and non-Canadian members.[17] The eleven Canadian members include the President, Chair and Vice-Chair. Additionally, the IDRC act also allows for the appointment of two members from the Senate or the House of Commons. The other ten

---

[15] Ibid., p.9.

[16] More information about CIDA is available on the website at http://www.acdi-cida.gc.ca/index-e.htm.

[17] International Development Research Centre, *Annual Report 2004-05*, http://www.idrc.ca/uploads/user-S/11307698851AnnualReport_2004-05_e2.pdf (accessed 02 March 2006), and the general website provides additional information on the Board of Governors, and their responsibilities.

members have been selected from developing and donor countries. There are eight non-Canadian board members from developing countries, and two from donor countries. This international composition makes IDRC's Board of Governors distinct among Canadian public corporations, yet ensures that IDRC programs meet both the needs of the developing world, and functions of the centre.

The primary function is outlined in IDRC's Corporate Strategy and Program Framework (CS+PF). This strategic program includes the directions for several divisions such as the program areas, finance and administration, human resources, and most recently communications.[18] It is in the program areas that the primary functions of IDRC are fulfilled, which include environment and natural resource management; information and communication technologies for development; innovation, policy and science; and social and economic policy. Within each of four areas are several project initiatives that support IDRC research and development. As such, IDRC pursues three objectives as outlined in its strategic plan:

- To strengthen and help mobilize the local research capacity of developing countries, especially in the program areas of Environment and Natural Resource Management, Innovation, Policy and Science, Information and Communication Technologies for Development, and Social and Economic Policy;
- To foster and support the production, dissemination and application of research results that lead to changed practices, technologies, policies, and laws that promote sustainable and equitable development, human rights, and poverty reduction;

[18] International Development Research Centre, *Corporate Strategy and Framework Program 2005-2010.*

- To leverage additional Canadian resources for research and development by creating, reinforcing, funding, and participating in partnerships between Canadian institutions in the developing world.[19]

As part of its mandate, IDRC also fosters the sharing of knowledge through communication with scientific, academic, and other communities in Canada and in developing countries. This mission of "Empowerment through Knowledge" has emerged since IDRC was established.[20] Part of the approach, and later a sub-function of IDRC, comes from the Information System Group that is responsible for maintaining, managing, and disseminating research information to other international research institutions, IDRC staff, and senior management, including the Board of Governors, and for ensuring effective management of research and development initiatives or programs.[21] For the first time, the 2005-2010 Corporate Strategy and Program Framework officially formalized this sub-function as part of the Communications strategy. Thus, the overall policy centres on the legislation, in which the objectives of IDRC are

> *to initiate, encourage, support and conduct research into the problems of the developing regions of the world and into the means for applying and adapting scientific, technical and other knowledge to the economic and social advantage of those regions and in carrying out these objectives:*
>
> *a) to enlist the talents of natural and social scientists and technologies of Canada and other countries;*

---

[19] International Development Research Centre, *Briefing Document* (February 2006), http://www.idrc.ca/uploads/user-S/11408155631Briefing_book_Jan2006_e.pdf p. 3-1 (accessed 24 April 2006).

[20] International Development Research Centre, *About IDRC – Briefing Book* website http://www.idrc.ca/en/ev-66231-201-1-DO_TOPIC.html (accessed 24 April 2006).

[21] International Development Research Centre, RIMS Division Library Archives, "Memorandum Regarding the Objectives, Style and Structure: a Possible Approach for the International Development Research Centre of Canada," http://idrinfo.idrc.ca/archive/corpdocs/117258/letters6.pdf (accessed 15 October 2005)

b) to assist the developing regions to build up the research capabilities, the innovative skills and the institutions required to solve their problem;
c) to encourage generally the coordination of international research; and,
d) *to foster cooperation in research on development problems between developed and developing regions for their mutual benefit.*[22]

Research Information Management Service (RIMS) Division: Mandate and Functions

The Research Information Management Service (RIMS) Division manages IDRC information resources and services through the Division's Library Archives and the Records Management Unit (RMU). The Library Archives contains IDRC project documents, as well as related resources such as images, and IDRC supported or produced publications, while the Records Management Unit focuses on the operational and administrative records generated and received by IDRC, in addition to managing records of historical value. Information of research value to IDRC projects is retained in the Library Archives, and is searchable through the RIMS Division online databases (BIBLIO and IMAGES). Records of historical value are the responsibility of the RMU. The Library Archives and Records Management Unit work cooperatively to acquire, maintain, and preserve the corporate memory and records of IDRC.

RIMS Division has the lead responsibility for information management within the centre headquarters and Regional Offices. RIMS Division manages the corporate records of IDRC according to government regulations and legislation. As such, RIMS Division's

---

[22] International Development Research Centre, RIMS Division Library Archives, Archiv Archer no. 105571, "Archives and records management: an IDRC 25 year retrospective of support" by Ronald Archer, originally published in *Janus* (1996 – Special – Tunis 1995), p. 110.

key roles are to develop and implement policies that ensure IDRC meets its accountability requirements as a research organization and as a Crown corporation, to provide support material for program initiatives, and to preserve the centre's corporate memory or assets.

The main functions of the Library Archives are to acquire and build a comprehensive permanent collection of materials relating to IDRC activities, and to make IDRC research and corporate information widely available to the public. The Library Archives collections include items produced by IDRC, reports and documents emanating from IDRC research projects and activities, and materials for and about IDRC. The Library Archives accomplishes its main functions through the:

- Identification, catalogue, classification, and index of archival materials in diverse formats in the bibliographic database BIBLIO, and the image database IMAGES
- Creation of microfiche copies of print materials, and preservation of both print material and microfiche
- Dissemination of archival materials through BIBLIO and IMAGES, and special collections accessible via the Internet and/or IDRC intranet
- Processing, archiving, and storing of print materials in the Archives Room
- Capture and archiving of electronic texts on a secure server (RIMS Division Library server)

As a separate yet complementary service, the main functions of the Records Management Unit are to manage the operational, administrative, and historical records of IDRC throughout the record lifecycle, and comply with *The Library and Archives of Canada Act (2004)* (previously *The National Archives of Canada Act (1987)), The Privacy Act (1985), The Access to Information Act (1985), The Emergency Preparedness Act/Emergencies Act (1988), The Security Policy* (The Treasury Board of Canada)

(2002), The Communications Policy of the Government of Canada (The Treasury Board of Canada) *(2002),* and *The Policy on the Management of Government Information* (The Treasury Board of Canada) *(2003)* in order to meet IDRC institutional, legal, and accountability requirements. The RMU accomplishes its main functions through the:

- Coordination, organization, schedule, and disposition of all IDRC official records and files
- Identification, classification, conservation, and IDRC staff retrieval of all records and files according to the corporate-wide Records Classification Framework
- Management of paper files and electronic records using the records management system iRIMS (integrated Recorded Information Management System) adopted by RIMS Division
- Transfer of historical IDRC records to the Library and Archives Canada (LAC) for permanent archiving

The mandate of the RIMS Division to acquire, manage, and make accessible the centre's corporate memory content (project research results and corporate business records, such as IDRC produced publications, outputs from IDRC funded activities, and other documents about or emanating from IDRC and its staff) includes all media types, such as paper, analog, audio/video, images, and digital. In order to meet this mandate, IDRC has identified the need to support electronic publishing, communication, and a strong knowledge base of print, digital, and other materials that reflect the priorities of IDRC and its partners, enhance significant research in all areas of program development, and strengthen the Library Archives current collection. These records have been identified as requiring preservation planning. RIMS Division has proposed that the planning incorporate a digital preservation management strategy that encompasses the physical and virtual preservation of the corporate memory content and access to it, and provides assistance, as outlined in the RIMS Division mandate.

Internship and Methodology

The International Development Research Centre identified the need for the development and implementation of a digital preservation strategy. The RIMS Division is responsible for ensuring that information of enduring value continues to be accessible to IDRC staff, project recipients, and researchers. Digital corporate memory content is managed through electronic information management software that staff, and recipients may access. Researchers and the public may also access unclassified information via the Library Archives online databases. When digital content is inaccessible through these means, physical hard copies may be provided. As part of this responsibility, RIMS Division also makes available expert assistance to staff and researchers utilizing the Library Archives' collection.

Marjorie Whalen, Director of the RIMS Division, proposed a digital archiving project to establish policies and procedures as well as implement a long-term digital preservation management strategy.[23] In her proposal, Whalen outlined the rationale for an internship within the RIMS Division. As RIMS Division has the responsibility for maintaining and providing continuous access to "the accumulated knowledge assets of the Centre," an internship was expected to provide the opportunity for "dedicated thinking to the issues of long-term preservation of digital content."[24] It was thought that a graduate student whose focus is on digital records and preservation would be able to

---

[23] International Development Research Centre, "Marjorie Whalen to Denys Vermette, Memorandum for Intern Proposal, 10 September 2003."
[24] Ibid.

study emerging trends, strategies, and research, and make appropriate recommendations on policy and sustainable procedures for IDRC. Furthermore, the internship would provide the student with practical experience in digital preservation while simultaneously enriching her professional credentials. The objective of my internship (June 2004 to January 2005), therefore, was to recommend a policy and related procedures that would result in a long-term preservation management strategy for IDRC digital objects, but which is not necessarily part of the archival record to be transferred to Library and Archives Canada (LAC).

As the successful candidate for the digital archiving internship, my academic learning in archival studies and experience working in both an archival and records management environment provided me with a strong foundation in issues related to long-term digital preservation management. Most of my internship focused on research into these issues. As part of my research, I had the opportunity to attend an international workshop on preservation management, as well as a facilitated discussion hosted by National Research Council Canada Institute for Scientific and Technical Information (NRC-CISTI) to establish a preservation strategy. The remainder of my internship was dedicated to the development of a report in which I outlined my recommendations for a preservation strategy. This report was given to Marjorie Whalen at the end of my internship. In addition to the report, I gave a presentation that outlined the results of my internship and the recommendations in my report. The presentation was attended by several IDRC staff, including Vice-President of Resources, staff of RIMS Division, the Director and the Manager of Infrastructure Support Information Technology Management Division (ITMD), a member of Bellanet, and other IDRC employees.

As the Centre Intern, I liaised with staff to establish sound archival practices that addressed the policy issues in the digital environment affecting the long-term preservation and management of the centre's corporate and knowledge assets for current and future researchers. During the internship, I advised the centre on practical issues related to preserving the centre's research projects, publications, and other substantive items relating to IDRC activities, while getting practical experience with digital preservation and archival methodology and practices.

Although the focus of my internship was digital preservation, RIMS Division also stressed the importance of making digital resources available to its staff and project recipients situated in the Ottawa headquarters, the Regional Offices, and other locations. Hence, the preservation recommendations I proposed also contain policy-related issues on matters related to access to the Library Archives' digital holdings. Digital archival materials are available through the centre's Internet website (www.idrc.ca) and its intranet or internal website, ECHOnet. The processes for disseminating information through the Internet or ECHOnet will make up a key component of the digital archiving and preservation management strategy, as identified in the original RIMS proposal for this project.

Since this thesis is a case study based upon my internship at IDRC, some personal communication with IDRC staff and others during the internship will be used. I have the approval of IDRC to use this information. As the accompanying bibliography indicates, the research for the preservation strategy is drawn mainly from websites and list serves. This information is easily accessible through the web.

The recommendations to IDRC for a best practices policy for long-term digital preservation management are informed mainly by digital preservation strategies such as migration, emulation, and encapsulation. This information is in publications, research papers, and reports made available by organizations such as the Digital Preservation Coalition, Electronic Resource Preservation and Access Network (ERPANET), and Joint Information Systems Committee (JISC). The literature review in chapter one provides more detailed analysis of their research, findings, and conclusions. In addition, the Cornell University Workshop on Digital Preservation Management and the Canada Institute for Scientific and Technical Information (CISTI) Digital Archiving Research Framework supplemented the theoretical information from consortiums and organizations that focused on high-level strategies and solutions. Consulting with colleagues within IDRC and the larger archival and library fields also provided research information related to the preservation of digital material. Finally, my academic course work and knowledge of electronic records management and archival practice for digital preservation also contributed to achieving the internship goals.

Challenges and Issues

Traditional archival methodology benefits from the fairly straightforward media or technology with which the record was created. The human eye does not require any type of machine in order to read traditional paper records. On the other hand, analogue records, such as audio- and videocassettes (i.e., VHS recordings, reel-to-reel tapes) and film (i.e., 8mm, 16mm, and 35mm film), are dependent on technology to render them readable. Fortunately, this technology has remained relatively stable over time. If the technology is obsolete, it may be rebuilt with minimal difficulty. Conversely, access to digital material[25] is dependent on the software and hardware technology with which it was created, and can only be rendered intelligible by that technology. In other words, for a user to access and understand digital material, the user is wholly dependent upon the technology to render the information in a comprehensible format. Archival methodology and practices, in order to preserve and maintain access to information, are now beginning to create and implement standards that take into account the role of technology.

These issues are made more significant and complex by rapid changes in technology, the proprietary nature[26] of software and hardware, and market-driven forces. Changes in both the software (operating systems, programs, and applications) and hardware (servers, hard drives, disk drives, and related system components) have led to the inaccessibility of digital material as technology becomes obsolete, and quickly replaced by newer, better systems. The systems are market-driven in order to provide enhanced features, such as increased functionality, higher speed, more memory, and so on. These enhanced features, demanded by the user, mean that the older technology is often abandoned by its manufacturer, and eventually no longer receives technical support, as is the case with Microsoft for Windows 95. Closely tied to the market is the proprietary nature of the technology. In order to compete in the market, each manufacturer must develop additional features to entice the user. Over time, these

---

[25] Digital material refers to any information that has either been digitally born, that is, the information was created, stored, and maintained in a digital format, or been created in another format, and digitized. The digitized version may or may not replace the original.

[26] Most commercial software and hardware is proprietary, in that the developers of software programs, such as Microsoft, own the rights and patents to the program applications. For long-term digital preservation management, open access to the encoding for program applications is necessary. An example of an open source application for an office suite (i.e., text document) is Open Office, http://www.openoffice.org/index.html (accessed 02 January 2007).

features may no longer be supported or compatible with other systems. Therefore, the digital information may become permanently lost.

These technology problems have made long-term preservation and access to digital material very difficult. In order to address these problems, collaboration in the field of information sciences (by, for example, archives, libraries, and other cultural heritage institutions) has led to the development of several preservation strategies. However, each of these strategies has advantages and disadvantages, and a viable strategy must be based upon the characteristics and needs of a particular institution. A preservation strategy has been developed for IDRC based on its organizational culture, needs, and resources. This rest of this chapter presents the development of the IDRC digital preservation management strategy.

The basis of the proposed approach to digital preservation at IDRC is the Open Archival Information Systems (OAIS) reference model (ISO standard 14721). Many digital preservation strategies and systems adhere to this model, which in turn allows for shared terminology and processes. Adopting the model's definitions for "long-term" and "long-term preservation" became the foundation for a comprehensive approach to digital preservation that addressed the needs of IDRC. According to the OAIS model, long-term preservation is "the act of maintaining information, in a correct and Independently Understandable form, over the Long Term." "Independently Understandable" refers to "information that has sufficient documentation to allow the information to be understood and used by the user without having to resort to special resources not widely available, including named individuals."[27] Long term is the "period of time long enough for there

---

[27] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, p.1-10.

to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future."[28] Achieving these aims implies that "metadata" (or descriptions of records which permit future retrieval and understanding of them) will be created and preserved, and the ability to render digital information comprehensibly across technological changes will be realized. These implications have been the focus of digital preservation strategies and systems.

While the above dealt specifically with rapid changes in technology and its consequent obsolescence, other challenges relating to digital preservation management can affect an institution's ability to preserve digital objects.[29] Costs, disaster planning, and resource commitment are important factors when planning a preservation strategy. Lack of financial, human, and technological commitment will render any digital preservation management policy and strategy ineffective. Digital preservation experts Anne Kenney and Nancy McGovern strongly urge that the key stakeholders who determine the availability of these resources be identified and that the resources be in place prior to the implementation of a digital preservation management policy and strategy.[30] Disaster planning is a component of any archival preservation policy, be it digital or otherwise, and should ensure that if a disaster occurs, a plan is in place to rescue and save archival records. The centre faces these challenges as well as some that are distinct, such as the preservation, access, and accurate rendering of semantic symbols; the complexity and functionality of the centre's digital records; and the relationship

[28] Ibid., p.1-11.
[29] Preservation Metadata: Implementation Strategies (PREMIS) defines an object as a discrete unit of information in digital form in the *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*, http://www.oclc.org/research/projects/pmwg/premis-final.pdf (accessed 25 May 2005).
[30] Cornell University, *Digital Preservation Management Workshop*, 19-23 July 2004.

between various divisions in regard to web archiving. The following sections discuss these challenges and other issues in greater detail.

Key Stakeholders

RIMS Division is the foremost IRDC stakeholder in long-term digital preservation management as it is responsible for management of digital information. The Division's key role includes providing information resources and services for all staff, delivering tools for the exchange of ideas and information, and ensuring continuous access to the accumulated knowledge assets of the centre. Aside from RIMS Division, other key stakeholders are:

- Authors, Creators, Originators[31]
- Information and Technology Management Division (ITMD)
- Senior Management
- Human and Financial Resources
- Designated User Community

Since preservation management must begin with the creation of the record, the authors, creators, and originators should be made aware of their important role in long-term preservation management of the records. These stakeholders play an essential role in the preservation strategy, and are essential in avoiding loss of data. With this group,

[31] The author or creator of the document may not be the person who electronically files the record. In that case, the person filing the record is referred to as the originator, and indicates him or herself as such in the appropriate metadata field.

early action in the preservation strategy occurs prior to and during the ingest stage.[32] The authors and creators create the corporate memory or assets identified for long-term preservation management and, along with the originator, are also responsible for taking action by electronically filing the documents with the appropriate metadata. This group, therefore, has responsibility for meeting the standards and requirements outlined in the preservation management policy, and must take an active role in maintaining the procedural accountability of a trusted digital repository.

The Information and Management Technology Division's (ITMD) role is to support the use of information technologies that comply with the standards and requirements of the digital preservation management strategy for IDRC. In tandem with RIMS Division, ITMD helps to ensure that the appropriate architecture and infrastructure exists for the ingestion, storage, access, and dissemination of digital material. This includes collaborating with other stakeholders to ensure that the prevailing technology management (hardware, software, and other services) meets both preservation and business processes for the centre, as well as providing adequate expert support for preservation management. As a stakeholder in the preservation management strategy, ITMD is responsible for the establishment and development of the technological architecture and infrastructure necessary to meet the preservation goals of RIMS Division policies.

Senior management is another key stakeholder in the digital preservation management strategy. Its role falls under the managerial and administrative functions of the OAIS model in which senior management is responsible for vetting the preservation

[32] Ingest is the term used in the OAIS reference model, and is synonymous with submission or acquisition. For the authors, creators, and originators, the ingest stage would be the act of electronically filing a record or saving it in the BIBLIO database or iRIMS.

strategy, as well as ensuring that the centre has the appropriate legal status, and human, financial and technical resources to implement a preservation strategy. Senior management, therefore, must demonstrate the commitment necessary for a successful and viable long-term digital preservation management strategy.

There must be adequate human and financial resources to sustain a viable preservation strategy. A staff member may be hired to manage the preservation processes and strategies, or the responsibilities may be divided among existing staff in one or more divisions or units. Funding is needed for staff and equipment, training, and other expenses as required. The hiring and funding of staff and equipment may be divided among various units, departments, or divisions within an institution.

The designated user community, along with RIMS Division, is another important key stakeholder. The designated user community is made up of staff, project recipients, and researchers (who may also be members of the general public). In the OAIS model, the designated community is "an identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities."[33] A Consumer is defined by the "role played by those persons, or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAIS repositories, as well as internal OAIS persons or systems."[34] In the case of IDRC, the designated user community expands upon the OAIS definition to include the persons or systems responsible for submitting information to the Library Archives and the RMU,

---

[33] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, p.1-10.
[34] Ibid., p.1-8.

as well as requesting and accessing information.[35] The designated user community plays an active role in driving the application of technology for the creation, use, and dissemination of information generated and received by IDRC. Hence, the designated user community may have a stronger influence over the digital preservation strategy than some of the other stakeholders. Many stakeholders are involved in the long-term digital preservation management strategy, and over time roles and responsibilities may change to reflect new priorities and modifications in the preservation management strategy.

Appraisal

The ability to capture digital material exceeds the ability to manage and preserve the information over a long period. This creates a twofold effect in that unneeded and duplicated information is continually being saved on media via software that is no longer supported or accessible at the current time. Without a proper appraisal policy, information is retained needlessly at the expense of preserving information of enduring value, while at the same time increasing the human, technical, and financial resources to preserve and render the information. As the costs of storage decrease, and the amount of space increases in terms of gigabytes, terabytes, and petabytes,[36] the tendency for many organizations is to preserve more and more data than is necessary. The downside of this strategy occurs as the amount of data becomes harder to manage, organize, describe, and index for successful retrieval and dissemination. RIMS Division has resolved this

---

[35] The OAIS model refers to the persons or client systems that provide information to be preserved as the Producer (p.1-12). For the simplicity of this report, the Designated Community and Producer have been merged into the common group of designated user community.

[36] A gigabyte is roughly equal to 1,024 megabytes (2^{30} bytes) or the equivalent of a pickup truck filled with books. A terabyte is 1,024 gigabytes, or one trillion bytes. An academic research library is approximately two terabytes. A petabyte is 1,024 terabytes, or one quadrillion bytes. All the American academic research libraries would equal to two petabytes. For more examples, see http://searchstorage.techtarget.com/sDefinition/0,,sid5_gci944596,00.html (accessed 4 May 2006).

challenge by implementing an appropriate appraisal policy that scales the collection according to the research value of the information. Hence, the digital preservation architecture will be in line with the development of the technical and managerial challenges in accessioning, managing, and providing access to digital materials.

Within the RIMS Division, appraisal determines what materials have research value and should be incorporated into the centre's Library Archives collection. Thus, appraisal assigns the value, significance or importance to records, and identifies records that require long-term or indefinite preservation.[37] Furthermore, appraisal decisions should be technology-neutral and not driven by the medium of the record, but apply the appraisal criteria across all records. This includes paper, digital, audio, visual, maps, and drawings, and any combination thereof. RIMS Division plays an important and *active* role in shaping the collective memory and legacy of IDRC. Consequently, an outline of the appraisal criteria RIMS Division uses is essential in understanding IDRC's functions and projects as these relate to the centre's corporate memory and the interactions between the centre and the general public, staff, project recipients, and researchers.

In the *Guidelines for Library Archives and Records Management Research Information Management Service Division (RIMSD) (Draft)*, the Library Archives acquires and builds the centre's permanent collection of materials relating to IDRC's activities. The records selected for the Library Archives include:

- IDRC produced publications, including books, journals and newsletters, published and unpublished;

[37] Terry Cook, "Appraisal Methodology: Macro-Appraisal and Functional Analysis – Part A: Concepts and Theory", (Library and Archives Canada, Summer 2000), http://www.collectionscanada.ca/information-management/061101_e.html (accessed 24 November 2004).

- Documents emanating from IDRC such as consultants' reports relating to research, evaluation reports and items describing IDRC activities, projects, programs and services;
- Items directly resulting from IDRC funding such as the scientific outputs of final technical reports, and other outputs such as conference or workshop reports and proceedings, theses and dissertations, and research reports by centre interns and awardees;
- Items authored by IDRC staff and governors, dealing with a subject related to international development, and written or published during the individual's tenure with IDRC, in both IDRC produced and non-IDRC produced publications; these may be speeches, journal articles or essays/chapters in books;
- Substantive items describing the work or history of IDRC or related to its programs or projects, but not produced by IDRC (these may be journal or newspaper articles about IDRC or its projects).[38]

RIMS Division has also identified records that are outside the scope of the Library Archives collection criteria. These include interim or progress reports, items remotely related to IDRC and its activities, items issued before or after an individual's term at IDRC, and book reviews written by IDRC staff or governors, unless the book is about IDRC. Items written by IDRC staff on a leave of absence have selective retention, and are only deposited in the Library Archives if the subject directly relates to IDRC research interests.

---

[38] International Development Research Centre, RIMS Division, *Guidelines for Library Archives and Records Management Research Information Management Service Division (RIMSD) (Draft)*, p.3.

The types of document described above are those that are needed to facilitate the centre's research mandate, but as a federal government body IDRC records are subject to the *Library and Archives of Canada Act*.[39] This means that any IDRC documents that are of long-term interest to the Canadian government and people, are identified under the Records Retention and Disposal Authority, and thereby transferred to LAC.[40] In the appraisal report of the Disposition Authority, two groups of records were identified as having archival value and were to be transferred to LAC.[41] The first group includes the registry subject files consisting of studies, research analyses, correspondence, and information and supporting documentation relating to projects. The records are arranged according to the program areas (as described in the organizational structure of IDRC) and are used to assess the value and success of similar, prospective projects. These records have a selective retention in which a sample of the files is to be transferred to LAC for permanent retention. According to the appraisal report, the regional and divisional project files, described here, represent a significant percentage of the funding allocated for the aid programmes that are supervised through one of IDRC's six regional offices. These files allow comparison of the smaller operational projects (under $100,000.00) and the major projects (over $100,000.00). The major project files is the second group of records that is identified for transfer to LAC. This group of records consists of three key reports for each of the major projects: *Project Summary, Project Completion Report* and *Final Report.* These records document the history of major and divisional projects

[39] See http://laws.justice.gc.ca/en/L-7.7/249401.html for the full act (accessed 02 January 2007).

[40] Library and Archives Canada, "Archival Appraisal Report on the Records of the International Development Research Centre, 15 April 1991," (ATIP request 29 March 2006).

[41] In order to comply with Library and Archives Canada's Disposition Authority, RIMS Division requires at least two copies of the documents – one for transfer to Library and Archives Canada, and another for the centre's permanent collection.

funded and supported by IDRC. The three key reports summarize the activities of each project, and provide information on the success of the project, the grant recipient, the project's relevant impact, and the value of the project to future initiatives. The Office of the Secretary and the General Counsel of IDRC require these files for ongoing reference in funding and supporting similar projects, and therefore, the files have great value, and are transferred to LAC for permanent retention. Overall, these two groups of "records have archival value because they document Canadian initiation of – and participation in – a unique vehicle for development aid delivery."[42]

Costs

Many organizations have identified the need for a scalable and workable financial solution for digital preservation that outlines the scope of the initial investment and ongoing costs; however, cost figures for small and medium sized institutions have yet to be determined.[43] Shelby Sanett of Amigos Library Services, Inc. and ERPANET have provided a cost analysis framework and tools, respectively. According to Sanett, her cost modeling predicts high overall costs that extend over a long period of time for acquisition and preservation-related activities, as well as access-related activities.[44] ERPANET also places emphasis on the initial investment and ongoing costs of a sustainable preservation

---

[42] Ibid.

[43] Michèle V. Cloonan and Shelby Sanett, "Comparing Preservation Strategies and Practices for Electronic Records", http://www.rlg.org/events/pres-2000/cloonan.html (accessed 05 October 2004). Cloonan and Sanett estimate that the digital preservation costs for a national institution may range from as low as $10,000 to $2.6 million US, per year. See also http://www.rlg.org/en/page.php?Page_ID=245 (updated 02 January 2006).

[44] Shelby Sanett, "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks", *RLG DigiNews*, vol. 7, no. 4, http://www.rlg.org/legacy/preserv/diginews/diginews7-4.html (accessed 09 November 2004).

strategy. In addition, this group also recognizes that information technology is still in its infancy, and that this affects preservation policies and investment decisions.[45]

Activities and costs related to digital preservation are reflected in the capital costs, direct operating costs, and indirect operating costs for the institution. When overlapped with the OAIS model, the costs can therefore be divided into three activities – ingest (or acquisition/appraisal), storage, and dissemination. Sanett refers to these activities as costs for preserving electronic records (initial costs, maintenance of the digital preservation architecture, and acquisition), costs for use (associated with perpetual storage), and user populations (dissemination). Sanett also outlines areas or activities in which costs would be incurred. She lists software development, equipment, hardware for preservation processing, research and development, facilities, and interface design for processing electronic records as Capital Costs. Appraisal, intellectual property issues/rights (or rights management) and payment of royalties, acquisition, processing, metadata development, preservation strategy implementation, physical storage space, maintenance (migration and/or refreshment), monitoring and evaluating, disposition of records, and staff expertise and time fall under direct operating costs. Overhead or indirect operating costs relate to expenses incurred through staff supervision, clerical support, benefits, training and unallocated work time, rent or lease of building space (office, onsite and offsite storage), utilities, amortization of capital costs, and general and administrative costs (human resources, operational expenses, funding development and grant writing, professional development, partnerships with other institutions, and policy development). The overhead costs are generally subsumed in the overall operational and

[45] ERPANET, *erpaguidance Cost Orientation Tool*, http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf, p.2 (accessed 09 November 2004).

administrative budget of the institution, but these costs should be calculated, as resources are still required. Most importantly, resources should be dedicated and committed to digital preservation prior to the implementation of the policy to ensure continued preservation and access.

While the capital costs for preserving digital records are incurred at the beginning, these costs may be spread out over a period of time, such as five years, which may also represent the replacement period of software and hardware, and should be calculated into the present value of the records. Both indirect and direct operating costs should also be brought to the present value,[46] and when combined with the capital costs, this would reflect the total cost of preserving the digital records. Other yearly costs are incurred through the operating costs related to the use of preserved digital records, and these costs, when combined with the total preservation cost, over the period of amortization represents the total present value for preservation and use of digital records. Sanett also analyzes costs using an activity-driven evaluative methodology. The process to determine the costs for acquisition, preservation, and access is as follows:

- Calculations for determining capital, indirect, direct, and total costs for preserving digital records remain the same as described above;
- Costs for institutional use/external access of preserved digital records are incurred on a yearly basis, and include costs of equipment, software, user training, facilities, interface design, storage, royalties, communications, access mechanisms, staff resources (for monitoring, appraisal and authentication following migration cycles, disposal, user query response and services,

[46] Sanett defines the present value as the current value of a sum of money expected to be received in the future.

records management), and indirect staff, amortization, and general and administrative costs;

- The total cost for acquisition, preservation, and access equals the total cost of preservation plus the total cost of institutional use/external access.[47]

The *erpaguidance* Cost Orientation Tool publication provides tabular information outlining the factors or issues, cost impact, and considerations for assessing the resources required for digital preservation.[48] The cost orientation tool is divided into seven sections – objects; people; standards; practices; systems, methods and technologies; laws and policies; and organization. Within the objects, factors include influence on creation, existing objects, complexity, preservation period, and appraisal (and/or value). Costs are related to the degree of influence of the archivist at the time of creation, the implementation of standards, the use of proprietary formats, insufficient contextual metadata, increased complexity of the record (particularly multimedia and dynamic records), the length of the retention period, and the application of sound appraisal practices. Costs related to people or human resources involve employing persons with qualified skills and experience, and training and professional development to update and maintain the skill sets for the job requirements (staff capacity). Well-trained, knowledgeable staff may require a higher salary than those unfamiliar with digital preservation, but over the long-term expert staff will reduce costs of poorly implemented policies and practices, minimize the risk of loss due to inadvertent error, and remain current with information technology trends. Implementing standards and best practices,

---

[47] Sanett, "The Cost to Preserve Authentic Electronic Records in Perpetuity."
[48] ERPANET, *erpaguidance Cost Orientation Tool*, pp.3-6.

in addition to workflow analysis, regular auditing of operations and processes, helps to lower costs and facilitate long-term planning and management of digital objects. Overall costs arise from the organizational culture with regard to workflow, operation and processes. Automation of work will help to decrease costs related to manual labour; however, consistency in the workflow remains important to maximize planning. Standards lower the overall effort, and facilitate long-term solutions. Systems, methods and technologies are the greatest factors related to cost. This section of the table includes the preservation method, validation of methods, sustainability, and digital preservation architecture (such as components, maintenance, operation, and facilities). Multiple preservation methods may be implemented, and each method has its own cost-related issues. Validation of the method is also necessary prior to implementation.

After the method (or methods) has been chosen, the sustainability of the method across several generations of technology is also essential for reducing costs. Encasing the preservation is the digital preservation architecture. The costs related to the architecture include maintenance and operation of the technical infrastructure, scalability (or flexibility), security and disaster planning, and level of functionality. A simple digital preservation architecture is less costly; however, the ability to preserve complex digital records with high functionality may be compromised, or the simpler architecture may lack the technical infrastructure of a rapidly changing IT environment. The final two factors, laws and policies, and organization, focus on rights management (intellectual rights, copyright, retention legislation, and privacy legislation) and the organization's commitment of financial, human and technological resources, as well as collaborative research and development. Proper records management should minimize the organization's expenses of preserving unnecessary records, breaches in privacy and other legal implications. Capacity and relationship building encourages cooperation and collaboration amongst colleagues, both external and internal. Hence, costs related to research and development may be shared, thereby reducing costs of having to "re-invent the wheel" multiple times.[49]

The ERPANET cost orientation tool attempts to provide a better view of costs, while distinguishing between the resources needed at the macro- and micro-level for each task in the digital preservation process. In summary, Sanett's and ERPANET's descriptions of costs share some features:

- Costs include all activities from the creation and capture of the digital objects to their disposal (either destruction, transfer to another institution or long-term preservation);
- Preservation is ongoing and active for the whole retention period of the object;
- Preservation is typically a long-term venture that needs to be financially sustainable.[50]

Disaster and Emergency Recovery Plan

A disaster and emergency recovery plan is necessary to minimize the impact of natural disasters such as flood, fire, or some other environmental or technological catastrophe. The first step should identify the risks and list the hazards that could jeopardize the building and collection of both non-digital and digital records. Once the hazards have been specified, a program or plan should be devised to decrease these risks.

---

[49] Chapter four discusses the importance of collaborative research and development within the broader scope of the digital preservation recommendations presented to IDRC.
[50] ERPANET, *erpaguidance Cost Orientation Tool,* p.2.

The disaster plan addresses the impact of natural and technological disasters, but does not eliminate their occurrences. Hence, a response plan should be developed so that a recovery strategy can be initiated. The emergency recovery plan sets the priorities of staff responsibilities during salvage and rehabilitation operations, staff access to the building and collection, and designates the priority or order in which the material in the collection should be rescued. A well-developed disaster and emergency plan, based upon sound principles, will alleviate the severity of the impact of disasters and incidents.

At this time, RIMS Division, in consultation with other IDRC divisions, is creating a disaster and emergency recovery policy and plan for implementation in the near future. As a key aspect of a digital preservation management strategy, preparedness in the eventuality of a disaster will help to minimize the risk of loss of information and costs related to recovery or rebuilding data and information. This includes the identification of vital records, which helps to ensure that those records necessary for the continuing operation of the institution (e.g., legal papers, reports and minutes of shareholders or board of governors, and in the case of IDRC, research material) are documented as high priority concerns so that business may resume with minimal loss or interruption.[51]

In addition, IDRC should ensure that the disaster and emergency recovery plan differs from the general policy of regular back-ups of servers, and the use of off-site storage. When backing up information, the tapes are recycled, and following each recycling period, the previous information is lost, and cannot be recovered. Off-site storage is used for records with longer retention periods, and records transferred to the

---

[51] Judith Read-Smith, Mary Lea Ginn, and Norman F. Kallaus, *Records Management, 7th Edition* (South-Western Thomson Learning, Inc., Mason OH, 2002), p.5.

LAC, as well as master microfiche copies of research and historical material retained by the RIMS Division Library Archives. It should be noted that off-site storage may be a component of the disaster and emergency recovery plan, but should not be considered a substitute for proper planning and policy in the event of a catastrophe.

Semantic Symbols

In promoting research in developing countries and communities, IDRC accepts research outputs in one of three official languages – English, French, and Spanish.[52] In addition, the IDRC Web Content Management System is implementing an Open Archival Initiative for grey- or pre-print literature, as well as other research information and material. Furthermore, some of this information may be written in a language other than English, French, or Spanish. As such, research literature, outputs, and material will be created in a growing number of languages and submitted to RIMS Division for long-term preservation. In time, the need to maintain and make accessible information in a number of languages will pose a unique long-term preservation challenge for the RIMS Division.

Character encoding in software determines how a character is rendered and displayed. In the early years of computer programming, each manufacturer programmed or encoded characters differently, making interoperability between software vendors difficult. The American Standard Code of Information Interchange (ASCII) (ISO 646) is the most basic text encoding format and the root of English language textual coding as well as some non-English languages. The inability of ASCII to properly render characters or semantic symbols of non-European languages such as Arabic, Chinese or

[52] The Library Archives also accepts archival materials in other languages such as Chinese, Indonesian, and Tagalog (Philippines) along with several other languages.

Korean has led to the development of a new Universal Code Set known as Unicode (ISO standard 10646). Unicode is a standardized set of characters that can accommodate a growing number of languages not included in the ASCII format. As Jürgen Bettels and F. Avery Bishop point out, for digital material, "Unicode represents a strategic direction in internationalization technology" for programming software that correctly renders data in most of the world's languages.[53] With the emphasis on international development and developing countries and communities, IDRC may benefit from implementing Unicode as a basic text format for the long-term preservation of research material written in non-European languages.

Web Archiving

IDRC's website contains information of enduring value to IDRC that requires long-term preservation management. Library and Archives Canada, among many other cultural memory institutions, has initiated projects on the acquisition and archiving of websites. One of the most successful efforts has been undertaken by the Internet Archive (http://www.archive.org), whose mandate is to capture and archive the entire Internet. The National Library of Australia has also developed a digital archiving system called PANDAS, which has been included in a number of key studies. Other leading experts in web archiving have included United Kingdom initiatives such as the Networked European Deposit Library (NEDLIB) and programmes in Denmark and Switzerland.

Web archiving activity has challenges similar to other digital preservation activities. The Library and Archives Canada pilot project for capturing and archiving

[53] Jürgen Bettels and F. Avery Bishop, *Unicode: A Universal Character Code*, http://python.planetmirror.com/ftp/macro/DEC/DTJ/DTJB02/DTJB02PF.PDF (accessed 16 March 2005).

websites revealed that rights management issues, software tools for capturing websites and converting proprietary formats into HTML, and technological issues relating to security and interoperability led to preservation problems in that some websites either could not be captured and/or adequately archived.[54] Unfortunately, due to these and other issues, LAC was unable to archive and make accessible the websites captured during its pilot project. Other barriers affecting the success of archiving websites relate to human and technological resources – the lack of infrastructure and dedicated staff time for capturing the large representative sample of Canadian websites.

In learning from LAC's experience, IDRC's Information Management and Technology and RIMS Divisions should ensure that a scalable and flexible infrastructure is developed and that existing web search crawlers and harvesters are tested to find those that best suit IDRC needs. Unlike LAC and countries in Europe (such as Denmark and Switzerland) whose mandate is to capture a substantial portion of the Internet, IDRC is only concerned with capturing the centre's website. Hence, some of the challenges related to the use of proprietary formats, password protected sites, rights management, and the "deep web" may be minimized. With respect to rights management and the deep web, IDRC policy on web archiving will need to address whether sites external to the centre shall be harvested, and if so copyright permissions will need to be obtained.

Library and Archives Canada tested SuperBot, Web Whacker, and Grab-a-Site as potential crawlers; however, it found these crawlers were unable to meet the requirements of the ambitious scope of their pilot project. The Library of Congress examined web crawlers that were open source, had documented prior use, and had an active community

---

[54] Susan Haigh and Roselyn Lilleniit, "A Strategy for Archiving Websites at the Library and Archives Canada," copy of paper presented at the *Preservation of Electronic Records Symposium 2003.*

of developers and users.[55] Other criteria, although not mentioned by the Library of Congress, should include a robust and flexible crawler with the ability to do broad or selective crawls. Jennifer L. Marill and her colleagues found that the open source NEDLIB Harvester and Heritrix adequately handled the tasks of large-scale harvesting for long-term preservation.[56] The NEDLIB Harvester has been developed and used by a European consortium of national libraries for long-term preservation; however, the harvester lacks the flexibility needed for complex permission environments and a direct user interface. Heritrix, developed by the Internet Archive, was released to the public in January 2004, and shows promise as a crawler. Separate access tools are required for presentation of sites harvested by both Heritrix and NEDLIB. Fortunately, the NWA tool set for NEDLIB is also compatible with Heritrix output. If after testing existing web crawlers, none are found suitable for harvesting IDRC websites, another option is the development of an in-house crawler.

Following the decision to select a web crawler, the method of harvesting needs to be determined. Netarchive.dk, a project for the preservation of Denmark's cultural heritage on the Internet, recommends using three different methods – automatic "snapshot," "selective," and "event-based" harvesting.[57] Snapshot harvesting refers to a static image in time, like a photographic snapshot. Selective harvesting captures pre-determined, or selected, websites; whereas, event-based harvesting focuses on the capture of special events, such as the inauguration of a new program initiative. In selective

---

[55] Jennifer L. Marill, Andrew Boyko, Michael Ashenfelder and Laura Graham, *Tools and Techniques for Harvesting the World Wide Web,* (Library of Congress), http://csdl.computer.org/comp/proceedings/jcdl/2004/2493/00/24930403.pdf (accessed 23 December 2004).

[56] More information relating to the NEDLIB Harvester is available at http://www.csc.fi/sovellus/nedlib/ and for Heritrix, please see http://crawler.arachive.org.

[57] Netarchive.dk, *About Netarchive.dk – 2nd Phase*, http://www.netarchive.dk/ (accessed 16 July 2004).

harvesting, particular sites are closely monitored, then captured following a modification to the site. Lars Clausen analyzed the viability and reliability of etags and last-modified-date fields as indicators of changed Web content.[58] His results demonstrated 99.94% accuracy in predicting a change in the Web content when the etag is missing and the last-modified-date is either missing or has changed. Clausen also demonstrated 63.3% accuracy of predicting no changes in the content under the same harvesting conditions.[59] Simply put, when the etag is missing and the last-modified-date is missing or changed, there is greater accuracy for the web crawler to predict a change in website, and therefore capture the website for permanent retention. If the etag and last-modified-date fields are used to predict no change in the website, the web crawler is more likely to miss any changes, and thus not capture a modified website. This approach is best when further combined with use of the "If-None-Match" and "If-Modified-Since" headers as the number of redundant downloads can be "reduced between one-half and two-thirds without missing significantly many changes."[60]

Metadata support and context are as important as the selection of a crawler and the method of harvesting. Metadata support has been developed through the Open Archives Initiative Protocol Metadata Harvesting (OAI-PMH) that can be used for accurate web content gathering of complex objects.[61] Herbert Van de Sompel and his colleagues have proposed an expanded view of OAI-PMH that supports not only the harvesting of metadata but also the resource harvesting of complex objects. Content,

---

[58] Lars R. Clausen, *Concerning Etags and Datestamps*, http://www.netarchive.dk/website/publications/Etags-2004.pdf (accessed 23 December 2004). The article is also available on an updated website, http://www.iwaw.net/04/Clausen.pdf (accessed 03 January 2007).
[59] Ibid., p.11.
[60] Ibid., p.12.
[61] Herbert Van de Sompel, Michael L. Nelson, Carl Lagoze, and Simeon Warner, "Resource Harvesting within the OAI-PMH Framework", *D-Lib Magazine*, vol. 10, no. 12 (2004), http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html (accessed 23 December 2004).

therefore, is cleanly mapped based upon an OAIS perspective of content transfer.[62] Included with the OAI-PMH framework is the capture of existing metadata, but not always the context of the website. The context of the web resources may be determined by the site itself, a subset of resources within the website, or a single site or document.[63] IDRC should define the boundaries of their web resources for effective and efficient long-term digital preservation management. The centre's preservation policy addresses four levels of context:

- As a stand-alone object, ignoring all hyperlinks;
- Within local context, considering the hyperlinks into and out of the website;
- As a set of linked web pages;
- As an entity in the broader technical and organizational context.[64]

Several sources relating to the archiving and preservation of websites are available through the Internet. Some websites include Digital Preservation Coalition (www.dpconline.org/graphics/), D-Lib Magazine (www.dlib.org/), JISC (www.jisc.ac.uk/), and Cornell University (www.library.cornell.edu/preservation/). Other resources may result from research and development, and collaboration with institutions such as LAC and the Canadian Institute for Scientific and Technical Information (CISTI). In addition, subscriptions to list serves such as WEB-ARCHIVE or DIGLIB can provide more practical information regarding web archiving.

---

[62] Ibid.

[63] Anne R. Kenney, Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze, and Sandra Payette, "Preservation Risk Management for Web Resources", *D-Lib Magazine*, vol. 8, no. 1 (January 2002), http://www.dlib.org/dlib/january02/kenney/01kenney.html (accessed 04 June 2004).

[64] Ibid.

Towards Developing a Digital Preservation Strategy

This chapter provides an overview of the digital preservation goals and challenges facing IDRC. As such, these challenges inform the technological approach taken in developing a digital preservation strategy for the centre. Chapter three, therefore, examines the technological component of digital preservation, and provides some recommendations for preserving digital records.

CHAPTER THREE

Digital Preservation Management Strategy Recommendations

Chapter three moves from the organizational context of the digital preservation strategy to the technical recommendations that were presented in the final report to RIMS Division. In this chapter, preservation strategies, file formats, technical architecture (or infrastructure), and metadata schemas are presented in greater detail. This includes recommendations, such as the proper use of media and storage of digital objects, to aid and improve current RIMS Division practices. The technical recommendations also examine the importance of metadata in digital preservation. Lastly, this chapter concludes with a *Final Report Post-Evaluation* that reflects upon my knowledge and experience, and how this formed my recommendations to the RIMS Division.

Digital Preservation Management Strategy Final Report: Overview

As part of my internship, I wrote a final report for Marjorie Whalen, the Director of RIMS Division and presented the report to IDRC staff, including Denys Vermette (Vice-President of Resources), Terry Gavin (Director of Infrastructure Support Unit, Information and Technology Management Division, ITMD), Allen Sinfield (Manager of Infrastructure Support Unit, ITMD), Todd Graham of Bellanet, and several other RIMS Division and IDRC staff. The recommendations were based upon research that I conducted during my internship, as well as the Cornell University Digital Preservation workshop that I attended and the Canada Institute for Scientific and Technical Information (CISTI) Digital Archiving Research Framework in which I participated. In both the final report and the presentation, I outlined recommendations for a digital preservation strategy, which included migration and conversion, a trusted digital repository, a systems architecture, and metadata.

The digital preservation strategy should be considered a roadmap for the RIMS Division in managing and preserving their digital records over the long term. This strategy, as outlined in the following sections of this chapter, was chosen for several reasons. The recommendations complied with current standards identified by other archival institutions that are developing and implementing a digital preservation strategy. One such standard is the Open Archival Information System (OAIS) reference model, which provides a blueprint or high-level model of a digital archives; yet is still technology neutral in that the model also applies to a traditional archives (as described in chapter one). Other standards include the use of open standard file formats for preservation, implementation of migration and conversion cycles to accommodate changes in technology, and the use of appropriate metadata for the management and preservation of IDRC digital records. Furthermore, the recommendations in the strategy were chosen because they align with the policies and standards being implemented by Library and Archives Canada. As IDRC is a federal Crown corporation, it is subject to the *Library and Archives of Canada Act*, and therefore is required to transfer records of permanent, historical value to the Archives as per the applicable records schedules. Additionally, many of the RIMS Division practices coincide with other studies that explore the activities and status of other archival institutions in Canada. In one such study, the findings were consistent with the best practices in the RIMS Division. These practices involve a technology-neutral approach to appraisal, acquiring digital objects or records in a variety of media, the use of digitization to make traditional or paper-based records accessible online, and managing the growth of storage space for the objects.[1] While the findings of the Provincial Archives of Alberta Digital Preservation Survey were reported nearly fourteen months after the completion of my internship, the survey summary illustrates how proactive the RIMS Division has been in its management of digital objects.

Recommendations

To implement any digital preservation strategy, it is important to understand that long-term digital preservation management requires a well-balanced organizational infrastructure, technological infrastructure, and a resources framework in order to build a trusted digital repository.[2] Anne R. Kenney and Nancy Y. McGovern describe the working relationship of the organizational infrastructure, technical infrastructure, and policy framework (resources) as a three-legged stool that supports digital preservation. The three-legged stool forms the foundation of the digital repository that ensures the authenticity, reliability, and integrity of the digital objects. When these elements are properly combined, institutions are able to establish trusted repositories that contain both a management and a preservation component. This allows for accountability through records management while simultaneously providing the mechanisms needed to ensure that digital objects will be renderable (that is readable) through time.

---

[1] Lori Podolsky Nordland, "Provincial Archives of Alberta's Digital Preservation Survey Summary", *Archives Society of Alberta Newsletter*, vol. 25, no. 2 (Winter 2006), http://www.archivesalberta.org/media/images/pdfs/Winter_Newslet_%202006.pdf (accessed 07 September 2006).

[2] Anne R. Kenney and Nancy Y. McGovern, *The Three-Legged Stool: Institutional Response to Digital Preservation.*

The organizational infrastructure expresses itself through a comprehensive policy framework; whereas, the technological infrastructure entails a flexible preservation plan that can incorporate changing technological platforms over time. A resources framework focuses on the staffing, technological, operational, and other costs associated with maintaining the organizational and technological infrastructure of a trusted digital repository. These are the core requirements necessary for the successful long-term preservation management of digital information.

In the final report to the RIMS Division, the technical infrastructure is based upon the *de facto* standard of converting digital documents into a recommended file format that is then migrated and refreshed every two to three years. The native, or original, file should be stored on an offline server and the current migrated version should be kept online for easy accessibility. Acceptable file formats for textual documents include PDF (or PDF/A), ASCII, and X-HTML; images should be in TIFF or PNG format; audio/visual files may be saved in WAV or MPEG. An integrated RAID (Redundant Array of Inexpensive Disks) server is one possibility for the storage architecture. However, storage media such as digital versatile disc, compact disc, floppy disk or similar media should be avoided as these storage devices are at high risk for technological obsolescence and deterioration from environmental hazards. Digital information should be stored and managed in a trusted digital repository that is OAIS compliant. The trusted digital repository should also contain sufficient metadata for authenticity, fixity, provenance, search, and retrieval of the digital resources. The metadata should also be available in a textual format such as XML, ASCII or UNICODE, with XML and ASCII as the preferred choice. Furthermore, the technical infrastructure of the digital preservation management strategy should be scalable to allow for future changes, as well as accommodate new user and system needs as these develop.

RIMS Division has several policies and procedures, or guidelines, and several other practices in place for managing IDRC corporate records. These policies and procedures should form the backbone of the organizational infrastructure of the digital preservation strategy; however, RIMS Division may need to update or develop new policies for activities such as the submission and acceptance of digital objects, as per the standards being identified and endorsed within the digital preservation community. In addition, RIMS Division has an appraisal policy for library materials and for determining official corporate IDRC records. When applied to the archival environment, these policies will help to ensure that the digital resources designated as having long-term value are properly captured, stored, and managed in the trusted digital repository. Finally, to achieve that critical success, the recommendations need to be supported through senior management, and have adequate human, financial, and technical resources committed to long-term preservation management so that once the digital records are adequately captured they can be preserved and made accessible over the long term.

Digital Preservation Strategies

Digital preservation has been described as the series of managed activities for ensuring continued access to digital objects.[3] A digital preservation strategy, in its broadest sense, refers not only to the series of managed activities, but also the policies, frameworks, and procedures that mould and give direction to the activities. In the general

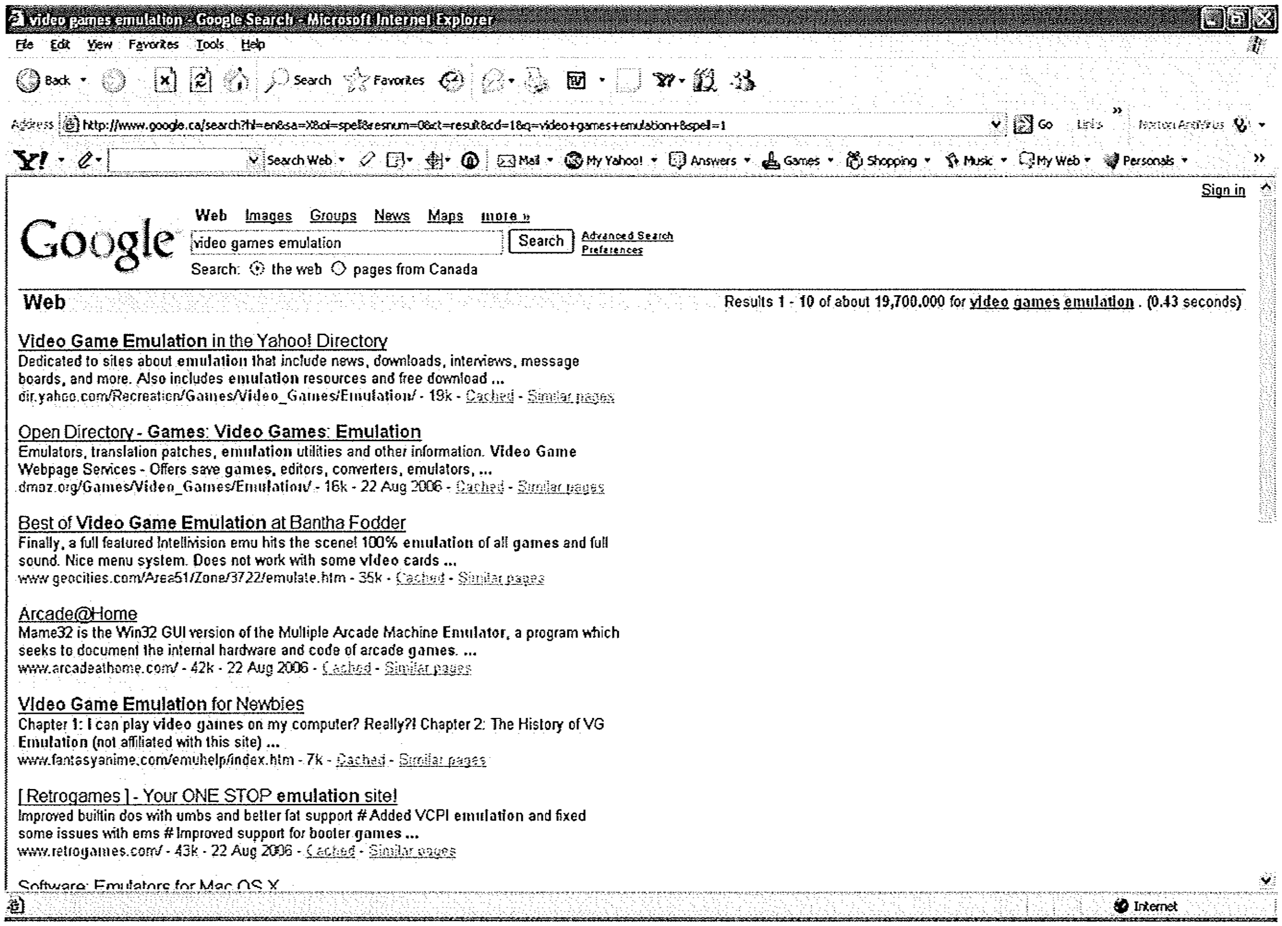
[3] Jones and Beagrie, *Preservation Management of Digital Materials.*

archival literature, a digital preservation strategy commonly refers to one of three approaches to managing digital objects beyond the lifespan of the technology that created the object. These three approaches are emulation, encapsulation, and migration. Each approach has been implemented to varying degrees, and each approach has its pros and cons, as discussed below.

Emulation has probably had its greatest success in the video game industry, in emulating games such as Pac Man, Donkey Kong, Frogger, and other similar games from the eighties and nineties. A quick Google search using key words "video game emulation" returns nearly 20 million hits. Included are sites that offer free emulation software that can be downloaded onto your PC, patches, updates, and new versions of old games, and of course, the list goes on. The image below is a snapshot of the Google search described above.[4]

[4] The search was done 23 August 2006.

Figure 4: Hits Received from a "Video Game Emulation" Key Word Search on Google



Emulation works by imitating other software programs or applications, as well as hardware systems, such as printers. In respect to digital preservation, emulation is used to overcome technological obsolescence by imitating obsolete hardware and software. Jeff Rothenberg, a proponent of the emulation approach, argues that emulation is the only reliable method "to recreate a digital document's original functionality, look, and feel."[5]

---

[5] Jeff Rothenberg, *Avoiding the Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* (Council on Library and Information Resources, Washington DC, January 1999), http://www.clir.org/PUBS/reports/rothenberg/pub77.pdf#search=%22%22Avoiding%20the%20Technological%20Quicksand%22%20%22digital%20preservation%22%20emulation%20Jeff%20Rothenberg%22, p.17 (accessed 23 August 2006). Another article that looks at the applicability of emulation as a digital

This allows the original software to be run on future, unknown systems, despite the original software being obsolete. Rothenberg outlines three pieces of information that need to be recorded, and four steps to be performed in order for emulation to work as a preservation strategy.[6] The three kinds of information are: 1) the digital object itself and the original software in which the object was created (including the operating system); 2) the specifications of the emulator for the object's original computing platform; and 3) the explanatory information such as metadata, provenance, and documentation for the software and (emulated) hardware. To preserve the digital objects, Rothenberg identifies four sequential steps: annotate, encapsulate, transliterate, and emulate. The first step, annotate, refers to the creation of any contextual information, or annotations, required to open and use the encapsulated information described above. In the second step, the three types of information are encapsulated with the digital object for preservation management. Step three, which Rothenberg has termed "transliterate," refers to the media refreshing cycles when the encapsulated information is copied to the new media. The last step is testing and using the emulator by running it on future systems. According to Rothenberg, these sequences of events should allow obsolete digital objects to be rendered on any future computer system.[7]

One of the best-known and successful examples of digital preservation using the emulation approach is the CAMiLEON (Creative Archiving at Michigan & Leeds:

---

preservation strategy is Stewart Granger, "Emulation as a Digital Preservation Strategy" in *D-Lib Magazine*, vol. 6, no. 10 (October 2000), http://www.dlib.org/dlib/october00/granger/10granger.html (accessed 23 August 2006).

[6] Rothenberg, *Avoiding the Technological Quicksand.*

[7] David Bearman critiques Rothenberg's approach in Bearman's article, "Reality and Chimeras in the Preservation of Electronic Records" in *D-Lib Magazine*, vol. 5, no. 4 (April 1999), http://www.dlib.org/dlib/april99/bearman/04bearman.html (accessed 23 August 2006).

Emulating the Old on the New) project through the Universities of Leeds and Michigan.[8] In this project, Margaret Hedstrom (University of Michigan Project Director) with Chris Rusbridge and Paul Wheatley (University of Leeds Project Director and Project Manager, respectively) worked together to investigate emulation as a means to retain the original functionality and "look and feel" of digital objects. The BBC Doomsday, a landmark multimedia resource to celebrate the 900th anniversary of the original Doomsday Book as well as to capture information about British life in the 1980s, was a key proof of concept test case for the CAMiLEON project. The project enabled Hedstrom, Rusbridge, and Wheatley to study the complexity of interactive objects and peripheral devices, and the necessary resources related to emulation. Over a four-year period (October 1999 to September 2003), the two teams were able to develop software to support the functionality of the BBC Doomsday and three other similar projects (Ecodisc, Volcanoes, and Countryside). Hedstrom, Rusbridge, and Wheatley emulated the Doomsday project using the "DomesEM" demonstrator system, an emulator developed to support the original program's complex multimedia functionality. However, the challenges of intellectual property rights, particularly copyright, were not solved with the emulation strategy. Access to the emulated version of the project may be subject to copyright, and therefore, access to the BBC Doomsday project may be restricted. While the project's bits and bytes have been digitally preserved, rights management could impact the manner in which people are able to access the wonderful information captured and organized by the BBC Doomsday.

Aside from the copyright or rights management issues, other disadvantages have been identified in using emulation as a digital preservation strategy. While Hedstrom,

---

[8] CAMiLEON, http://www.si.umich.edu/CAMILEON/index.html (accessed 23 August 2006).

Rusbridge, and Wheatley were able to emulate the functionality of the Doomsday project, this may not always be the case. The success of emulation depends upon the complexity of the digital object, which may mean that only part of its functionality or look and feel can be recreated. In addition, emulators are subject to technological obsolescence, and therefore, new emulators need to be built as changes in software and hardware occur. These changes can have a twofold effect – not only is the technology that document creators use changing, but also the emulators themselves need to be updated to be usable on new computers. With each new paradigm shift in technology (e.g., floppy to compact disks, or WordStar to Microsoft Word), a new emulator must be built to recreate the digital objects. If the emulator runs on hardware that undergoes a technological change (e.g., a Commodore 64 to IBM), the emulator itself has become obsolete, and a new one needs to be built. Finally, there must be rigorous documentation of the hardware and software requirements, including any coding modifications developed to assist in the operations of the software and hardware. Consequently, emulation can be a resource intensive approach to digital preservation.

Another digital preservation strategy is encapsulation. Encapsulation groups together the necessary information required to preserve a record. The information may include metadata, viewers or readers, and other types of discrete files. Like emulation, encapsulation has also been a successful strategy used for digital preservation. In Australia, the State of Victoria government has implemented the Victorian Electronic Records Strategy, or VERS, that consists of a framework of standards, guidelines, and consultancy and implementation projects "centered around the goal of reliably and authentically archiving electronic records."[9] The foundation of VERS is the VERS Encapsulated Object (VEO) format for capturing, managing, and preserving digital objects. The VEO was first piloted between 1999 and 2002, as part of the VERS@DOI (Department of Infrastructure) Pilot Implementation Project. In this pilot project, an encapsulated object was formed using XML (eXtensible mark-up language), which conformed to the VERS metadata schema. The digital object is converted to one of the preservation file formats (PDF/A, JPEG, JPEG-2000, MPEG-4) identified in the VERS Standard (Version 2). Together, the digital object and metadata are wrapped to create the VEO. The VEO is managed in the Public Record Office Victoria (PROV) Digital Archives, which became operational in August 2005.

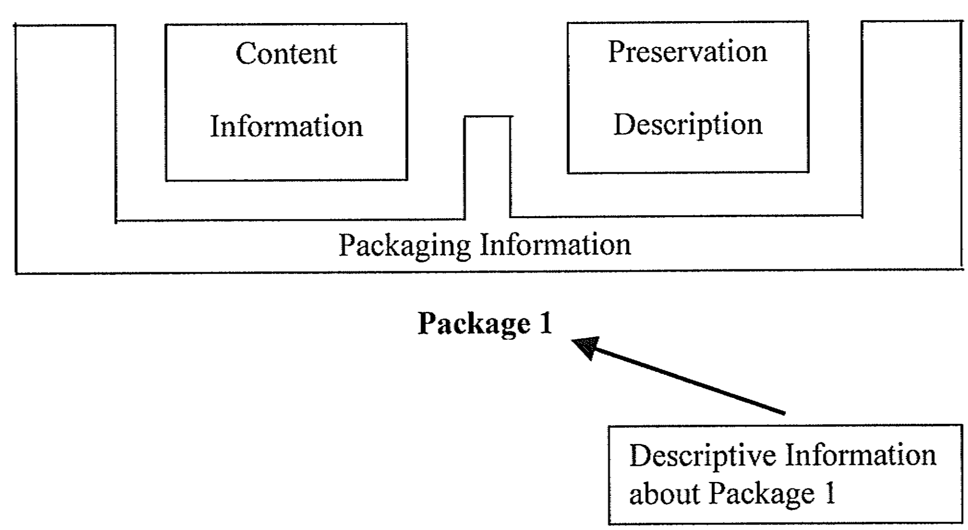
Encapsulated objects are represented in the OAIS reference model (described in chapter one). The information package, defined as content information and associated preservation description information to aid in preservation, is wrapped within the packaging information.[10] This is a conceptual container of the record, its associated representation information, the information required to preserve the record, and the information that "binds" the content information and preservation description information together (see Figure 5).[11] According to the OAIS reference model, "The Content Information and PDI [Preservation Description Information] are viewed as being encapsulated and identifiable by the Packaging Information."[12] The descriptive information about the package may be linked to the information package, or included within the encapsulated object. In the OAIS reference model, a link between the

---

[9] Victorian Electronic Records Strategy (VERS), http://www.prov.vic.gov.au/vers/vers/default.htm (accessed 01 September 2006).
[10] Consultative Committee for Space Data Systems, p. 1-11.
[11] Ibid., p.2-5.
[12] Ibid.

conceptual container (or package), and the descriptive information is suggested by the use of the arrow (see Figure 5). In this case, linking the descriptive information to the encapsulated object allows for multiple objects to share descriptive information. When this approach is taken, the descriptive information may be stored in a registry, such as those used for metadata. In addition to managing the encapsulated object within the digital archives, the links between the description information and object must also be maintained to ensure access to the files. The descriptive information may also be included in the encapsulated object, as mentioned above. This produces larger files since the descriptive information would require duplication for each related file. Since encapsulated objects maintain all the required information for the digital object's long-term management and preservation as one object, resources are more focused on the elements and information that is essential for access and preservation.

Figure 5: Information Package Concepts and Relationships



As a digital preservation strategy, encapsulation is subject to technological obsolescence, and therefore requires some form of long-term management similar to emulation. In addition, metadata may be difficult to update and maintain, particularly if it is wrapped within the encapsulated object. Any descriptive information, such as the Rosetta Digital Stone, that is wrapped with the object requires an analogue or human-readable copy that includes information on how to access the document and the outer layer of the encapsulated object.[13] Even if links to metadata and software are created, there is no guarantee that these links will effectively and efficiently maintain access to and preserve the object. While the encapsulated approach allows for more flexibility in managing the digital object, this approach may be described more as a secondary preservation strategy that would be implemented in conjunction with emulation.[14]

One of the most common methods of managing and preserving digital objects includes a combination of migration and conversion. Migration is the *de facto* standard for most archival institutions. According to the Provincial Archives of Alberta's Digital Preservation Strategy Survey, most of the participating archival and library institutions are implementing some form of migration and conversion as part of their digital preservation strategy.[15] Procedures for simple migration are well established, and as

[13] The Rosetta Digital Stone contains descriptive information required to interpret digital content. This information is kept separate from the object to avoid duplication of information and for more efficient use of storage space. For more information see Alan R. Heminger and Steven B. Robertson's article, "Digital Rosetta Stone: A Conceptual Model for Maintaining Long-term Access to Digital Documents," http://www.ercim.org/publication/ws-proceedings/DELOS6/rosetta.pdf (accessed 07 September 2006).
[14] Jones and Beagrie, *Preservation Management of Digital Materials,* and see also National Library of the Netherlands' Digital Preservation Officer, Hilde van Wijngaarden's presentation entitled "Different Approaches to Digital Preservation," http://rdd.sub.unigoettingen.de/conferences/ipres04/van_wijngaarden/Different%20approaches%20to%20digital%20preservation.pdf#search=%22Encapsulation%20disadvantages%20%22digital%20preservation%22%22 (accessed 07 September 2006).
[15] Lori Podolsky Nordland, "Provincial Archives of Alberta Digital Preservation Survey: Technology – Part I", *Archives Society of Alberta Newsletter*, vol. 25, no. 3 (Spring 2006),

technology advances and the number of platforms diminishes, migration may increase in its ability to preserve the functionality of complex or dynamic documents while simultaneously decreasing the required resources and time. However, this is not to say that migration is the "silver-bullet" for digital preservation. Written policies, procedures, and guidelines need to be developed, quality checks need to be performed, migration procedures require rigorous documentation, and preservation metadata should be captured to ensure that digital objects are properly migrated and converted.

Migration, as defined by Maggie Jones and Neil Beagrie in the *Preservation Management of Digital Materials: A Handbook*, is:

> A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology.[16]

The proposed IDRC digital preservation management strategy is based upon migration and conversion. To date, this strategy proves to be an effective method of preserving the digital record while maintaining access to the information as technology advances from one system or generation to the next. Migration of simple formats, when done automatically, is also cost-effective as procedures are well established in this area, and may reduce the expense of manual labour. When choosing migration as the preservation strategy, a migration path needs to be available. If not, additional costs may

---

http://www.archivesalberta.org/media/images/pdfs/spring_2006_newsletter.pdf (accessed 07 September 2006).

[16] Jones and Beagrie, *Preservation Management of Digital Materials,* p.12. Conversion of file refers to changing the format of a file. For example, converting a Word document to a PDF. Migration is "updating" the file version, in which a file is migrated from Word 95 to Word 2000.

occur in migrating the record. It is, therefore, recommended that the chosen preservation file format should be interoperable with other file formats when choosing commercially available software or, if creating an in-house application, that a migration path be incorporated. Migration should be done every two to three years, or if a major change in software is occurring, a migration cycle should be run.

The original or native document should be retained, as the possibility of future technology being able to accurately and reliably render the document may exist. In the interim, an access or research copy should be created and saved in one of the recommended long-term preservation file formats, such as TIFF, PDF/A, or ASCII, so that existing and future interpretation tools may continue to render the files. A current trend for conversion and migration of records is to separate the intellectual content from the presentation.[17] This means that for textual documents, for example, the content is converted to a format such as ASCII or UNICODE to maintain full-text searchability; whereas, the "look-and-feel" of the document is presented in a TIFF format. The user sees the presentation or TIFF format on the screen, but searches the content copy. Thus, three copies are stored, maintained, and preserved in this scenario – the content copy, the presentation copy, and the native or original document. Consequently, this approach allows for better search results, and minimizes the number or frequency of erroneous hits. Since storage is relatively inexpensive, a PDF/A copy may also be kept. A PDF file is full-text searchable if the document is created using a PDF editor or printed using a PDF driver such as cutePDF®.[18] If a document is scanned using imaging software, then saved

---

[17] Personal Communication, Kyle Brentnell, President of Reuel Inc., an Internet solutions company, and Intermediate Systems Analyst (IT) with RIMS Division (2004).
[18] cutePDF is free software for creating PDF documents from Word and other file formats. The website is accessible at http://www.cutepdf.com/ (accessed 14 September 2006)

as a PDF image, the PDF is not searchable unless OCR (optical character recognition) software is applied.

Following the migration of files, a checksum[19] or a similar computed value should be completed to verify that the migration has not corrupted any files. This procedure can also be used to certify that the file has maintained its authenticity, integrity, and reliability. In addition, the results of the migration should be recorded in the file's metadata.

Recommended File Formats

Recommended file formats for long-term digital preservation are based upon selection criteria developed by international experts from the fields of archival studies and library sciences. The criterion emphasizes file formats that are open source or specification, are a standard (such as ISO), are ubiquitous, have demonstrated stability, are interoperable with other software and hardware, and viable. Sustainable file formats and standards are two aspects of a comprehensive long-term digital preservation management strategy. Since file formats are subject to technological and software obsolescence, the choice of file formats for both creation and preservation has implications for any long-term preservation strategy. When considering the file format and standards, this choice is determined by the functional requirements of the file and its creating process. For example, textual files, as compared to images or audio/video files, have different functions, and therefore, require different formats and standards for preservation. As well, files can range from basic text to complex, highly interactive web

[19] A checksum is an algorithmic value used to check transmitted data for errors.

resources. The basic text file, such as a Word document, can be managed and preserved for long-term accessibility in a straightforward manner. More sophisticated resources that fully exploit available technology require additional preservation management processes than what is needed with the simple text document. In other words, the more complex the file, the harder the file is to manage for long-term digital preservation.

When proposing choices for submission to the trusted digital repository, a minimal number of file formats will reduce the cost and complexity of long-term digital preservation management. These criteria should be implemented during data creation; however, the criteria can be applied if converting files to a preservation file format some time after the file has been created. File conversion to a managed and sustainable format in this way is expensive and complex, and generally less successful than when implemented at the creation stage. Furthermore, better control over the file helps to ensure its authenticity and reliability within a legal context and from an archival perspective.

The following provides a guideline for choosing file format and standards for long-term digital preservation management. This is general advice and not meant to be exhaustive of the various file formats in existence at this time. In my final report to IDRC, I described a range of criteria to assist RIMS Division to make informed choices about file format issues. The selection criteria and issues for sustainable file format and standards are divided into ten sections (see below). The last three criteria should be considered if implementing a migration strategy for any or all files. Whichever strategy or combination of strategies is implemented, a minimal number file formats will reduce the cost and complexity of long-term digital preservation management.

1. **Open Standards**

   File formats, in which the technical specification has been made available in the public domain, are considered to be open. Access to the detailed technical information about the file format is required for preservation management. Examples of open formats include JPEG and ASCII. In addition, formats may be proprietary or non-proprietary. Some proprietary formats, such as PDF, have published the technical specifications for collaboration with other software developers. Non-proprietary formats, such as ASCII, can result in some loss in structure, context, and functionality; however, the range and sophistication of open standard formats is increasing, and their use is strongly recommended.

2. **Ubiquity**

   Ubiquity refers to the use of widely available popular formats with broader and longer-lasting support from software suppliers than software created for a limited clientele. Popular formats are also more widely supported by a number of software programs, and therefore are preferred.

3. **Stability**

   The format specification should not be subject to frequent or significant changes. New versions should be backwards compatible.

4. **Metadata Support**

   File formats that link to metadata support are preferable. The metadata may be generated automatically by the creating application, entered by the user, or a combination of both. Metadata has value for the active use of the file, and for its long-term digital preservation management. Metadata provides information on the file's provenance and technical properties. Its value depends both upon the degree of support provided by the software environment used to create the file, and the extent to which externally-stored metadata is used in its place. Metadata also assists in recovering files created on obsolete systems, which then can be rendered by current computer systems.

5. **Feature Set**

   The features and functionality of formats should be selected which support the required purpose or corporate/business processes. Over-specified formats should be avoided at these formats are complex, and expensive to manage and preserve.

6. **Interoperability**

   Formats that are supported by a wide range of software or are platform-independent are considered to be interoperable. Furthermore, the ability to exchange files supports long-term sustainability of the data by facilitating the migration of the data from one technical environment to another.

7. **Viability**

   Error detection functions such as Checksum, Cyclic Redundancy Check (CRC), or other error correcting abilities search for file corruption that may occur during transmission, migration or at some other event. Formats that provide these error checking and correcting facilities are more viable, and therefore preferable.

8. Authenticity

Authenticity must be demonstrated for various legal, administrative, and historical reasons. When selecting a file format, the format must preserve the content (data and structure) of the file, and any inherent contextual, provenance, referencing, and fixity information.

9. Processability

Certain types of data must retain their processability, or functionality, to have any reuse value. For example, processability would include the retention of hyperlinks in a document after it has been converted to its long-term preservation format. Authenticity, however, requires that the file not be altered through reprocessing. To maintain authenticity, for example when a text document is converted to image-only PDF file, a processable version of the file should also be maintained.[20]

10. Presentation

The formatting of a file may have significant information value. If the authenticity of a file requires the preservation of its original "look and feel" (such as fonts, colours, and layout), then the ability of a file format to support this through migration will be a crucial consideration.[21]

---

[20] Processability refers to the ability to retain the dynamic features of a document, such as hyperlinks, or raw data contained in databases. In some cases, a static document may be disseminated to a user, and in this case a formatted copy is released (see also page 100 of this chapter, footnote 30). Processability does not refer to the processes involved when migrating a document, but is one of the criteria to consider if migration or conversion is being used as a digital preservation strategy.

[21] The ten selection criteria were developed by Adrian Brown of the National Archives (UK), and published (19 June 2003) in the National Archives' *Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation*, http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf (accessed 18 November 2004). These criteria are reproduced with permission by the UK National Archives.

The above criteria are important in selecting a file format for preservation. Yet this is only one aspect of the digital preservation strategy as other requirements may also affect the archives' choices. According to William G. LeFurgy, Digital Initiatives Project Manager for the Library of Congress Office of Strategic Initiatives, National Digital Information Infrastructure and Preservation Program's (NDIIPP) Project Management Office, a digital preservation management strategy must also address these three basic requirements:

1. The needs of the document producers. Files must be easy to create, compatible with workflow processes, and flexible enough to include images, sub-documents, and other components.

2. The needs of document users. Files must be reliable, appropriately functional, and discoverable from different approaches (e.g., index terms and full text).

3. The needs of archives. Files must be based on non-proprietary and stable technology and suitable for recordkeeping guidelines issued to producers (e.g., guidance for records creation, maintenance, and disposition). In addition, files must support metadata for access, provenance, and preservation.[22]

For textual documents, files should be converted to PDF, X-HTML, or ASCII. Each of these preservation formats has its own advantages and disadvantages for preservation, and therefore, the decision on which format to choose should be based upon the requirements of the file, and the needs of the users and the archives. However, long-term digital preservation management of text files may encompass the use of various formats, and several options exist. The three more common formats – ASCII-based, PDF, and X-HTML – will be analyzed. Nonetheless, other formats may provide better preservation management and accessibility in the future, and these should be regularly

[22] William G. LeFurgy, "PDF/A: A new digital preservation format", *Society of American Archivists Newsletter* (May 2003), http://www.archivists.org/saagroups/gov/newsletters/May2003.asp (accessed 18 November 2004).

reviewed and/or considered. For example, Adobe has created a PDF/A format that has just recently been approved as an international standard for all textual documents requiring long-term preservation (ISO 19005, as explained below).

ASCII

Advantages	*Disadvantages*
- non-proprietary	- loses the "look and feel"
- open source (access to the technical specifications)	- loses any functionality and complexity
- not compressed	- compromises the authenticity of the record
- not encrypted	

The American Standard Code for Information Interchange (ASCII) is the most basic text-encoding format. Developed by the American National Standards Institute, this format is the root of English language textual coding as well as some non-English languages. Furthermore, this format is supported by various software applications and suppliers, which increases ASCII interoperability and ubiquity. ASCII is considered to a best practice choice when only the intellectual content is required for long-term digital preservation.

PDF

Advantages	*Disadvantages*
- open specification	- proprietary
- able to preserve some functionality of a file or digital object	- includes features such as encryption
- normalization, widely deployed, and integrated into several work environments	- complex subset of features and functionality
- backward compatibility	- loss of processability or renders the file different from the original, consequently compromising the authenticity and

	integrity of the file or digital object
- longevity and trustworthiness of the file format	- cannot be used for complex files fully utilizing available technology

PDF/A

Advantages	*Disadvantages*
- ISO 19005	- some elements may be subject to copyright or patents
- self-contained and self-documenting	- cannot contain audio/visual content in the file
- mandated metadata is based upon standards, uses XMP	- loss of functionality still exists
- encryption is disallowed	- logical structure of the document must be incorporated at creation
- suitable for long-term preservation of page-orientated textual (or primarily) documents	

Although PDF is a proprietary file format, it has proven stability, longevity, and is considered "trustworthy." PDF has had only five major revisions since is initial release, retained excellent backward compatibility, and has remained an open-source product. In such cases as this, file formats that are both open-source and proprietary may be a successful compromise for a digital preservation management strategy. Although PDF tends to be the default archival format, it does have limitations. PDF/A was developed to address these issues. Just recently, PDF/A has become a recognized international standard (ISO 19005) for both the exchange of digital information and for the long-term preservation of digital objects.[23] It was developed for archival and records management purposes with a specified and limited, stable subset of PDF. Thus, PDF/A excludes any

[23] International Organization for Standardization, "New ISO standard will ensure long life of PDF documents", http://www.iso.org/iso/en/commcentre/pressreleases/archives/2005/Ref974.html (accessed 07 September 2006).

features that are considered to be incompatible for archival viability, such as encryption. PDF/A also includes extensible metadata platform (XMP) as part of its standard. XMP metadata would be "embedded in each file as plain text, which both lessens the possibility of loss and simplifies access to the metadata."[24] This format addresses most of the digital preservation requirements as outlined by William G. LeFurgy (see above), including the ability to retain simple functions, and some significant properties that are critical to preserve the authenticity and integrity of the file.

X-HTML

Advantages	*Disadvantages*
- open source, non-proprietary	- some loss in the level of complexity
- history of stability, longevity, and ubiquity	- SGML is not adaptable for all Internet environments
- example of Standard Generalized Markup Language (SGML) application that is ISO Standard 8879	- interoperability problems as used for a specialized market
- regarded as a standard publishing language of the Internet (World Wide Web)	- over-specified format that is complex and costly to manage and preserve
- limited subset that is "user-friendly", including increased flexibility to the combinations of new and existing feature sets	
- allows for alternate ways of accessing the Internet	

Extensible Hypertext Markup Language (X-HTML) is XML based and designed to work in conjunction with XML-based user agents. X-HTML documents can be viewed, edited, and validated using standard XML tools. In addition, X-HTML documents can utilize applications, such as scripts and applets, which rely upon either the HTML Document Object Model or the XML Document Object Model. Furthermore, this

[24] LeFurgy, *PDF/A: A new digital preservation format.*

file format is widely used for web-based documents, and content developers can utilize the XML family, with the benefits of backward and future compatibility for any and all content. XML retains the power and flexibility of Standard Generalized Markup Language (SGML) without its complexity.

HTML is an example of a language defined in SGML, and used in electronic document exchange, document management, and document publishing. HTML addressed the problem of SGML complexity by specifying a small set of structural and semantic tags suitable for authoring relatively simple documents. In addition to simplifying the document structure, HTML added support for hypertext, and later included multimedia capabilities. The rapid invention of new elements for use within HTML has led to inoperability problems as highly specialized markets have adapted HTML file formats.

For image documents, PNG and TIFF are recommended as both support lossless compression. File formats that have lossy compression should be avoided, as data can be lost when converting from the native format into the preservation format. However, if a lossy compression format, such as JPEG, is the native format, then the archivist should maintain the image in this format, and migrate and refresh as outlined in the preservation strategy.

Cornell University Library has also developed several additional recommendations for long-term digital preservation management of image files.[25] These recommendations build upon the selection criteria and issues outlined at the beginning of

[25] Anne R. Kenney, Oya R. Reiger, et al., *Report of the Digital Preservation Policy Working Group on Establishing a Central Depository for Preserving Digital Image Collections* (Cornell University, March 2001), http://www.library.cornell.edu/preservation/IMLS/image_deposit_guidelines.pdf (accessed 18 November 2004).

this section. File formats can vary in terms of resolution, bit-depth, colour capabilities, support for compression, and metadata. In regard to compression, the preference is for uncompressed files or files with lossless compression, where no information is discarded in the compression technique. Files in which the least significant information is averaged or discarded are described as being lossy. In addition, other technical considerations include enhancement or image processing, web support, and quality control. In addition, long-term viability of images necessitates consistent capture, and that these methods of capture are well documented.

TIFF

Advantages	*Disadvantages*
- well established with widespread use leading to increased ubiquity	- no major developments since purchased by Adobe in 1992 (has remained dormant)
- supported by a wide range of processing applications	- very difficult to create a fully TIFF compliant reader, resulting in the appearance of TIFF sub-standards
- highly flexible and platform-independent	- supports compression algorithms that are used by proprietary, licensed, and patented software
- ideal use for information exchange between prepress applications	- compression files may be lossy
- *de facto* standard	- strictly used for bitmap data, does not contain text or vector data
- lossless image	- not compatible with JPEG files, especially for prepress use
	- cannot contain more than four GB of data

At the time that Cornell University Library published its report on preserving digital image collections, no clear archival format was recommended although preference has been given to Tagged Image File Format (TIFF) files 5.0 and 6.0.[26] Since then, the recommended imaging requirements are now the required guidelines for depositing

[26] Ibid.

digital images in the Cornell University Library's repository. Furthermore, Library and Archives Canada has implemented TIFF as the preservation format for its image collection.

PNG

Advantages
- open source
- efficient, patent-free, lossless compression
- support for metadata
- replacement for GIF format, and potential TIFF substitute
- gaining support among the leading browsers, including W3C and IETF
- expected to be released as ISO/IEC International Standard 15948
- allows colour profile information to be stored within the image, enabling accurate colour reproductions

Disadvantages
- does not support animation
- not as flexible as other file formats, such as JPEG 2000
- some applications that write PNG images are subject to lossy compression
- not well supported by Adobe Photoshop and Internet Explorer for Windows

The Portable Group Network Graphics (PNG) file format was developed to replace the Graphics Interchange Format (GIF) and TIFF. PNG has improved upon many of the qualities of GIF including its functionality and avoidance of incorporating licensed compression technology in the file format. In addition, PNG provides better quality control for accurate colour reproduction, brightness (gamma correction), and transparency for use in graphic design. The metadata set is simple, yet can be expanded as necessary if more structured metadata is required in the future.

JPEG 2000

Advantages	*Disadvantages*
- open specification	- coding is an international standard, and can be purchased through the ISO and ITU-T websites
- ISO/IEC Standard 15444, and ITU Standard T.800	- common JPEG supports lossy compression
- most popular image for the Internet	- lossless version (JPEG-LS) is rarely used
- efficient, license-free wavelet compression	- several different version with different feature sets
	- JPEG-2000 includes some patented technology
	- simple metadata support (however, blocks of text may be added, and mapping to other metadata standards, such as Dublin Core, is being done)

The Joint Photographic Experts Group (JPEG) developed the common JPEG file format in the late 1980s. Later the JPEG-2000 file format was introduced as a replacement for the JPEG/JFIF format commonly used on the Internet. The JPEG-2000 contains patented technology and features; however, a baseline version has been created using license- or royalty-free technology. This baseline version utilizes lossy compression. The lossy compression is more efficient than other common compressions, but at the expense of the quality of the image resulting in some visible distortion. Support for the JPEG-2000 file format continues to be intermittent, yet improvement has been noted. However, plug-ins are still required for most applications, such as Adobe's Photoshop and JASC's Paint Shop Pro, and support remains limited with Internet Explorer and Netscape Navigator. Nonetheless, following the standardization of JPEG-2000, this may change and new versions of software applications may offer built-in support.

Audio/visual digital preservation management has consisted of transferring or copying analog tapes onto a more accessible digital medium. For files that are digitally born, and for which an analog copy does not exist, these digital files remain highly susceptible to software and technological obsolescence. In addition, audio/visual files are highly complex contributing to their increased difficulty for long-term digital preservation. Currently, WAV and MPEG are the two recommended file formats for preservation. Only MPEG is an ISO standard, but WAV is recommended by Electronic Resource Preservation and Access Network (ERPANET). These file formats have been developed to help minimize the risk of information loss; however, technological monitoring for audio/visual files should be stressed as audio/video files are relatively new for file creation at IDRC, and use of these files is increasing.

Carl Fleischhauer, project coordinator for the Office of Strategic Initiatives of the Library of Congress, identifies four issues relating to the preservation management of digital audio/visual files.[27] These are "selecting the target format for reformatting, determining the quality of the reformatted copy, shaping the information package and the importance of metadata, and analyzing longevity in a "media-less" environment."[28] In selecting a target format, disclosure, adoption, transparency, self-documentation, "fidelity", and sound field support are six important factors in choosing the bitstream structure and file type. As noted with text and image file formats, these six factors focus on open-standard or open-specification, interoperability, and metadata support criteria. Fidelity examines the file format's support for high resolution of sound; whereas, sound

[27] Carl Fleischhauer, "The Library of Congress Digital Audio Preservation Prototyping Project" presented at the *Sound Savings: Preserving Audio Collections Proceedings from the Symposium*, http://www.arl.org/preserv/sound_savings_proceedings/fleischhauer.html (accessed 24 August 2004).

[28] Ibid.

field support determines whether the format represents stereo or even surround sound. The second and third issues focus on reformatting and copying analog files to a digital format; however, the fourth issue, longevity of file formats, has greater application as a guideline for long-term digital preservation management strategy through the adoption of the Open Archival Information System (OAIS) reference model. Currently, the Library of Congress has adopted Microsoft's WAV file format for its ability to meet the adoption, disclosure, and fidelity tests (Macintosh users have adopted Audio Interchange File Format to equal effect).

WAVE (AIFF for Macintosh users)

Advantages	*Disadvantages*
- standard file format for use on IBM-compatible PCs for both pro-audio and multimedia productions - does not compress files and there is no loss of information - recommended by erpaNET	- AIFF is less frequently used as an archival formation because files are very large

MPEG

Advantages	*Disadvantages*
- ISO/IEC Standard 11172 - MPEG is a popular format for DVD, establishing widespread use - supported by other software suppliers	- proprietary - allows for data compression - several versions of the format are available

The Motion Pictures Experts Group (MPEG) developed a set of compression standards for moving images (such as film, video, and animation). As a result, many versions of full MPEG-1 audiovisual players are available, and are capable of encoding and editing audio/visual files. In addition, MPEG will include Scalable Video Coding that allows for high scalability features and high compression. At this time, an archival quality format does not exist; however, as the format is well established and supported, a standardized format may be developed.

These file formats are examples that have been recommended for long-term digital preservation of electronic images. Unlike textual documents, as discussed above, images may utilize more complex technology, resulting in a greater degree of difficulty in rendering the various software created for specific uses of images. Audio/visual formats consist of the both the bits that comprise the image and the header information on how to read and interpret the file.

The number of file formats that meet both corporate and archival criteria should be minimized, and data creation should be, whenever possible, restricted to these formats. When selecting file formats, the selection criteria and issues, mentioned above, help to identify the necessary standards needed to meet long-term digital preservation management. In addition, regular monitoring of technology for changes in file formats should also be part of the preservation strategy. The development of new standards, obsolescence of existing formats, and changes in proprietorship, for example, will impact the choice of a file format, or require a migration to new file format to ensure long-term digital preservation.

After choosing the file format(s), continued monitoring should occur to detect any errors, or loss of information over time. If the viability of the format fails, the authenticity, reliability, and integrity of the file are questioned, and the preservation strategy is unsuccessful. Related to viability of the format is accessibility. If the file format lacks interoperability with other systems and applications, the file is inaccessible, and the preservation strategy, once again, fails. In short, file formats "must meet the requirements for both preservation of authenticity and ease of access."[29] Furthermore, the National Archives (UK) notes that:

> There is also a subtly different conflict between the need for *processable* and *formatted* data formats. From a preservation and re-use perspective, data must be maintained in a processable form. For the purposes of access, however, control of the formatting may well be the most important criteria, and in some cases it may actually be desirable for the data not be processable by end users. In some cases it may only be possible to reconcile these differences by using different formats for preservation and presentation purposes.[30]

Finally, the choice of the file format should meet the needs and requirements of IDRC. The functional requirements of the file and the creating processes affect the choice of a preservation format that is to be used. In addition, RIMS Division should not overlook the costs of implementation in their decision on file formats for long-term digital preservation management.

Within the RIMS Division's Library Archives, several formats have been identified for textual, image, and audio/visual files. At the time of my internship, RIMS Division accepted Adobe Portable Document Format (PDF), Microsoft Word, PowerPoint (and slideshow), Excel, WordPerfect, text file, and rich text for textual documents. Reports and other records that are scanned into the Library database, BIBLIO, are kept in PDF, and those documents that are submitted electronically are maintained in their original or native format. Graphic and web-based formats, such as hypertext Cascading Style Sheet and HTML, are also housed in the Library Archives.

---

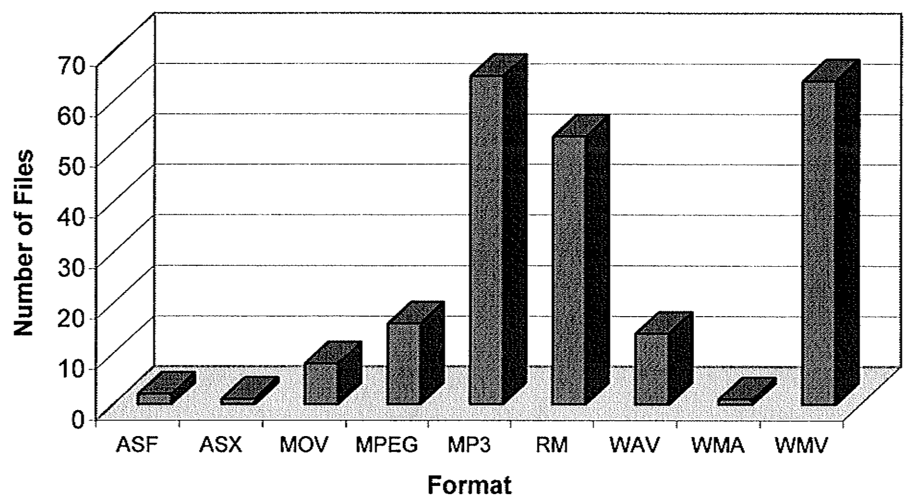
[29] Brown, *Digital Preservation Guidance Note 1.*
[30] Ibid.

The IMAGES database contains JPG files, but the Library Archives also includes images in TIFF, PNG, and GIF. In addition to these formats, the Library server contains audio and video files in Advanced Streaming Format (ASF) and Advanced Stream Redirector (ASX), QuickTime (MOV), Real Media (RM), Waveform Audio (WAV), and Windows Media Audio and Video Files (WMV and WMA, respectively).[31] Digital audio and video files are less common than textual documents and images; however, the number of a/v files will increase quickly as more information is captured in this type of media. In October 2004, most of the a/v files were in MP3 format, even though WAV is considered more appropriate for preservation and long-term management (see Chart 1). Overall, the Library Archives holds mostly textual documents and images, with a smaller percentage in audio and video formats. While many of these formats are proprietary, and all are subject to technological obsolescence, RIMS Division will either need to look towards a more open source format, or implement a migration and conversion strategy, should they wish to continue to access these formats over the long-term.

---

[31] Description of the formats was current to 07 October 2004.

Chart 1: Distribution of A/V Formats



Digital Preservation Management Architecture

The digital preservation management architecture is the system and processes for managing the digital material that is submitted for long-term digital preservation. The architecture supports the digital collection from acquisition to storage, and finally to dissemination, which includes integrated access to digital and print resources. In order to do so, persistent and unique identifiers, a metadata registry, and a hierarchical data storage system or model are necessary for filing, organizing, and retrieving digital material. These components should be interoperable, and based upon standards such as ISO 15489 (Records Management) and the United States Department of Defence 5015.2.[32] Presently, most digital preservation management architectures continue to evolve, and new recommendations and systems are being developed.

---

[32] International Organization for Standardization, *Information and Records Management Standard 15489*, http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=31908&ICS1=1&ICS2=140&ICS3=20 (see also http://cio.doe.gov/RBManagement/Records/PDF/n01-19-

At the core of the digital preservation architecture is an established repository that is reliable and has sufficient storage. According to David Vernon and Oya Reiger of Cornell University, once the storage requirements exceed five terabytes, a hierarchical storage system should be considered.[33] This approach implements a fully redundant storage system to manage a wide variety of workloads in a dynamic environment, which in addition allows for fast storage and retrieval.[34] The National Library of Australia, in addition to above, selected a hierarchal model for its ability to reflect "the complex nature of some collection items" in which several documents form a series or collection when combined together, as is the case for IDRC project initiatives.[35] However, two drawbacks exist for using a redundant array of inexpensive disks storage system: the short life expectancy of three to five years, and that the cost of long-term storage can be higher than initially expected.[36] In regard to sufficient storage, this should be scalable and flexible to meet the requirements of the preservation management architecture and strategy. In other words, additional space should be easily added without disrupting the database management and records already stored in the repository.

Aside from the servers for iRIMS, and the BIBLIO and IMAGES databases, the Library Archives stores its digital information on gold-layered compact discs.[37] While CDs and other types of removable storage (e.g., digital versatile discs [DVD], floppy

---

1.pdf#search=%22ISO%2015489%22) (access 19 September 2006); and United States of America Department of Defence, *Records Management Application*, http://jitc.fhu.disa.mil/recmgt/standards.html (accessed 08 January 2005).

[33] David. R. Vernon, and Oya Y. Rieger, "Digital Asset Management: An Introduction to Key Issues (Revised 2/02 – Revision 2)," *Office of Information Technologies IT Architecture Initiative*, http://www.cit.cornell.edu/oit/Arch-Init/DigAssetMgmt.pdf, p.4 (accessed 18 November 2004).

[34] Ibid.

[35] Dr. Warwick Cathro and Tony Boston, *Development of a Digital Services Architecture at the National Library of Australia*, http://www.nla.gov.au/nla/staffpaper/2003/cathro1.html (accessed 20 December 2004).

[36] Vernon and Rieger, "Digital Asset Management."

[37] Gold/gold layer (solid gold reflective layer and phthalocyanine-based dyes) CD-R is the most stable for storage as the gold layer and phthalocyanine-based dyes have the least amount of chemical reaction.

disks, and USB drives) are commonly used, these media are not considered reliable for archival collections.[38] In a recent paper on the risks associated with use of CDs and DVDs, Kevin Bradley of the National Library of Australia explains that these discs are subject to failure, and thus, do not have the reliability to manage digital information over the long-term.[39] Failure of the discs can occur at the time of production (quality control), or while the disc is stored or being used. Poor quality control is a result of the lack of commercial standards in producing CDs; however, costly testing for manufacturers' defects can help to ensure that reliable blank discs are used when information is to be recorded. On the other hand, environmental hazards can occur at any time, and cause irreparable damage.[40] Fortunately, these risks can be minimized through proper handling and care. Further compounding the environmental hazards and the lack of quality control, CDs and DVDs, as well as other media, are vulnerable to technological obsolescence. As Bradley argues, integrated digital mass storage systems are still

---

[38] Other types of storage media include magnetic disks, magnetic tape, and solid-state media. Although other institutions have or are using this media for the storage of digital information, the Library Archives' policy is to migrate or copy corporate records onto CD and/or DVD. Consequently, the discussion is limited to this media. See also the Canadian Conservation Institute, *Modern Information Carriers* and the National Archives (UK), *Digital Preservation Guidance Note 3: Care, Handling and Storage of Removable Media*, http://www.nationalarchives.gov.uk/documents/media_care.pdf (accessed 26 September 2006).

[38] Kevin Bradley, "Risks Associated with the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections – Strategies and Alternatives", Memory of the World Programme, http://portal.unesco.org/ci/fr/files/22734/11571181531CD_DVD_for_UNESCO_final__2006.pdf/CD%2BDVD%2Bfor%2BUNESCO%2Bfinal%2B%2B2006.pdf, p.6, (accessed 26 September 2006).

[38] Vernon and Reiger, "Digital Asset Management," p.5

[38] Cathro and Boston, *Development of a Digital Services Architecture at the National Library of Australia.*

[39] Bradley, "Risks Associated with the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections – Strategies and Alternatives".

[40] Similar to paper and other analogue records, CDs and DVDs are subject to decay from environmental hazards such as heat, ultra-violet light, dust, and humidity, to name a few. Additional information is available from the Canadian Conservation Institute (e.g., *Modern Information Carriers*) as well as various internet sites (NIST Special Publications, *Care and Handling of CDs and DVDs*, http://www.itl.nist.gov/div895/carefordisc/CDandDVDCareandHandlingGuide.pdf#search=%22CD%20DVD%20storage%20conditions%20environment%22; National Archives (UK), *Digital Preservation Guidance Note 3: Care, Handling and Storage of Removable Media*) (all websites accessed 26 September 2006).

recommended, and when compared to expensive testing for quality control and obsolescence, these systems remain more affordable than CDs and other media.[41]

Aside from the hardware technology used for the architecture, software for database management is also important in managing the repository. Similar issues regarding the choice of file format for long-term digital preservation management may affect proprietary software. Control over the storage application "is one of the reasons the largest data archives in the world use a consortium derived storage application that assures access to needed software and stores data on media in a means that open source code tools can extract."[42] Like the file format guidelines, software for database management should be open standard, and organizations such as the Digital Preservation Coalition (DPC) are working to develop standards in this area. In the meantime, to avoid loss of access to the database structure that is used to organize and file digital records, the database management software should be regularly reviewed, and if necessary, migrated.

The applications used to manage, organize, and file the digital records for RIMS Division include BIBLIO, IMAGES, and iRIMS. These databases could be considered the primary entry point for the designated user community in accessing records. The library catalogue and records management servers link to the delivery systems that could refer to the document itself or a digital surrogate, via the Library Archives' persistent identifier framework. Other entry points would include the Internet, ECHOnet (IDRC Intranet), and Web Content Management System (WCMS). In either case, these services should be interoperable with other external search and retrieval processes or methods. This means that every record stored in the repository should have its own persistent and

---

[41] Bradley, "Risks Associated with the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections," p.6.
[42] Vernon and Reiger, "Digital Asset Management," p.5

unique identifier so that the database application can locate, retrieve, and disseminate the requested record. As for websites in this regard, the persistent web address can be made up of two components – the name of the resolver, which can be the server, and the persistent identifier. For example, the National Library of Australia uses the following form: http://nla.gov.au/[persistent identifier], in which http://nla.gov.au/nla.mus-an7412026-s3 is the persistent web address for page 3 of the 1905 edition of the score of Waltzing Matilda, published by Turner & Henderson.[43] The Library Archives also employs a similar practice with a persistent identifier for the records in BIBLIO. The IDRC report on *Research for Policy Influence: A History of IDRC Intent* has its unique identifier as http://idrinfo.idrc.ca/archive/corpdocs/119955/History_of_Intent.pdf. In this example, the resolver is http://idrinfo.idrc.ca/archive/corpdocs, and the persistent identifier for the record is 119955/History_of_Intent.pdf. Persistent and unique identifiers, therefore, ensure that every record in the collection is citable in a persistent manner, and that the identifier will resolve to the location of the record.

The preservation management architecture, as suggested above, must also incorporate a means of ensuring that the records are searchable. A full-text search is preferable in order to reduce the number of erroneous hits or retrievals presented to the user. Minimally, the metadata should at least be properly indexed for use by public search engines such as Google. In addition to public search engines, the National Library of Australia also includes federated discovery services such as Kinetica, PictureAustralia and the Register of Australian Archives and Manuscripts (RAAM).[44] By using

[43] Cathro and Boston, *Development of a Digital Services Architecture at the National Library of Australia.*
[44] Cathro and Boston use the word "federated" "to signify that metadata has been gathered from many contributors to create a search service supporting simultaneous discovery of information resources residing in the collections of all the contributors."

components of the National Library of Australia's architecture called the "Metadata Repository and Search System," the Library is, therefore, able to support a number of federated and specialized discovery services, allowing for increased searchability and retrieval through several interoperable systems.[45]

The digital preservation architecture is determined by several factors. These factors include the diversity of content, format types and media, storage, and cost. IDRC research outputs have diversified from the traditional paper report to digital documents with images and hyperlinks. Multimedia reports have been submitted to the Library Archives and the complexity of such records will increase in the future. Content is also becoming more diversified through language and technological capability. IDRC funds projects in several developing countries, and reports are submitted in English, French, and Spanish. IDRC also accepts reports in other languages and the encoding for accurately recognizing the bitstreams will become a preservation challenge since ASCII is currently unable to render certain semantic symbols. To assist with preservation management, the number of format types should be kept to a minimum, and the types chosen for long-term preservation should be, at the very least, ubiquitous, open standard or open source, and viable. The fragility of media storage is another challenge, and selection should be limited to Redundant Array of Inexpensive Disks (RAID) or Storage Area Network (SAN).[46] Storage has become relatively inexpensive, and challenges arise in determining whether information should be preserved offline, nearline, or online. In addition, the storage capacity of the architecture should be easily increased without

[45] Cathro and Boston, *Development of a Digital Services Architecture at the National Library of Australia.*
[46] RAID is also known as a redundant array of independent disks. This is a storage technique in which disks are connected in a series of multiple hard drives to increase the capacity and integrity of data. Storage area network is a high speed network of storage devices, connecting multiple servers.

disrupting the current databases. Cost has remained a challenge in digital preservation management since only pilot projects have been done, and extrapolation of cost is nearly impossible. These factors, as discussed in chapter two, are challenges that need to be addressed in the digital preservation management strategy and policy.

In addition to choosing a RAID or SAN storage architecture, the digital preservation strategy must ensure that the digital objects are housed or stored in a trusted digital repository (TDR). The OAIS reference model provides some guidelines for the proper storage, protection of authenticity, and access to the objects. This may be accomplished at various levels (e.g. administrative, operational, or technical) within the organization, or may be implemented through the adoption, or adaptation of a strategy or system such as the Victorian Electronic Records Strategy (VERS). At the Digital Preservation Management workshop, instructors Anne R. Kenney and Nancy Y. McGovern stressed seven attributes for a trusted digital repository that include:[47]

1. OAIS compliance
    - The recommended means and methods for implementing the model and ensuring that the implementation complies with the model
2. Administrative responsibility
    - To advise on the adoption of appropriate standards, respond to the demands of evolving standards and requirements, and document decisions and actions as needed in support of organizational objectives

---

[47] Cornell University, *Digital Preservation Management Workshop,* http://www.library.cornell.edu/iris/tutorial/dpm/foundation/tdr/index.html (accessed 26 July 2003); and RLG-OCLC report on *Trusted Digital Repositories: Attributes and Responsibilities*, available at http://www.rlg.org/legacy/longterm/repositories.pdf (accessed 25 January 2007). The description of the seven attributes has also been abridged from both websites. For a full description of each of the attributes, please see the above links.

3. Organizational viability
    - Defining feasible technical requirements (transfer protocols and specifications) for depositor agreements, and advising on and providing text to address the technical aspects of policies, procedures, and practices
4. Financial sustainability
    - To provide sound recommendations on the technical infrastructure to support the program, the cycle of replacement for hardware software, the appropriate technical solutions for preservation strategies, and supporting technologies for archival storage with associated cost information for these aspects
5. Technological and procedural suitability
    - To develop plans that will meet each requirement for the infrastructure or the necessary preservation strategies, and to ensure that the plans are executed successfully and fully with complete documentation and in compliance with established policies and standards, including the organization's responsibility to actively seek and evaluate appropriate preservation solutions and provide the means to implement them
6. System security
    - Requires technical solutions, ongoing upgrades and enhancements, and means for auditing processes
7. Procedural accountability
    - To respond to ongoing managerial needs, to anticipate those needs when possible, and to develop mechanisms for automatically generating and

capturing the necessary documentation. In some cases, compliance will require direct technical input and documentation.

Implicit within the seven attributes above are several features that support the search, retrieval, and management of authentic and reliable digital objects. While some of these features appear to be redundant within the TDR, as well as within this chapter, they are important in ensuring the success of the digital preservation strategy. Overall, a well-designed preservation infrastructure should have:

- A persistent and unique identifier for every digital object
- Sufficient metadata for search, retrieval and preservation
- Formats and quality levels that facilitate long term preservation
- Policies and procedures for compliance with standards and legislation (e.g., ISO 15489-1, CGSB "Electronic Records as Evidence", Library and Archives Act)
- Good technical support
- Financial and human resources that are committed to long-term digital preservation management
- Scalable and flexible metadata and storage repositories
- Quality control and assurance checks to validate digital objects and to check for error or corruption in data.

RIMS Division is in the beginning stages of developing its digital preservation architecture. Marjorie Whalen, in partnership with Terry Gavin, Director of

Infrastructure Support Unit in the Information and Technology Management Division, has developed a five-year plan that outlines the information management and technology requirements for the centre. In the Operational Plan, they note that the "choices made with respect to IT, in terms of resources, components, and sustainability, have an impact on the Corporate Information System [in which…] there are comparatively long-term implications for the centre's technology directions and the integration of new or enhanced technologies."[48] Consequently, RIMS Division and ITMD have identified a number of factors that need to be considered, including, but not limited to, the diverse requirements of the head office and regional offices with varying information technology capacities, fiscal responsibility and risk management to ensure that new technologies and the information created are adequately managed and accessible, and staff capacity to absorb the challenges and partake in the training related to changes in technology. The operational policy is driven by the needs of the users and of the institution – two of the three requirements for a digital preservation architecture, as described by William LeFurgy, earlier in this chapter. Furthermore, the policy defines the stakeholders and sets out their levels of responsibility and commitment, in addition to identifying the digital records, objects or assets that are to be captured and maintained within a trusted digital repository. In other words, this policy is a foundation document for IDRC to capture and preserve the records according to organizational, legal, and operational values. These requirements, albeit simple in nature, provide the backbone for a trusted digital repository and the architecture needed to create a digital preservation strategy.

---

[48] International Development Research Centre, *Operational Plan – IM & IT, Chapter 10: IT and IM management Draft 4* (07 September 2004), p. 5.

Metadata Schemas

Metadata, literally "data about data," has several meanings for information management professionals responsible for the creation, description, management, and preservation of digital objects.[49] As a result, librarians, records managers, and archivists understand this term differently, and apply the definition in various ways. In other words, metadata is defined according to the context in which it is used. In coming together to use metadata in a consistent manner, working groups and consortiums are developing more concise definitions for the application of metadata. Most definitions include the functionality required to ensure the renderability, understandability, and viability of information.

Within the archival environment, preservation metadata has become more than just a catalogue or index to digital objects; it is a statement of the information needed to support the digital preservation process.[50] More specifically, preservation metadata makes reference to each digital object and provides the associated administrative, descriptive, technical, and use information needed to render (or make readable) the digital object, provide the necessary contextual information to understand the content information, manage the digital object to ensure its authenticity, reliability and integrity, and to administer any access and use conditions.[51] Additionally, digital objects are technology-dependent and highly mutable (easily altered and subject to environmental decay), and hence, preservation metadata covers a range of information used for resource discovery, retrieval, and relationship mapping between the object and other collections.

---

[49] Susan S. Lazinger, *Digital Preservation and Metadata: History, Theory, Practice* (Libraries Unlimited, Engelwood CO, 2001).
[50] PREMIS, *Data Dictionary for Preservation Metadata,* p. ix.
[51] Lazinger, *Digital Preservation and Metadata.*

In regard to use conditions, intellectual property rights management helps to ensure sufficient rights for digital preservation activities. In other words, metadata helps to organize, manage, and maintain digital information so that the information may be retrieved and used. Accordingly, preservation metadata can be expressed as the recommended set of elements that provides the meaning, context, and structure of a digital object and its attributes.[52]

Overall metadata encompasses the technical, contextual, and administrative attributes of each digital object managed and stored in a trusted digital repository. This information plays a key role in finding, managing, controlling, and understanding each object, and from an archival perspective, metadata also provides information on the document's provenance, or the relationship of the document to its creator, creating institution, and creating processes, including other contextual documentation. From the archival perspective, the metadata function is to ensure the authenticity, integrity, and reliability of the record. And more specifically, preservation metadata has particular importance as it allows for the "re-creation and interpretation of the structure and content of digital data over time."[53] This assists in recovering files created on obsolete systems, which then can be rendered by current computer systems. Hence, metadata is important for the preservation and access of digital information.

---

[52] For more definitions on preservation metadata, see also National Library of Australia, http://www.nla.gov.au/preserve/pmeta.html; Anne Gilliland- Swetland, *Setting the Stage*, http://www.slis.kent.edu/~mzeng/metadata/Gilland.pdf; Rebecca Guenther, "PREMIS – Preservation Metadata Implementation Strategies Update 2: Core Elements for Metadata to Support Digital Preservation", *RLG DigiNews*, vol. 8, no. 6 (15 December 2004), http://www.rlg.org/en/page.php?Page_ID=20492&Printable=1&Article_ID=1693; and PREMIS (all sites accessed 24 March 2006, except PREMIS, accessed 25 May 2005).

[53] Traugott Koch, "Metadata topics", *DELOS brainstorming* (Jaun les Pins 2005-12-05), http://www.ukoln.ac.uk/ukoln/staff/t.koch/pres/Brainst200512-MDc.html (accessed 26 September 2006).

With increasing research and development into metadata, standards have now become essential tools for managing not only digital information, but also the metadata elements themselves. One example is the Metadata Encoding and Transmission Standard (METS), a standard for encoding descriptive, administrative, and structural metadata through standardized profiles to provide sufficient detail for creating and processing metadata documents. Metadata standards and schemas are improving the interoperability and collaboration of information between organizations, while providing a consistent approach to describing information, and supporting the life-cycle management of digital objects. Consequently, preservation metadata plays a central or key role in the digital preservation strategy.

While several different metadata standards exist, Michael Day of UKOLN (UK Office for Library Networking) proposes a metadata registry-based approach to help in producing metadata standards and increase interoperability between schemas for long-term digital preservation.[54] Day argues that such a registry would function as an authoritative source for terms and vocabulary, in which the registry would populate the metadata required in a trusted digital repository, while also exchanging metadata with other archival and library repositories when information packages are exported. Conversely, the existence of multiple metadata for digital objects will require qualitative decisions to determine the authoritative metadata. This challenge can be minimized through the use of unique identifiers so that proper metadata can be identified and combined with other schemas. Another approach suggests storing the metadata in a database that links the metadata to the digital object. In this case, digital objects are

---

[54] Michael Day, *Integrating Metadata Schema Registries with Digital Preservation Systems to Support Interoperability: a Proposal*, http://www.siderean.com/dc2003/101_paper38.pdf (accessed 20 June 2004).

linked to the database that contains the same metadata. Therefore, that metadata is recorded once, reducing the resources required for storing the same metadata in multiple copies and/or places. Another metadata schema proposes that the metadata be embedded or wrapped with the digital object. The metadata and the digital object are combined, and constitute a single information package. In some instances, institutions have combined these initiatives in which some metadata is automatically generated and embedded with the digital object, some metadata is stored in a database with links to the digital object, and some metadata is entered into a registry for increased interoperability between schemas.

The Dublin Core Metadata Initiative (DCMI) began in 1995 in order to develop standards for information resources. Since then the scope of the DCMI has expanded to "developing metadata standards for discovery across domains, defining frameworks for the interoperation of metadata sets, and facilitating the development of community- or disciplinary- specific metadata that are consistent with [metadata discovery and metadata frameworks]."[55] As such the DCMI has been active in the development and maintenance of standards, tools, services and infrastructure to support the management of the DCMI metadata registry, and education outreach for other metadata communities.[56] The Dublin Core metadata is not intended to replace other metadata schemas, but instead to be interoperable with other standards using the same resource description or semantics. Hence, the Dublin Core has been described as a "generic standard" that can be easily adapted to fit more complex metadata schemes.[57] The following element set promotes

---

[55] Dublin Core Metadata Initiative (DCMI), http://dublincore.org/about/ (accessed 07 January 2005).
[56] Ibid.
[57] Dublin Core Metadata Initiative, *Information and documentation – The Dublin Core metadata element set* (ISO Draft Paper, 26 February 2003), http://www.niso.org/international/SC4/n515.pdf (accessed 06

global interoperability through a suggested controlled vocabulary and conventions intended for syntactic specification of elements for ease of encoding schemes. These elements have been approved as an ISO Standard 15836.

The Dublin Core element set consists of 15 elements:

Contributor: An entity responsible for making contributions to the content of the resource
Coverage: The extent or scope of the content of the resource
Creator: An entity primarily responsible for making the content of the resource
Date: A date of an event in the lifecycle of the resource
Description: An account of the content of the resource
Format: The physical or digital manifestation of the resource
Identifier: An unambiguous reference to the resource within a give context
Language: A language of the intellectual content of the resource
Publisher: An entity responsible for making the resource available
Relation: A reference to a related resource
Rights: Information about rights held in and over the resource
Source: A reference to a resource from which the present resource is derived
Subject: A topic of the content of the resource
Title: A name given to the resource
Type: The nature or genre of the content of the resource

The Open Archival Information System (OAIS) is a functional high-level model for managing and yielding the archival information package (AIP) into an understandable form.[58] The AIP contains content information and its associated preservation description information (PDI). The latter information is one of four distinct information objects (content information, preservation description information, packaging information, and descriptive information), and is responsible for supporting the long-term retention of the

---

January 2005), and Day, *Integrating Metadata Schema Registries*, p. 3. The Dublin Core Metadata Initiative views the simplicity of its metadata set as both a strength and weakness. While the metadata set does not have the capability for supporting semantic and functional richness of more complex schemes, the metadata set does lower the cost of creating metadata and promotes interoperability.

[58] Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS),* (see also chapter 1 for more detail on the OAIS reference model).

content information. In addition, the PDI functions to document the OAIS preservation process. This metadata framework identifies four requisite categories – reference, context, provenance, and fixity. Within the OAIS reference model, the preservation metadata schema, as well as the other information objects, emphasizes a conceptual or theoretical approach, rather than a well-defined element set such as the Dublin Core.

Preservation description information records the "*identity, relationships, history and integrity* of the archived Content Data Object [or the bitstreams]."[59] This metadata forms the infrastructure that supports the processes associated with digital preservation management. In the OAIS model, the reference information, context information, provenance information, and fixity information are consistent with a comprehensive and structured approach to long-term digital preservation. Reference information identifies and describes the mechanisms and unique or persistent identifiers for the content information, as described in the digital preservation management architecture. The context information documents the relationship of the content information to its environment, including the purpose of the content information, and its relationship to other content information objects. Provenance information encompasses the history of the content information, its creator, creating institution, the chain of custody and control, and any changes that may have occurred during its lifecycle. Fixity information authenticates the content information by providing the proper mechanisms and keys to ensure that changes occur in a documented manner. These categories, therefore, maintain

---

[59] OCLC/RLG Working Group on Preservation Metadata, *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects* (Dublin, Ohio, June 2002), http://www.oclc.org/research/projects/pmwg/pm_framework.pdf, p. 28 (emphasis in original) (accessed 07 January 2005).

the viability, renderability, and understandability of digital resources for a long period of time.[60]

An example of an OAIS compliant metadata schema is PREMIS (Preservation Metadata: Implementation Strategies), which was developed by an expert working group that convened "to consider the types of information falling within the scope of preservation metadata… [that included] a comprehensive metadata framework describing the information necessary to carry out, document, and evaluate digital preservation processes."[61] The implementation of this metadata framework is based upon the "synthesis of four existing preservation metadata schemes, developed by the CURL Exemplars in Digital Archives project (CEDARS), the National Library of Australia (NLA), the Networked European Deposit Library (NEDLIB), and the Online Computer Library Center, Inc. (OCLC), respectively."[62] The objectives of this group include the creation of core preservation metadata elements with broad applicability, a data dictionary for standardized terms, analysis of alternative metadata strategies and their interoperability, the development of best practices, and the promotion for the collaboration of cooperative sharing of information on preservation metadata.[63]

The Records Continuum Research Group (RCRG), located at Monash University (Australia), is working towards the development of a metadata standard. The SPIRT (Strategic Partnerships with Industry – Research & Training) Recordkeeping Metadata project "aims to comprehensively specify and codify recordkeeping metadata in ways that enable it to be fully understood and deployed both within and beyond the records and

---

[60] Ibid., p.1.
[61] PREMIS, http://www.oclc.org/research/projects/pmwg/ (accessed 08 January 2005).
[62] OCLC/RLG Working Group on Preservation Metadata, *Preservation Metadata and the OAIS Information Model*, p.3.
[63] PREMIS, http://www.oclc.org/research/projects/pmwg/.

archives profession."[64] In other words, metadata should, similar to the OAIS model, provide the intellectual control that enables accessibility to reliable, authentic, and meaningful records over time. In order to do so, the metadata needs to be identified, categorized, and presented in a formal, standardized manner that supports proper recordkeeping practices from the point of creation to the end of the record's lifecycle. Additionally, the RCRG stressed the development of compatible and interoperable metadata schemas to effectively manage records in a distributed network and the broader metadata community.

The Austalian Recordkeeping Metadata Schema (RKMS) is the standardized set of structured recordkeeping metadata elements, and the framework for developing and specifying the standard, which also includes the way in which mapping metadata sets have established interoperability for semi-automated translation between schemas.[65] In conforming to the ISO 9000 specifications, the metadata schema consisted of various layers, each with their own types of information.[66] The handle layer contains the record identification metadata and the content discovery metadata. These layers function to identify and make accessible the record. In the identification metadata, the "Transaction-Domain-Identifier" and "Transaction-Instance-Identifier" document the information relating to the type of product or service that emanated the record, including the creating organizational unit, and the time and date of creation, as well as providing a unique identifier. The terms and conditions layer includes metadata relating to restrictions

---

[64] Monash University Records Continuum Research Group, *Recordkeeping Metadata Project*, http://www.sims.monash.edu.au/research/rcrg/research/spirt/about.html (accessed 30 November 2004).
[65] Monash University Records Continuum Research Group, *Recordkeeping Metadata Project*, http://www.sims.monash.edu.au/research/rcrg/research/spirt/deliverables.html (accessed 30 November 2004).
[66] Sue McKemmish and Wendy Duff, *Metadata and ISO 9000 Compliance*, http://www.sims.monash.edu.au/research/rcrg/publications/smckduff.html (accessed 30 November 2004).

status, access conditions, user conditions, and disposition (retention) requirements. Metadata relating to the operational system, software applications, and hardware dependencies provides the necessary information to render or make understandable the electronic record. This metadata interprets the bitstreams into human-readable symbols. Additional metadata is required for the structure or form of the record. For an electronic document, the structure is controlled by software functionality, and may help guard against loss of information as records are migrated. In other words, this metadata helps to preserve the "look-and-feel" of the record. The contextual layer "provides information about the transaction that created the record including the product or service that the record documents."[67] This layer of metadata identifies the policies and procedures controlling the product/service, and the relationship of the record to the product involved, thereby allowing for transparent business practices. Content and use history refer to the content of the record and to an audit trail of information relating to the use of the record. The use history layer contains document control metadata, similar to the fixity information of the OAIS model. The RKMS are designed to identify the metadata necessary for the preservation and accessibility of reliable, authentic, and meaningful records.

In October 2000, RIMS Division commissioned XIST Information Services and Technology, Inc. to develop a metadata user's guide for RIMS Division.[68] This guide recommends the adaptation of the fifteen elements outlined in the Dublin Core Metadata Initiative, as well as adding an additional five elements that are IDRC specific. Each element is given a name and an identifier, in which the identifier is used to identify the

---

[67] Ibid.
[68] Tim Wayne, *International Development Research Centre Meta Data User's Guide* (XIST Information Services & Technology, Inc., October 2000).

specific metadata represented in the content. An example is <META NAME="DC.Title" CONTENT="IDRC Homepage">, in which the metadata element content is "IDRC Homepage" and is identified as a Dublin Core title element. Modifications to the Dublin Core element set include:

- The addition of:
  - **Development Activity Title** – title of the development activity as assigned by the agency responsible [Identifier: CEFDA.ActivityTitle]
  - **Development Activity ID** – unique identifier assigned by the agency responsible for the development activity [Identifier: CEFDA.ActivityID]
  - **IDRC web publisher** – person or organization responsible for making the [resource] publicly available [Identifier: idrc.publisher]
  - **Date of first IDRC availability** – date the [resource] was first made publicly accessible by IDRC [Identifier: idrc.datepublished]
  - **Subject: Keyword** – a short list of keyword or keyphrases that describe the [resource] [Identifier: DC.Subject]
  - **Coverage.Geographic** – country and/or region on which the content of the [resource] is focused [Identifier: DC.Coverage.Spatial]
- The division of:
  - Dublin Core element DATE into:

      - **Date Available** – date content was first made public [Identifier: DC.Date.Available]
      - **Date Valid** – date content is valid [Identifier: DC.Date.Valid]
  - Dublin Core element RIGHTS into:
      - **Use contstraints: Intellectual Property Rights** – information about Intellectual Property rights held in and over the [resource] [Identifier: DC.Rights]
      - **Use constraints: Other** – Information about other (non intellectual property) constraints on the use of the content [Identifier: DC.Rights]
- Dublin Core TYPE is renamed **Resource Type** [Identifier: DC.Type]

While RIMS Division has implemented a modified Dublin Core metadata set for the cataloguing, indexing, search, and retrieval, they have yet to implement a preservation metadata framework. An OAIS compliant strategy is recommended, and includes core elements that document the context and history (provenance), fixity information relating to security and access, as well as establishing an audit trail of changes such as checksum data, and reference information that contains information relating to software and hardware technology. Metadata regarding the migration of files and the refreshment of media should include the date and time when the content information was migrated or refreshed, and the identification of the system or person responsible for the migration and refreshment cycles. Paul Wheatley of the University of Leeds suggests that this type of change metadata should be a core element in any schema.[69] The file format for the metadata schema should be open source such as UNICODE, ASCII, or XML. Metadata Encoding and Transmission Standard (METS) is one such example of an XML document format and metadata schema.[70] In some cases, the metadata may be better retained in a registry as Michael Day notes, or may be saved in a database with links to the content information. Finally, most of the metadata is automatically generated; however, some elements of the core metadata will need to be manually entered.

Final Report Post-Evaluation

When I presented my final report to Denys Vermette (Vice-President of Resources), Marjorie Whalen and RIMS Division staff, and others from various IDRC program areas, I struggled with the complexity of the digital preservation challenges faced by the centre. Since then, my understanding of the broader issues has improved, and I am better able to apply technical solutions to the user and organizational needs. Consequently, I more fully grasp the "bigger picture" in which digital preservation lies. As I reflect on my presentation and final report, brief indications of a more comprehensive strategy exist in these documents, but it is not until now that I have become cognizant of the relationship of the digital preservation strategy to corporate policy.

Both the presentation and final report mention policy development, identification of key stakeholders, and legal and operational recordkeeping requirements to support a

---

[69] Paul Wheatley, *Institutional Repositories in the context of Digital Preservation* (Digital Preservation Coalition, March 2004), http://www.dpconline.org/docs/DPCTWf4word.pdf (accessed 09 November 2004).
[70] Metadata Encoding and Transmission Standard (METS), http://www.loc.gov/standards/mets/ (accessed 08 January 2005).

digital preservation strategy. These are important; however, the rationale as to *why* policies, key stakeholders, and recordkeeping requirements were necessary was absent. Simply put, policies provide the mandate to implement and enforce the digital preservation strategy. Key stakeholders champion the digital strategy by authorizing the policies and putting them into effect. And finally, the legal and operational recordkeeping requirements justify the need for a digital preservation strategy. Even though these are crucial to the success of the digital preservation strategy, the end-users can have a stronger influence over the digital preservation strategy than other stakeholders, or operational requirements – an important point for IDRC as they are strongly research-based, and have regional offices with varying degrees of technological complexity.

While the organizational culture of IDRC plays an important role in developing and implementing a digital preservation strategy, the strategy remains strongly affected by the technology used to create the digital objects. A sound digital preservation strategy incorporates the regular practice of reviewing, monitoring, and modifying guidelines and best practices as new information and technology become available. As mentioned throughout this thesis, digital information and the software and hardware used to create it are subject to rapid technological obsolescence. In order to keep digital objects viable, technology needs to be monitored so that the digital objects can be managed over the long-term, which is beyond the lifespan of the technology used to create the object. As one author so perceptively noted, there are: "three things in life that are certain – death, taxes, and that your hard drive will fail [and become obsolete]."[71] At the minimum, best

---

[71] Arthur, "The End of History".

practices in digital preservation management will at least minimize the calamity of the failing hard drive!

IDRC continues to be proactive in its approach to the long-term preservation and management of its digital records, and this critical success factor, which includes support from senior management, will increase the likelihood of RIMS Division's success in implementing a digital preservation strategy. The centre's emphasis on the importance of research and collaboration will also help RIMS Division to address organizational needs more effectively as it works toward full implementation the digital preservation strategy. The emphasis on collaboration has helped me grasp the unique challenges related to digital preservation, and merits further discussion in the next chapter.

CONCLUSION

In this short concluding chapter, the importance of research and collaboration is presented as part of the development of the digital preservation strategy that was recommended for IDRC. At Cornell University's *Digital Preservation Management Workshop* and during the facilitated discussion of NRC-CISTI's Digital Archiving Research Framework (DARF), two themes were taught and/or evolved out of these events. The first theme is the necessity for research in order to understand the challenges of digital preservation. The second theme is the value of collaboration within an institution as well as with external organizations and individuals to address long-term the management and preservation of digital objects. As RIMS Division implements and sustains its digital preservation strategy, it will find that research and collaboration will provide the tools to maintain the long-term viability of its digital corporate memory.

Margaret Hedstrom notes that "there is a dark side to the rapid growth in digital information. The technologies, strategies, methodologies, and resources needed to manage digital information in the long term have not kept pace with innovations in the creation and capture of digital information."[1] Technological obsolescence and the increasing growth of digital collections in libraries and archives are outpacing our ability to bring the necessary human and financial resources to manage and preserve them. Thus digital archiving research initiatives, such as those identified in this thesis, are often collaborative partnerships among archival and library organizations, computer sciences,

---

[1] Margaret Hedstrom, *It's About Time: Research Challenges in Digital Archiving and Long-term Preservation, Final Report, Workshop on Research Challenges in Digital Archiving and Long-term Preservation, April 12-13, 2002* (August 2003), http://www.digitalpreservation.gov/about/NSF.pdf (accessed 04 January 2007), p.vii.

and information sciences can be developed.[2] The OAIS reference model is a good example of scientific, technical and social organizations coming together to collaborate on digital archiving problems.

Cornell University's 2004 *Digital Preservation Management* workshop also emphasized the importance of collaboration during the session on the five organizational stages of digital preservation.[3] The two latter stages stressed the need for partnerships within the organization and with external partners. Stage four, Institutionalize, focuses on incorporating support on an organization-wide basis. For example, information management and information technology divisions may form an inter-institutional committee on digital preservation management. As the organization moves toward greater collaboration with other organizations, institutions, and/or individuals, the organization then enters into stage five, Externalize. This last stage is characterized by formal and informal arrangements with others outside the institution. The institution may either be the "lead" or a "participant" in these partnerships. The research consultation hosted by NRC-CISTI also stressed the importance of collaborative approaches to digital archives research.[4] Although the recommendations from the consultation aligned more with stage five of the *Digital Preservation Management* workshop organizational stages, the need for research and collaboration in developing a digital preservation strategy was considered a priority by CISTI's Advisory Board in meeting the first goal outlined in

---

[2] Ibid.

[3] Cornell University, *Digital Preservation Management Workshop,* 19-23 July 2004.

[4] National Research Council Canada – Canada Institution for Scientific and Technical Information (NRC-CISTI), "Digital Archives Research: Recommendations of the Digital Archives Research Framework Consultation of September 27th 2004, Ottawa ON" (final report, 22 June 2005), http://cisti-icist.nrc-cnrc.gc.ca/darf-ran/darf-ran-eng.pdf (accessed 05 January 2007).

their 2005-2010 Strategic Plan.[5] Following my participation in the research consultation, I submitted a summary report to RIMS Division, which recognized that:

> 6. [The] IDRC model for digital preservation should encompass the ability and flexibility [...] [t]o collaborate and network with other institutions
>
> 7. CISTI is in the process of developing a digital archiving research framework, and future collaboration with this organization will benefit IDRC in its development of a long-term preservation management strategy.[6]

Within IDRC, RIMS Division works with other divisions to provide access to IDRC reports and projects for staff, research recipients, and the general public. In addition, RIMS Division has the primary responsibility for the permanent preservation of non-digital and digital information. This means that RIMS Division should continue to develop supportive partnerships with IDRC information technology (IT) staff. During my internship, I had the opportunity to develop several collaborative relationships with IT staff, and thus gain a better understanding of the technical aspects of digital information and technology. I also networked with others in the archival, library, and information management fields. The collaboration enabled me to have a fuller understanding of long-term digital preservation management, and to provide informed recommendations to RIMS Division.

Nestled within research and collaboration is an understanding of information technology. Adam Jansen, Digital Archivist at the Washington State Digital Archives, provides the following advice for archivists: "If you don't know it, learn it" and "Speak

---

[5] Ibid., 2. Introduction, 2.1 Context.
[6] International Development Research Centre, RIMS Division, "TRIP REPORT, CISTI Digital Archive Research Framework Consultation" (28 September 2004).

geek."[7] Part of understanding the digital preservation challenge means understanding the technology that is used to create records. This includes grasping the terminology, and being able to apply it within an archival environment. At IDRC, I began to learn about and gain experience in the information technology environment.

Others in the information management and technology fields are also discovering the need to communicate better with each other on the challenges facing digital preservation. A March 2006 thread on the E-RECS list serve illustrates the impact of information technology on the responsibilities of today's information management professional (see posting below). This discussion showed how records managers and archivists are broadening their knowledge base to communicate more effectively with their information technology counterparts. Colman Murphy of DocuShare (Xerox) posted the following message:

> Which do YOU feel is happening more in your environment?
>
> A) The IT department is re-assigning their resources to learn more about Records Management, so they can do a better of job of supporting RM/ERM initiatives
>
> B) RM professionals are taking the time to learn more about the technologies and applications that will impact their ability to do their jobs in the months and years ahead.
>
> This may sound like a philosophical question, but it immediately impacts the kinds of educational materials we will develop to support our RM software sales.
>
> I appreciate any and all feedback.
>
> Regards,
> Colman Murphy
> DocuShare Product Manager[8]

---

[7] Adam Jansen, "Digital Preservation in the 20th Century" presentation given at the 2005 Government of Alberta *Managing Information Assets in the Public Sector* (18 October 2005).

[8] Colman Murphy, "Is IT Learning about RM, or is RM Learning about IT??" 28 March 2006 posting on the E-RECS list serve. Available in the E-RECS Archives at http://listserv.albany.edu:8080/cgi-bin/wa?A1=ind06&L=erecs-l#98 (accessed 05 November 2006).

Murphy received nine responses that answered his question directly, and one that dealt with the overall strategic direction of electronic records management. Of the nine responses, one person answered that his IT department is re-assigning resources to learn more about records management (RM); five people indicated that records managers ought to learn more about IT; and three responded that the resources in their institutions were equally split between IT and RM. Bernard Chester, Principal of IMERGE Consulting, Pacific Northwest Office, was one of two respondents who commented: "I'd say that the predominance is RM learning about IT. I think it is understandable -- everyone has a PC on their desktop and one at home, but how many of us get exposed to records management in school or in normal course of business?"[9] Larry Medina's response to Murphy's question supports this view: "IT is finding a need to understand how RIM [Records and Information Management] works and how business uses the information [....] and RM needs to be aware of the policies and practices being applied to data storage to ensure proper retention is being achieved" as well as learning about automated taxonomy and classification.[10] While it appears that records managers and archivists are learning more about information technology, it also seems that information technologists are learning about records management and archives. My internship experience supports this observation, as both information managers and information technologists I encountered agreed that "if you don't know, learn it."

[9] Bernard Chester to Colman Murphy, "Re: Is IT learning about RM, or is RM learning about IT??" posted March 28, 2006 on the E-RECS list serve. The archive of this email is available in the E-RECS Archive at http://listserv.albany.edu:8080/cgi-bin/wa?A2=ind06&L=erecs-l&H=1&P=20521 (accessed 05 November 2006).

[10] Larry Medina to Colman Murphy, "Re: Is IT learning about RM, or is RM learning about IT??" posted March 30, 2006 on the E-RECS list serve. The archive of this email is available in the E-RECS Archive at http://listserv.albany.edu:8080/cgi-bin/wa?A2=ind06&L=erecs-l&H=1&P=22389 (accessed 05 November 2006).

Bibliography

Primary Sources:

Brentnell, Kyle. Personal communication.

Chester, Bernard to Colman Murphy, "Re: Is IT learning about RM, or is RM learning about IT??" posted to the E-RECS list serve on 28 March 2006.

Cornell University, *Digital Preservation Management Workshop*, 19-23 July 2004.

----. *Digital Preservation Management Workshop*, "Introduction," http://www.library.cornell.edu/iris/tutorial/dpm/foundation/index.html.

----. *Digital Preservation Management Workshop,* "Trusted Digital Repositories," http://www.library.cornell.edu/iris/tutorial/dpm/foundation/tdr/index.html.

----. *Digital Preservation Management Workshop*, "OAIS Terms," http://www.library.cornell.edu/iris/tutorial/dpm/terminology/oais.html.

----. *Digital Preservation Management Workshop*, "Digital Preservation," http://www.library.cornell.edu/iris/tutorial/dpm/terminology/preservation.html.

Creative Archiving at Michigan & Leeds: Emulating the Old on the New (CAMiLEON), available at http://www.si.umich.edu/CAMILEON/.

----. http://www.si.umich.edu/CAMILEON/index.html.

Digital Preservation Coalition (DPC), available at http://www.dpconline.org/graphics/index.html.

Dublin Core Metadata Initiative (DCMI), available at http://dublincore.org/about/.

----. *Information and documentation – The Dublin Core metadata element set* (ISO Draft Paper, 26 February 2003), http://www.niso.org/international/SC4/n515.pdf.

ERPANET, *erpaguidance Cost Orientation Tool*, (September 2003), http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf.

International Development Research Centre. *Corporate Strategy and Framework Program 2000-2005*, http://idrinfo.idrc.ca/archive/corpdocs/116539/CSPF_2000-2005.htm.

----. "Marjorie Whalen to Denys Vermette, Memorandum for Intern Proposal," 10 September 2003.

----. *Operational Plan – IM & IT, Chapter 10: IT and IM management Draft 4* (07 September 2004).

----. "TRIP REPORT, CISTI Digital Archive Research Framework Consultation" (28 September 2004).

----. *Annual Report 2004-05*, http://www.idrc.ca/uploads/user-S/11307698851AnnualReport_2004-05_e2.pdf.

----. *The International Development Research Centre: A Brief History* (2005), http://www.idrc.ca/uploads/user-S/11394255461History_rev_e.pdf.

----. *Corporate Strategy and Framework Program 2005-2010*, http://www.idrc.ca/uploads/user-S/11250758901CSPF_2005_e.pdf.

----. *About IDRC – Briefing Book* website (2006), http://www.idrc.ca/en/ev-66231-201-1-DO_TOPIC.html.

----. *Briefing Document* (February 2006), http://www.idrc.ca/uploads/user-S/11408155631Briefing_book_Jan2006_e.pdf.

International Development Research Centre, RIMS Division Library Archives. "Notes For The Prime Minister's Remarks To The Canadian Political Science Association Banquet Carleton University, Ottawa, June 8, 1967," http://idrinfo.idrc.ca/archive/corpdocs/010474/Pearso5.pdf

----. "A.F.W. Plumptre, Feasibility Study on the establishment of IDRC, December 22, 1967," http://idrinfo.idrc.ca/archive/corpdocs/117258/01e.pdf.

----. "Memorandum Regarding the Proposed International Development Research Centre, January 24, 1968," http://idrinfo.idrc.ca/archive/corpdocs/117258/letters1.pdf.

----. "House of Commons Debates (First Session – Twenty Eighth Parliament 17 Elizabeth II, Volume 1, 1968) Ottawa 1969," http://idrinfo.idrc.ca/archive/corpdocs/069849/03aE.pdf.

----. "Working Paper on a Strategy for the Establishment of the International Development Research Centre of Canada, 26 May 1969," http://idrinfo.idrc.ca/archive/corpdocs/117258/letters3.pdf.

----. "Archiv No. 341.1 IDRC I581, International Development Research Centre of Canada: Memorandum to Cabinet Supporting Proposed International Development Research Centre and Draft of an Act to Establish the IDRC of Canada, 30 June 1969," http://idrinfo.idrc.ca/archive/corpdocs/069849/02e.pdf.

----. "Memorandum Regarding the Objectives, Style and Structure: a Possible Approach for the International Development Research Centre of Canada" (9 January 1970), http://idrinfo.idrc.ca/archive/corpdocs/117258/letters6.pdf.

----. "Archiv Archer no. 105571, Archives and records management: an IDRC 25 year retrospective of support."

----. *Guidelines for Library Archives and Records Management Research Information Management Service Division (RIMSD) (Draft).*

International Research on Permanent Authentic Records in Electronic Systems (InterPARES), available at http://www.interpares.org/.

----. InterPARES 1, Preservation Domain, http://www.interpares.org/ip1/ip1_ptf.cfm.

----. InterPARES 2, Project Summary, http://www.interpares.org/ip2/ip2_index.cfm.

International Organization for Standardization. *Information and Records Management Standard 15489*, http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=31908&ICS1=1&ICS2=140&ICS3=20.

Joint Information Systems Committee (JISC), available at http://www.jisc.org.uk/.

Library and Archives Canada, "Archival Appraisal Report on the Records of the International Development Research Centre, 15 April 1991," (ATIP request 29 March 2006).

Medina, Larry to Colman Murphy, "Re: Is IT learning about RM, or is RM learning about IT??" posted to the E-RECS list serve on 30 March 2006.

Metadata Encoding and Transmission Standard (METS), available at http://www.loc.gov/standards/mets/.

Monash University Records Continuum Research Group (RCRG), available at http://www.sims.monash.edu.au/research/rcrg/index.html.

----. *Recordkeeping Metadata Project*, http://www.sims.monash.edu.au/research/rcrg/research/spirt/about.html.

----. *Recordkeeping Metadata Project*, http://www.sims.monash.edu.au/research/rcrg/research/spirt/deliverables.html.

Murphy, Colman, "Is IT Learning about RM, or is RM Learning about IT??" posted to the E-RECS list serve on 28 March 2006.

National Archives and Records Administration (NARA), Electronic Records Archives (ERA), available at http://www.archives.gov/era/.

----. "The National Archives and the San Diego Supercomputer Center Sign Landmark Agreement to Preserve Critical Data" Press Release (28 June 2006), http://www.archives.gov/press/press-releases/2006/nr06-119.html.

National Institute of Standards and Technology (NIST). *Care and Handling of CDs and DVDs*, Special Publications, http://www.itl.nist.gov/div895/carefordisc/CDandDVDCareandHandlingGuide.pdf#search=%22CD%20DVD%20storage%20conditions%20environment%22.

National Library of Australia, "Preservation Metadata," http://www.nla.gov.au/preserve/pmeta.html.

National Research Council Canada – Canada Institution for Scientific and Technical Information (NRC-CISTI). *Digital Archives Research: Recommendations of the Digital Archives Research Framework Consultation of September 27th 2004, Ottawa ON* (final report, 22 June 2005), http://cisti-icist.nrc-cnrc.gc.ca/darf-ran/darf-ran-eng.pdf.

Netarchive.dk. *About Netarchive.dk – 2nd Phase*, http://www.netarchive.dk/.

Online Computer Library Center/Research Library Group (OCLC/RLG). *Trusted Digital Repositories: Attributes and Responsibilities An RLG-OCLC Report* (Mountain View, California, May 2002), http://www.rlg.org/legacy/longterm/repositories.pdf.

Online Computer Library Center/Research Library Group (OCLC/RLG) Working Group on Preservation Metadata. *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects* (Dublin, Ohio, June 2002), http://www.oclc.org/research/projects/pmwg/pm_framework.pdf.

Preservation Metadata: Implementation Strategies (PREMIS). *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group* (May 2005), http://www.oclc.org/research/projects/pmwg/premis-final.pdf.

United States of America Department of Defence. *Records Management Application*, http://jitc.fhu.disa.mil/recmgt/standards.html.

University of British Columbia, School of Archival, Library and Information Studies (SLAIS), available at http://www.interpares.org/UBCProject/index.htm.

University of Indiana *Sound Directions: Digital Preservation and Access for Global Audio Heritage*, available at http://www.dlib.indiana.edu/projects/sounddirections/.

University of Pittsburgh Functional Requirements for Evidence in Recordkeeping Project, available at http://www.archimuse.com/papers/nhprc/.

Van Garderen, Peter to Heather Lyons, "Re: Long-term preservation of electronic records" posted to the E-RECS list serve on 17 October 2006.

Victorian Electronic Records Strategy (VERS), available at http://www.prov.vic.gov.au/vers/vers/default.htm.

Wayne, Tim. *International Development Research Centre Meta Data User's Guide* (October 2000), XIST Information Services & Technology, Inc.

Secondary Sources:

Arthur, Charles. "The End of History" (30 June 2003), *The Independent*, London, UK.

Barry, Rick. "Making a Difference: Comments on Electronic Records Management R&D Projects at Ohio State University, Indiana University and City of Philadelphia: SAA Annual Meeting August 29, 1996, Session 6," www.mybestdocs.com/barry-r-saa1996-er-rd.html.

Bearman, David. *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*, (Archives & Museum Informatics, Pittsburgh, 1994), http://www.archimuse.com/publishing/electronic_evidence/ElectronicEvidence.Intro.pdf#search=%22%22Electronic%20evidence%3A%20strategies%20for%20managing%20records%20in%20contemporary%20organizations%22%22.

----. "Reality and Chimeras in the Preservation of Electronic Records", *D-Lib Magazine*, vol. 5, no. 4 (April 1999), http://www.dlib.org/dlib/april99/bearman/04bearman.html.

Bettels, Jürgen and F. Avery Bishop. *Unicode: A Universal Character Code*, http://python.planetmirror.com/ftp/macro/DEC/DTJ/DTJB02/DTJB02PF.PDF.

Bradley, Kevin. "Risks Associated with the Use of Recordable CDs and DVDs as Reliable Storage Media in Archival Collections – Strategies and Alternatives", Memory of the World Programme (2006), http://portal.unesco.org/ci/fr/files/22734/11571181531CD_DVD_for_UNESCO_final__2006.pdf/CD%2BDVD%2Bfor%2BUNESCO%2Bfinal%2B%2B2006.pdf.

Brown, Adrian. *Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation* (UK National Archives, 19 June 2003), http://www.nationalarchives.gov.uk/preservation/advice/pdf/selecting_file_formats.pdf.

----. *Digital Preservation Guidance Note 3: Care, Handling and Storage of Removable Media* (UK National Archives, 27 June 2003), http://www.nationalarchives.gov.uk/documents/media_care.pdf.

Cathro, Dr. Warwick and Tony Boston. *Development of a Digital Services Architecture at the National Library of Australia*, http://www.nla.gov.au/nla/staffpaper/2003/cathro1.html.

Clausen, Lars R. *Concerning Etags and Datestamps*, http://www.netarchive.dk/website/publications/Etags-2004.pdf.

Cloonan, Michèle V. and Shelby Sanett. "Comparing Preservation Strategies and Practices for Electronic Records", http://www.rlg.org/events/pres-2000/cloonan.html.

Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)* (January 2002), http://public.ccsds.org/publications/archive/650x0b1.pdf.

Cook, Terry. "It's 10 O'Clock: Do You Know Where Your Data Are?," *The Technology Review* (January 1995).

----. "Appraisal Methodology: Macro-Appraisal and Functional Analysis – Part A: Concepts and Theory", (Library and Archives Canada, Summer 2000), http://www.collectionscanada.ca/information-management/061101_e.html.

Day, Michael. *Integrating Metadata Schema Registries with Digital Preservation Systems to Support Interoperability: a Proposal*, http://www.siderean.com/dc2003/101_paper38.pdf.

Duff, Wendy. "Ensuring the Preservation of Reliable Evidence: A Research Project Funded by the NHPRC", *Archivaria* 42 (Fall 1996).

Fineberg, Gail. "Capturing the Web: Staff Briefed on National Digital Preservation Plan," *Library of Congress Information Bulletin*, vol. 62, no. 4 (April 2003), http://www.loc.gov/loc/lcib/0304/digital.html.

Fleischhauer, Carl. "The Library of Congress Digital Audio Preservation Prototyping Project" presentation, *Sound Savings: Preserving Audio Collections Proceedings* *from the Symposium*, http://www.arl.org/preserv/sound_savings_proceedings/fleischhauer.html.

Gilliland- Swetland,Anne. *Setting the Stage*, http://www.slis.kent.edu/~mzeng/metadata/Gilland.pdf.

Guenther, Rebecca. "PREMIS – Preservation Metadata Implementation Strategies Update 2: Core Elements for Metadata to Support Digital Preservation", *RLG DigiNews*, vol. 8, no. 6 (15 December 2004), http://www.rlg.org/en/page.php?Page_ID=20492&Printable=1&Article_ID=1693

Granger, Stewart. "Emulation as a Digital Preservation Strategy", *D-Lib Magazine*, vol. 6, no. 10 (October 2000), http://www.dlib.org/dlib/october00/granger/10granger.html.

Haigh, Susan and Roselyn Lilleniit. "A Strategy for Archiving Websites at the Library and Archives Canada presentation, Preservation of Electronic Records Symposium 2003" (copy of paper).

Hedstrom, Margaret. *It's About Time: Research Challenges in Digital Archiving and Long-term Preservation, Final Report, Workshop on Research Challenges in Digital Archiving and Long-term Preservation, April 12-13, 2002* (August 2003), http://www.digitalpreservation.gov/about/NSF.pdf.

Hedstrom, Margaret and Sheon Montgomery. *Digital Preservation Needs and Requirements in RLG Member Institutions*, (Mountain View, CA; 1988), http://www.rlg.org/preserv/digpres.pdf.

Heminger, Alan R. and Steven B. Robertson. "Digital Rosetta Stone: A Conceptual Model for Maintaining Long-term Access to Digital Documents", http://www.ercim.org/publication/ws-proceedings/DELOS6/rosetta.pdf.

International Organization for Standardization. "New ISO standard will ensure long life of PDF documents", http://www.iso.org/iso/en/commcentre/pressreleases/archives/2005/Ref974.html.

Iraci, Joe. *Modern Information Carriers (Participant Manual),* (Canadian Conservation Institute, 2000).

Jansen, Adam. "Digital Preservation in the 20th Century" presentation, 2005 Government of Alberta *Managing Information Assets in the Public Sector*, (18 October 2005).

Jones, Maggie and Neil Beagrie. *Preservation Management of Digital Materials: A Handbook* (British Library; London, 2001).

Kenney, Anne R. and Nancy Y. McGovern. *The Three-Legged Stool: Institutional Response to Digital Preservation* (March 2005), http://www.library.cornell.edu/iris/dpo/docs/Cuba-arknym_final.ppt.

Kenney, Anne R., Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze, and Sandra Payette. "Preservation Risk Management for Web Resources", *D-Lib Magazine*, vol. 8, no. 1 (January 2002), http://www.dlib.org/dlib/january02/kenney/01kenney.html.

Kenney, Anne R., Oya R. Reiger, David Block, Erla Heyns, Peter Hirtle, George Kozak, Joy Paulson, and Tom Turner. *Report of the Digital Preservation Policy Working Group on Establishing a Central Depository for Preserving Digital Image Collections* (Cornell University, March 2001), http://www.library.cornell.edu/preservation/IMLS/image_deposit_guidelines.pdf.

Koch, Traugott. "Metadata topics", *DELOS brainstorming* (Jaun les Pins 2005-12-05), http://www.ukoln.ac.uk/ukoln/staff/t.koch/pres/Brainst200512-MDc.html.

Krolicki, Kevin. "NASA Data Point to Mars 'Bugs,' Scientist Says," *The Long Now Foundation Press*, (updated 27 July 2001), http://www.longnow.org/10klibrary/darkarticles/ArtNASAMars.htm.

Lazinger, Susan S. *Digital Preservation and Metadata: History, Theory, Practice* (Libraries Unlimited, Engelwood CO, 2001).

LeFurgy, William G. "PDF/A: A new digital preservation format", *Society of American Archivists Newsletter* (May 2003), http://www.archivists.org/saagroups/gov/newsletters/May2003.asp.

Levy, David M. *Scrolling Forward: Making Sense of Documents in the Digital Age* (Arcade, New York, 2003).

Marill, Jennifer L., Andrew Boyko, Michael Ashenfelder and Laura Graham. *Tools and Techniques for Harvesting the World Wide Web,* (Library of Congress), http://csdl.computer.org/comp/proceedings/jcdl/2004/2493/00/24930403.pdf.

McKemmish, Sue and Wendy Duff. *Metadata and ISO 9000 Compliance,* http://www.sims.monash.edu.au/research/rcrg/publications/smckduff.html.

Podolsky Nordland, Lori. "Provincial Archives of Alberta's Digital Preservation Survey Summary", *Archives Society of Alberta Newsletter*, vol. 25, no. 2 (Winter 2006), http://www.archivesalberta.org/media/images/pdfs/Winter_Newslet_%202006.pdf.

----. "Provincial Archives of Alberta Digital Preservation Survey: Technology – Part I", *Archives Society of Alberta Newsletter*, vol. 25, no. 3 (Spring 2006), http://www.archivesalberta.org/media/images/pdfs/spring_2006_newsletter.pdf.

Read-Smith, Judith, Mary Lea Ginn, and Norman F. Kallaus. *Records Management, 7th Edition* (South-Western Thomson Learning, Inc., Mason OH, 2002).

Rothenberg, Jeff. *Avoiding the Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* (Council on Library and Information Resources, Washington DC, January 1999), http://www.clir.org/PUBS/reports/rothenberg/pub77.pdf#search=%22%22Avoiding%20the%20Technological%20Quicksand%22%20%22digital%20preservation%22%20emulation%20Jeff%20Rothenberg%22.

Sanett, Shelby. "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks", *RLG DigiNews*, vol. 7, no. 4 (15 August 2003), http://www.rlg.org/legacy/preserv/diginews/diginews7-4.html.

Van de Sompel, Herbert, Michael L. Nelson, Carl Lagoze, and Simeon Warner. "Resource Harvesting within the OAI-PMH Framework", *D-Lib Magazine*, vol. 10, no. 12 (2004), http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html.

van Wijngaarden, Hilde. "Different Approaches to Digital Preservation" presentation, National Library of the Netherlands, http://rdd.sub.unigoettingen.de/conferences/ipres04/van_wijngaarden/Different%20approaches%20to%20digital%20preservation.pdf#search=%22Encapsulation%20disadvantages%20%22digital%20preservation%22%22.

Vernon, David. R. and Oya Y. Rieger. "Digital Asset Management: An Introduction to Key Issues (Revised 2/02 – Revision 2), *Office of Information Technologies IT Architecture Initiative*, http://www.cit.cornell.edu/oit/Arch-Init/DigAssetMgmt.pdf.

Wheatley, Paul. *Institutional Repositories in the context of Digital Preservation* (Digital Preservation Coalition, March 2004), http://www.dpconline.org/docs/DPCTWf4word.pdf.