



**DRAFT**

**DIGITAL INFORMATION AT  
LIBRARY AND ARCHIVES CANADA:  
AN OVERVIEW OF PROGRESS AND ISSUES**

**Library and Archives Canada  
September 28, 2005**

# DIGITAL INFORMATION AT LIBRARY AND ARCHIVES CANADA: AN OVERVIEW OF PROGRESS AND ISSUES

INTRODUCTION .....	3
<b>1. ACQUISITION AND MANAGEMENT OF BORN-DIGITAL DOCUMENTARY HERITAGE .....</b>	<b>3</b>
<b>1.1. Status.....</b>	<b>3</b>
1.1.1. Mandate and policies .....	3
1.1.2. Selection .....	4
1.1.3. Scale of collection, rate of ingest .....	4
1.1.4. Metadata .....	6
1.1.5. Collection management .....	6
<b>1.2. Issues, Challenges and Opportunities related to born digital materials.....</b>	<b>6</b>
<b>2. DIGITIZATION OF DOCUMENTARY HERITAGE HOLDINGS .....</b>	<b>9</b>
<b>2.1. Status.....</b>	<b>9</b>
2.1.1. Mandate and policies .....	9
2.1.2. Funding, selection and planning .....	9
2.1.3. Scale of collection, rate of production .....	10
2.1.4. Production processes and standards .....	10
2.1.5. Descriptive Metadata .....	11
2.1.6. Partnerships.....	12
<b>2.2. Issues, Challenges and Opportunities related to Digitization .....</b>	<b>12</b>
<b>3. PRESERVATION OF DIGITAL DOCUMENTARY HERITAGE HOLDINGS.....</b>	<b>13</b>
<b>3.1. Status.....</b>	<b>13</b>
3.1.1. Mandate and policies .....	13
3.1.2. Inventory .....	13
3.1.3. Documentation .....	13
3.1.4. Preservation strategy and processes .....	13
3.1.5. Technical and Preservation Metadata .....	14
<b>3.2. Issues, Challenges and Opportunities Related to Preservation .....</b>	<b>14</b>
<b>4. ACCESS TO DIGITAL DOCUMENTARY HOLDINGS .....</b>	<b>15</b>
<b>4.1. Status.....</b>	<b>15</b>
4.1.1. Mandate and policies .....	15
4.1.2. Descriptive Metadata .....	16
4.1.3. Access currently offered .....	16
4.1.4. Access: Near and Future Directions.....	18
<b>4.2. Access Issues, Challenges and Opportunities.....</b>	<b>18</b>
<b>5. CONSIDERATIONS TOWARD A NATIONAL STRATEGY .....</b>	<b>20</b>
5.1. Overall .....	20
5.2. Digitization considerations .....	20
5.3. Digital preservation considerations .....	21
5.4. Access considerations.....	23
<b>CONCLUSION .....</b>	<b>24</b>

# DIGITAL INFORMATION AT LIBRARY AND ARCHIVES CANADA: AN OVERVIEW OF PROGRESS AND ISSUES

## INTRODUCTION

This document provides an overview of the current capacity of Library and Archives Canada (LAC) to collect, digitize, manage, preserve and provide access to digital information. It covers both digital content that LAC acquires, and digital content that LAC creates through digitization.

It has been prepared by staff from a number of parts of the institution. Contributors were asked to describe the current state of various areas in which digital capacity is required, and then to assess the main issues and challenges in a way that could be useful to LAC and other stakeholders as we collectively consider development of a Canadian digital information strategy. Contributors are listed at the end of the paper.

## 1. ACQUISITION AND MANAGEMENT OF BORN-DIGITAL DOCUMENTARY HERITAGE

### 1.1. Status

#### 1.1.1. Mandate and policies

The mandate of the Library and Archives Canada includes “to preserve the documentary heritage of Canada for the benefit of present and future generations” and to “be a source of enduring knowledge accessible to all, contributing to the cultural, social and economic advancement of Canada.”

The *Library and Archives Canada Act* that received Royal Assent on April 22, 2004 gave new collecting powers to LAC with regard to Internet publications. It mandates LAC to acquire and hold in its collection Canadian Internet publications under Legal Deposit, and, for the purposes of preservation, it allows LAC to collect a representative sample of Canadian websites. As in the past, the LAC Act also stipulates that LAC is to appraise the records of the Government of Canada (GoC) and preserve those that are archival. This responsibility includes the archiving of GoC records in electronic format.

Becoming a truly digital institution is one of the key objectives of the new institution. This means that LAC must become as adept in collecting documentary heritage materials in digital form as it has long been with analog media. In order to achieve this transformation, change will be required at many levels--policy, procedures and work processes, technology, and staff skills. Specifically, effort is underway to:

- establish priorities, and identify and allocate available resources;
- resolve policy issues and establish policies relating to the development of the digital collection (e.g. digital formats collected; parameters for acquisition of websites);
- review and update current LAC direction/guidance for creators of documentary heritage materials and in particular, federal government information;
- develop the necessary technical and procedural infrastructure for the acquisition of private and public sector e-publications, Government of Canada (GoC) e-records, and websites;
- develop a policy and a strategy for the creation of digital collection content and the creation of tools to make this content known and accessible;

- identify the skills needed by LAC staff, assess skills gaps and develop and deliver required training; and
- resolve resourcing issues to ensure sufficient human resources are in place for a sustainable digital collection program.

A Collection Development Framework has been drafted for the new institution. See <http://www.collectionscanada.ca/collection/024/index-e.html>

### **1.1.2. Selection**

The Collection Development Framework governs the selection and acquisition of digital materials. Under the Act, the acquisition of Internet publications as part of legal deposit is intended to be comprehensive, while the acquisition of web sites is intended to be representative ('Sampling from Internet', Section 8(2)). The two new clauses in the LAC Act are quite separate and distinct and, consequently, LAC must be clear about what authority is being invoked to select and acquire Canadian electronic resources. The Legal Deposit of Publications Regulations, which will take effect January 1, 2007, will further shape and limit what is acquired.

Government and private records, both digital and paper-based, are selected according to approved guidelines which are currently in place for acquiring such records. About 1% to 3% of the total GoC output is archived by LAC.

The Collection Development Committee is presently working on preparing a Collection Development Policy for digital collections at LAC, with a target completion date of Fall 2005. A document "Guidelines for the Selection of Canadian Web Resources" was drafted in July 2005 and is being used by the staff in the Legal Deposit Internet Unit in their daily selection work.

### **1.1.3. Scale of collection, rate of ingest**

A comprehensive report on the size of LAC's digital collections and the rate of ingest was prepared in March 2005.

#### *Internet publications*

LAC has been collecting Internet publications since 1994-95 when the former National Library began a pilot project using the Electronic Publications Pilot System (EPPS). Internet publications are acquired, catalogued and made available online through the AMICUS national database (<http://www.collectionscanada.ca/amicus/>), but also as a separate searchable site on LAC's web site. The current size of the Electronic Collection (<http://www.collectionscanada.ca/electroniccollection/>) is over 17,000 titles, over 64,000 serial issues, and millions of files. About 70% of the titles are federal government documents.

#### *Digital theses*

LAC currently holds 45,000 e-theses and last year, using OAI (Open Archives Initiative) protocol, harvested 700 e-theses from four Canadian universities. This annual ingest is expected to grow as more universities adopt an all-electronic approach to their theses programs and to the deposit of theses with LAC. The Theses Canada Portal (<http://www.collectionscanada.ca/thesescanada/>) and AMICUS as well as the Networked Digital Library for Theses and Dissertations at Virginia Tech University (<http://scholar.lib.vt.edu/theses/>) provide access to these resources. A further 11,000 Canadian theses are digitized under contract each year, but under the terms of the contract with the publishing company, they will not be available for free access for some time.

#### *Web sites*

Experimentation with the collection of web sites began a few years ago, first under the aegis of the Government On-line Task Force and now as a project under the current Digital Collection

Catalytic Initiative (DCCI). LAC has now acquired, catalogued and made accessible in the Electronic Collection and on AMICUS some 200 web sites. Due to its previous experience in handling digital on-line materials, staff in the Legal Deposit Internet Unit were ear-marked for handling web sites and, as of Spring 2005, the acquisition of web sites has been mainstreamed into the daily work of the Unit.

A project is currently adapting the International Internet Preservation Consortium (IIPC) web crawling bulk harvesting tool and is now experimenting with global harvesting of Government of Canada web sites. Our intent is to develop keyword searching capability and metadata extraction tools for bulk harvested sites. We will investigate a means to highlight key websites within bulk harvests so that they can be catalogued and made accessible in AMICUS.

#### *GoC electronic records*

As early as 1975 the Machine Readable Archives Division of the former National Archives was responsible for the appraisal, acquisition and preservation of nationally significant social science research datasets.

Today, LAC continues to appraise the records of the Government of Canada, including electronic records, databases and datasets, and archives those that are appropriate for long-term preservation. The volume of acquisition of e-records is to date relatively small although this flow will change as the record of government becomes increasingly digital.

Archival electronic records are generally acquired on physical media, then registered, accessioned, converted to preservation media and stored on tape. LAC is examining other methods of transfer including using the Internet and government Secure Channel for an online upload process. In addition LAC is examining the entire issue of access restrictions for government records with the hope that donor departments will indicate the accessibility of the record before transferring to LAC.

LAC's Government Information Management Office is leading development of a GoC records management metadata set which will in turn serve as the basic metadata for electronic records transferred to LAC.

#### *Digital audio*

LAC is currently investigating born digital audio and has initiated discussions with Musicrypt, a Canadian music distributor. In consultation with CRIA and CIRPA, major music associations, LAC is examining the notion of secure online acquisition of digital audio files which are designated as Canadiana.

#### *Digital television, radio, cinema and multimedia*

LAC captures broadcast media in real time off satellite on a daily basis. A pilot project using Virage software is currently underway to experiment with managing and storing digital video files. LAC must shift its focus from managing AV in physical format to managing it as file-based digital content.

#### *Digital photography*

(i) Government – LAC collection of digital photography in its GoC collection is very small. Only six signed authorities in the Records Disposition Authorities Control System (RDACS) system refer to electronic photographic records. It is assumed that the amount of digital photography acquired from GoC departments will increase in the future. Photography stored in proprietary database systems is likely to pose a significant preservation and access problem for LAC.

(ii) Non-governmental: Only a handful of collections and fonds containing electronic photography have been acquired, and some of these are not 'born digital' but are scans of negatives or prints held elsewhere in our collections. The amount of photography in our collections that has been

generated by “photo-quality” printers, or by commercial digital photographic printers, is wholly unknown.

#### *Geomatics data*

Geomatics records acquired from government departments includes geospatial data that encompass digital maps and hydrographic charts in both vector and raster data formats, synthetic aperture radar imagery and side-looking aperture radar imagery. Current quantity is approximately 3-4 Terabytes in size. Over the next year it will likely grow by a couple of hundred gigabytes.

#### **1.1.4. Metadata**

Metadata for born digital materials is generally created by LAC librarians, archivists, technicians or media specialists, and the process accounts for a significant resource expenditure for the institution. Given the resource-intensive nature of description, the notion of repurposed metadata, using metadata from the creators or from other metadata agencies is a long held tradition at LAC and continues for digital materials. New projects to acquire digital publications and electronic records all include a metadata component. Descriptive metadata is mandatory within the Terms and Conditions of electronic record transfer and will be part of the Regulations for Legal Deposit of Internet publications. For the long term, LAC is developing the capacity to use ONIX metadata from publishers, for conversion and loading to AMICUS.

In general, the descriptive and subject metadata standards employed for born digital acquisitions are the same used for traditional materials, for both published and archival holdings. See sections 1.2, 2.1.5, 3.1.5 and 4.1.2 for further discussions of metadata.

#### **1.1.5. Collection management**

Born digital collections are managed in different ways depending on whether the digital resources are online or off-line. (The same is true of digital collection items which have been digitized from analog sources.)

The Electronic Publications Pilot System (EPPS) has been used to acquire and manage online Internet publications (and now web sites) since the mid 1990's, but this system is not scaleable and staff use many labour-intensive manual procedures to acquire and manage these digital files. Implementing legal deposit for Internet publications will require a new digital content management system to cope with the increased volume and manage the burgeoning collection effectively.

Born digital acquisitions from GoC departments and private donors are copied to tape and stored in the vaults at GPC.

Born digital video is being managed in a pilot system using Virage software.

### **1.2. Issues, Challenges and Opportunities related to born digital materials**

Some issues, challenges and opportunities are specific to born digital e-resources, but many are generic to all types of digital materials. These include:

#### *Building human resource (HR) capacity*

Currently, LAC lacks operational/technical capacity to mount, read and access some of its digital assets. It lacks tools and processes to manage and preserve access to digital content effectively; many processes are labour-intensive and not sufficiently automated to be scaleable.

Expertise in digital information management and preservation is uneven across operational areas.

LAC needs to build human resource capacity to handle the anticipated influx of digital materials to our collections with the overall shift to digital and the implementation of legal deposit for electronic publications in 2007. The operational areas with responsibility for digital materials have made estimates of resource requirements and have submitted this information to LAC Management Board. Also, the training of staff to use the new technologies as they become available (and as LAC implements them) is of crucial importance.

### *Systems to handle digital materials*

Within the past year, LAC struck a number of transformational activities, including the Digital Collection Catalytic Initiative (DCCI). Its mandate, outlined in LAC *Directions for Change*, is to begin mainstreaming digital--organizationally, culturally and technologically. Along with examining significant digital policy and resourcing issues, the DCCI has identified that the institution must develop corporate and robust technical capacities, standards and best practices for acquiring, managing, preserving and making accessible Canada's digital documentary heritage. Working with operational areas throughout LAC, the group is developing functional requirements and projects for the implementation of systems which can better acquire and manage our wealth of digital content.

### *Policies*

LAC must determine organizational responsibilities and inter-relationships so that a holistic, comprehensive and cohesive approach toward handling digital materials can be taken. This fall, LAC will begin to develop new policy and practices for the selection and acquisition of digital materials (under the aegis of the Collection Development Committee) and for the description of Web-based materials (under the Intellectual Access Committee).

There is existing guidance for acceptable file formats in "Guidelines for Computer File Types, Interchange Formats and Information Standards" ([www.collectionscanada.ca/information-management/0612/061204\\_e.html](http://www.collectionscanada.ca/information-management/0612/061204_e.html)). While currently directed at GoC departments, this document is being updated and broadened to serve a broader range of digital information producers.

### *New work processes*

Intricately related to new policies and new systems is the development of new work processes. While traditional media have been described for access by LAC staff, increasing volumes of digital information suggest that the need to look for automated processes such as metadata extraction tools and full-text search, and new approaches such as the automatic ingest and use of producer-generated metadata included with the digital item itself. The Published Heritage branch has conducted a benchmarking study of innovation and work practices at peer national institutions in other countries.

### *Becoming a Trusted Digital Repository (TDR)*

In August 2005, RLG issued a draft document "An Audit Checklist for the Certification of Trusted Digital Repositories" (<http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>). This draft is open for public comment until 15 January 2006 and LAC is considering using this tool to evaluate our institutional readiness as a TDR. The Reference Model for an Open Archival Information System (OAIS) provides a *de facto* standard and base outline for assessing technical readiness, and LAC is currently mapping its digital processes to this model.

There are also issues that are specific to born-digital materials. These include:

### *Distinguishing e-publications, e-records, and web sites and providing guidance on managing and archiving them*

Over the past few years, LAC has received many requests from other federal government departments, provincial ministries, universities and libraries for assistance in determining both policies and practices for archiving web content, and have been asked what is an Internet publication versus a web site versus a GoC e-record. LAC is actively discussing and working through these issues as part of its overall collections policy activities.

### *Frequency of capture*

LAC must establish policies and procedures for determining the frequency of capture for born digital materials which are often frequently updated (e.g. websites, newspapers, geomatic data). It must define the optimum moment at which a given type of born digital publication should be acquired, and be able to track publication patterns (if they exist) to support additional captures as necessary reflect major changes over time.

### *Rethinking metadata*

The creation of metadata for born digital material is costly and time-consuming. The new mantra may be 'acquire all, describe some'. This is a major access/cataloguing policy issue which raises many questions. Should LAC use keyword searching to make vast amounts of born digital materials accessible and then only catalogue a selection of seminal e-resources? How will this be determined and decided? Due to the use of sophisticated search engines, Internet publications do not require the same types of descriptive elements as analog publications. Will LAC use our own version of an 'access record' as has been recently put forward by Library of Congress?

LAC is also looking at ways of using creator or producer-supplied metadata. Under the new Act, LAC now requires e-publishers and GoC departments who submit e-records to submit metadata when they submit their e-materials. Currently, student authors of electronically-harvested Canadian theses provide basic metadata including subject keyword descriptors and an abstract; these are converted to MARC records and included in AMICUS.

LAC also provides some Cataloguing-in-Publication (CIP) data to publishers for digital publications, on request.

### *Contact with information producers*

LAC has spent time and energy in contacting and building good relationships with both government and non-government information producers, and these must be continued and strengthened so that, for digital materials, sound practice in content creation, transfer protocols, and metadata can be fostered early in the lifecycle.

### *Decentralized collecting responsibility*

Can LAC establish partnerships that will allow the collecting of born digital materials to be shared with other institutions? What attributes and responsibilities would this imply for both LAC and those partners? There is particular interest in pursuing partnerships for digital provincial government publications. (See "Becoming a Trusted Digital Repository" below).

### *Acquiring content in databases*

Collections of digital content that may merit preserving at LAC are often stored in and accessible only through proprietary databases. Currently, some archival databases have been acquired as flat ASCII files that can be read by any database software. LAC does not currently acquire publicly-accessible databases as part of the Electronic Collection.

### *Permission to preserve and allow access*

LAC clearly has the mandate to *acquire* Internet publications and web sites, but currently asks permission to *make accessible* both government and non-government digital material in our

repository (even though in most cases the material is already publicly accessible). This is a time-consuming process. LAC has asked Public Works and Government Services Canada (PWGSC) for a blanket permission that would permit us to make accessible without requesting individual permissions all GoC Internet publications and web sites acquired. How to address this issue with non-governmental publishers and web site owners remains to be determined.

## 2. DIGITIZATION OF DOCUMENTARY HERITAGE HOLDINGS

### 2.1. Status

#### 2.1.1. Mandate and policies

In order to be “a source of enduring knowledge accessible to all, contributing to the cultural, social and economic advancement of Canada” as set out in the *LAC Act*, the institution has taken full advantage of digital technologies to make our holdings available. The new organization is continuing to invest in digital conversion technologies to promote awareness of and access to Canada’s cultural heritage.

Corporate LAC policy is under review and development for all areas that touch on digitization, including Collection Development, Intellectual Access, Rights Management, Information Technology and Preservation. The following basic principles will be reflected:

- Support effective and efficient creation of digital collections with a common and scalable service framework and technical architecture.
- Implement widely-recognized best practices and standards to create digital collections with long-term viability.
- Monitor developments in digitization technologies, standards, and practices.
- Select and create digital collections with enduring value.
- Augment LAC’s digitized collections with content from donors, creators, and government departments.
- Participate in studies that assess the role of digital collections in supporting learning, teaching, and research.
- Promote financial accountability and business planning.<sup>1</sup>

#### 2.1.2. Funding, selection and planning

Digitization projects are funded primarily by Canada Culture Online Program (CCOP) (<http://www.pch.gc.ca/progs/ccop/>), through its Canadian Memory Fund. This fund is devoted to federal institutions including all federal departments, agencies and Crown corporations holding key collections relevant to Canada’s culture and heritage.

Digitization project initiation, proposals, cost estimates and implementation plans are most commonly led by the Programs and Services Sector in collaboration with Care of Collection Branch and content specialists. Project selection is guided by Canadian Culture Online selection criteria. Additional selection criteria include the condition of the materials, copyright, format of the material, existence of other copies, user needs and interests, and capacity of the digitization platform.

Digitization is also undertaken on as-requested basis as a result of Access to Information and Privacy (ATIP) and Interlibrary Loan requests. In addition, digitization is undertaken as planned

---

<sup>1</sup> Cornell University Digital Production Services <http://dcaps.library.cornell.edu/index.htm>

migration for analog materials, such as audio-visual holdings. Migration of analog materials to digital formats serves both preservation purposes and fosters access. Recent work by the Digital Collection Catalytic Initiative has produced a decision tree for digitization that rolls in factors such as user needs, requests for materials, and condition of the original.

### **2.1.3. Scale of collection, rate of production**

LAC has been developing a digital platform that is capable of creating millions of digital objects per year. It is expected that the current collection of digitally converted collection materials will rise from the current figure of over 4 million images and a production rate of ca. 800,000 images per year, to yearly additions of 2 to 3 million digital objects. The digitization of books and audiovisual materials will grow the collection to petabytes of data. (Moving image files are particularly huge: one hour of broadcast video equals 108 gigabytes and LAC has approximately 100,000 hours of it to convert.)

The LAC digital collection is already immense, at approximately 856 terabytes. The certain (although not yet quantified) growth in the acquisition of digital materials in the coming years will greatly increase this already large-scale collection.

### **2.1.4. Production processes and standards**

The Library and Archives of Canada has made considerable investment into a digital platform that supports digitization for three primary purposes: planned collection-based digitization for preservation and access, thematic project-based digitization to support access and interpretative programs, and on-demand digitization to support service. Most of the digitization services are supported within the Care of Collections Branch of the Library and Archives Canada. Other digitization efforts are within Access to Information and Privacy, Interlibrary Loans, and the Cartography, Architecture and Geomatics areas.

LAC is very much “digital” in a number of specialized areas. The digital environment within LAC is fundamentally divided into the following specialized areas: digital video, digital audio, aspects of moving images (converting film to a digital video signal), document imaging, and finally digital imaging. LAC generally follows a standard digitization process that involves nine functions:

- Prepare (may include conservation treatment)
- Benchmark
- Digitize (actual stage of conversion)
- Process (includes all post-conversion procedures)
- Metadata (some preservation and technical metadata is captured; descriptive metadata is created or pre-existing records are adjusted)
- System (fit into Information Technology architecture)
- Interface (application development, web development)
- Deliver (create derivatives, determine use of plug-ins or specialized applications, and manage files)
- Access (consider different access channels and usability)

In addition, the digitization production process is further tailored based on particular collection material types. An internal document provides a complete overview of the digitization guidelines, “Technical Guidelines for Digitizing Archival Library Collections Materials for Electronic Access: Creation of Production Master Files – Raster Images”. Production processes and specific digitization workflows have been optimized for the following collections:

*Microfilm Collection*

At 800,000 plus images per year, the majority of digital objects created to date have been reproduced from the microfilm collections. The work processes and workflow have evolved as our experience and understanding of the requirements and technology have matured. The digitization of microfilm is considered a specialized area of mass digitization within the digital platform. All aspects of the production process are handled within the same unit and are uniquely optimized to create the web-ready content. Digital scanning equipment/software is also optimized for the scanning of microfilm reels.

#### *Print and Drawings, Fine Art, and Photographic Collections*

Due to the complexity and nature of these types of collection materials the work processes and guidelines for capture depend upon the purposes the digital object will serve. The work processes are divided into capture, file management, quality assessment, post-processing and derivative file creation. Each work process is managed by specialty teams to ensure control and efficiency. The digital capture technology that is utilized is very broad, ranging from inexpensive consumer-type flatbed scanners to highly-specialized digital cameras and digital camera backs. The range of equipment is necessary due to the various preservation, access and client demands.

#### *Cartographic Collection / Oversize Materials*

These types of collection materials have very specific attributes and thus the work processes and technical requirements are unique to the materials being captured. The digital equipment is highly specialized and designed specifically to capture very large documents at very high resolutions.

#### *Audio and AV Collection Materials*

A longterm migration program is underway to convert audio-visual records to digital form in order to avoid the obsolescence of their original analog format. As well, the collection of Canadian 78 RPM musical recordings is being digitized for preservation and access through the Virtual Gramophone website (<http://www.collectionscanada.ca/gramophone/>).

#### *Bound Volumes and Serials*

Over the past year, LAC has been working with the Internet Archive and University of Toronto Libraries to test automated page-turning scanning systems to digitize monographs and journals into high-resolution files and to automatically correlate these digital images with descriptive metadata. Technological options are continually developing and we believe that recent advancements in book scanning technologies and demands for enhanced access to bound volumes require LAC to continue to seek to increase production rates for published volumes, and to build in-house capacity in this area.

#### *Textual Documents*

LAC recently commissioned a Document Lifecycle Strategy Assessment to examine the workflow processes and costs associated with producing photocopies of textual records for research and personal usage. The primary goals were to achieve greater process efficiency, reduce turnaround time and achieve faster document delivery, and manage costs effectively.

The report's findings and recommendations support LAC's desire to move towards a high-volume enterprise-wide electronic document management, content management and document imaging solutions to create, manage, store and retrieve textual heritage collection materials.

### **2.1.5. Descriptive Metadata**

Descriptive metadata for LAC digitized images is usually based on the existing descriptive metadata for the original materials: e.g. item and file level descriptions for archival holdings. Descriptive metadata for published materials in AMICUS will be leveraged for digitized books

and has been used for digitized Canadian sheet music. The objective is to repurpose existing metadata by simply adding a link to the digital object on the existing metadata record for the original material. Ideally, this process would be automated.

More discussion of LAC practices related to descriptive metadata is found under Access (Section 4.1.2).

### **2.1.6. Partnerships**

A collaborative approach to digitization has been adopted by LAC. We have partnered nationally and internationally with organizations such as University of Toronto, Archives nationales du Québec, canadiana.org, Statistics Canada, Direction des Archives de France, Public History Inc., and many more. Most recently, the National Archives of Ireland has proposed a partnership to facilitate the digitization and indexing of the 1901 and 1911 census records.

## **2.2. Issues, Challenges and Opportunities related to Digitization**

### *File formats*

File format choices at the time of digitization have long-term implications. Currently, LAC uses Tagged Image File Format (TIFF version 6.0) for production master files that are uncompressed, and then it also creates and stores up to ten derivative versions. It is looking instead toward creating and storing one master file such as JPEG 2000 from which all other file formats to support usage can be rendered. But any change in practice will also require changes to our applications that manage and serve the files.

### *Aging digitization platform*

The LAC has built a digitization platform that has proven to be robust and productive. However, new technologies, hardware, digital file formats and workflows need to be introduced and streamlined with our changing delivery systems and increased appetites for digital content online.

### *Efficient workflows*

Achieving the maximum production rate at the lowest possible cost is an ongoing challenge, and merits ongoing applied research effort. For example, with bound volume digitization, LAC has been exploring with University of Toronto and other partners how to maximize the automation of both scanning and metadata processes in order to achieve a viable cost point for a more ambitious book digitization project than has previously been conceived for Canada.

### *Selection policy*

LAC's current digitization activities, although significant, are not framed in a coherent institutional or national strategy. Such a strategy will need to take into account user demand, needs and interests; Copyright; the suitability of material for conversion in terms of condition, size, etc.; and a host of other factors.

### *Funding*

While Canada Culture Online Program (CCOP) funding has resulted in the development of some very important LAC digital collections for Canadians, the degree of our dependence on a single source of external funding affects the both the scope and scale of our projects and threatens the long term viability of the LAC program.

### 3. PRESERVATION OF DIGITAL DOCUMENTARY HERITAGE HOLDINGS

#### 3.1. Status

The former National Archives began acquiring digital information and converting it to standard format in 1974. In the early 1990s, the former National Library began acquiring published electronic information on CD, diskette etc., but did no evaluation or migration of this content, partially owing to Copyright. It added online electronic publications in 1993, and both former organizations, before the development of LAC, had experimented with the acquisition of websites.

##### 3.1.1. Mandate and policies

LAC's mandate supports preservation of digital material, and the new LAC Act strengthens its acquiring and preserving role through provisions such as the ability to "rescue" at-risk government records and to acquire online electronic publications through legal deposit.

The preservation policy of the former NL (1989) encompassed digital assets, while the preservation policy of the former NA (2001) distinguished between "technology-dependent" records (including digital assets) and "human-readable" records and incorporated preservation principles based on this distinction. When the new LAC preservation policy is developed, it will likely retain the distinctions inherent in the NA policy.

In addition to the Preservation Policies, both former organizations had developed policies specific to electronic assets; *Networked Electronic Publications: Policy and Guidelines* (NL) and the *Electronic Records Policy* (NA) were written in the 1990's but only partially implemented.

##### 3.1.2. Inventory

The Digital Collection Catalytic Initiative has prepared a partial inventory of types and amount of digital objects at LAC--and the range is large--but there is no single complete inventory. Assets are held in different databases, on different servers, under the control of various operational areas. The size of the collection is variously reported in terms of storage space, or in terms of the number of titles (publications).

##### 3.1.3. Documentation

Internal procedures are well-documented for archival records. The value of digital records is appraised using the same criteria that are applied to other records. The transfer of digital records of the Government of Canada is detailed in written terms and conditions documents, whereas acquisition of digital records from private sources is more informal. Guidelines exist regarding preferred file formats (Archives), and best practices for publishers (Library). Written acquisition agreements exist with creators of published electronic records.

##### 3.1.4. Preservation strategy and processes

Conversion of digital records into standard logical formats, and their subsequent migration as necessary over time, has been articulated as LAC's digital preservation strategy. In practice, conversion is not always implemented at the operational level. Nonetheless, LAC recognizes that all its preservation work with digital material must be based on non-proprietary standards.

Similarly, LAC recognizes that digital materials not stored on servers must be kept on physical storage media that are not at risk of imminent obsolescence. Any newly-acquired materials on formats such as diskettes or older forms of tape storage are transferred to current physical

formats such as CD, DVD or digital linear tape. Over time, all materials must either be transferred to a server environment or regularly moved onto newer physical formats as storage technology evolves. LAC has backlogs of material requiring logical format conversion, physical format conversion, or both.

LAC's IT infrastructure supports the minimum requirements for digital preservation (e.g., by IT standard security and backup procedures).

LAC has huge collections of television, radio, cinema and multimedia in many formats. Equipment for acquiring and managing AV content at LAC is aging and the existing infrastructure is geared towards handling AV in physical formats. There is a high risk of deterioration and loss of media and content due to chemical processes and the disappearance of expertise and playback systems. The media are not eternal and specific actions need to be taken in order to guarantee the survival of this heritage for future generations. LAC must initiate ambitious and expensive plans to preserve the material in digital form. It is clear that, unless some major programs are developed, most of the audiovisual content will not be saved on time, due to the cost of preservation and digitization and the time needed to realise this work. Within the Information Technology Branch, the Digital Media Preservation Technology Section has developed a digital preservation infrastructure for AV holdings to improve the quality of preserved digital content and to reduce dependency on old formats/equipment.

### **3.1.5. Technical and Preservation Metadata**

Preservation metadata is not systematically collected for acquired digital material; however, different metadata is collected in various operational areas according to the requirements of the documents. Records of conversions are routinely kept. Metadata that will support preservation is captured for some publications as part of the descriptive process, but it is incomplete.

In the realm of digitizing still images, the Digital Collection Catalytic Initiative has introduced an XML Schema and accompanying *Data Dictionary – Technical Metadata for Digital Still Images* (NISO Z39.87, AIIM 20-2002) for use by the Care of Collections Branch for digitizing still images. Standardizing the information will allow users to develop, exchange, and interpret digital image files. The dictionary has been designed to facilitate interoperability between systems, services, and software as well as to support the long-term management of and continuing access to digital image collections. While such technical metadata is only a subset of the complete suite of preservation metadata necessary to achieve the long-term viability of a digital asset, it has often been called the first line of defense against losing access. Technical metadata assures that the information content of a digital file can be resurrected even if traditional viewing applications associated with the file have vanished. Furthermore, it provides metrics that allow machines, as well as humans, to evaluate the accuracy of output from a digital file. In its entirety, technical metadata supports the management and preservation of digital images throughout the different stages of their life-cycles.

There remains the need to develop a corporate metadata set that addresses the preservation of both born-digital and digitized content. The PREMIS (Preservation Metadata Implementation Strategies) Data Dictionary (<http://www.oclc.org/research/projects/pmwg/>) will be reviewed by a working group of the LAC Digital Preservation Working Group with the view of analyzing current preservation data elements in use, current preservation systems and adopt a common set of technical/preservation metadata elements to be implemented at the corporate level.

## **3.2. Issues, Challenges and Opportunities Related to Preservation**

A Digital Preservation Working Group now provides a forum for information exchange, creation of a metadata standard, development of systems specifications, pilot projects, etc. Subgroups will focus on areas of particular interest (e.g., persistent naming, which is underway). LAC has the opportunity to create an environment that will allow the migration, preservation and repurposing of its electronic and digital assets independent of formats and associated technologies to ensure current and long-term access.

LAC faces a number of challenges related to digital preservation, including:

#### *Managing and preserving both digital and legacy collections*

To date, the advent of digital has not meant that the need for LAC to acquire, manage and preserve traditional library/archival material has diminished. Even if we might anticipate that, at some point in the future, all acquisition may be of digital information, there will always remain several hundred years worth of analog content in the LAC collection. LAC must allocate the required resources to preserve new digital materials while still managing legacy collections in traditional media.

#### *Preservation metadata*

A local metadata set is currently applied to portions of the collection, but LAC is currently examining and benchmarking PREMIS to bring a corporate, cohesive approach. Whether preservation metadata will be stored with descriptive metadata in systems such as AMICUS or in the future Digital Content Management systems has not been determined.

#### *Automating preservation functions*

Appropriate metadata, registries of file formats, and system to support automatic migration or emulation are prerequisites to the long-term goal of automating preservation functions.

#### *Authenticity and provenance*

Consideration of the preservation of authenticity in electronic materials from all sources has to be integrated into LAC operations, beginning with selection decisions. This is difficult. For example, the maintenance of the provenance of digital photography in a composite archival record is much more difficult than with conventional photography as it is far easier to create a copy of an image file. Work on this issue is being undertaken by InterPARES II (see [www.interpares.org](http://www.interpares.org)). It is also significantly more possible to manipulate and alter digital content, raising the possibility of unauthorized manipulations or repurposing.

## **4. ACCESS TO DIGITAL DOCUMENTARY HOLDINGS**

### **4.1. Status**

#### **4.1.1. Mandate and policies**

One of the LAC's *Directions for Change* is "Access is the Primary Driver." It is partially described as:

Users' actual and anticipated use of our resources must be explicitly brought to bear on all activities undertaken by the institution. The implications are profound. LAC will acquire for access, preserve for access, describe for access, digitize for access, drive policy toward access, innovate with technology for access, and ensure that the ways in

which we provide access effectively meet users' needs. Everybody at LAC is in the access business.”<sup>2</sup>

LAC has begun extensive policy work toward an overarching Access Policy that will include in its scope provisions for access and copying of digital resources.

#### **4.1.2. Descriptive Metadata**

LAC is in the process of developing a Metadata Policy Framework which will outline all standards used for the management of all parts of the collection. The primary descriptive standards are AA2/Anglo-American Cataloguing Rules 2<sup>nd</sup> edition, Rules for Archival Description and MARC : Machine Readable Cataloguing. Some work has been done with XML EAD: Encoded Archival Description, XML MARC, Dublin Core, and Electronic Theses and Dissertations Dublin Core (ETD-DC). Our current redevelopment of our corporate metadata repositories will rely heavily on the Metadata Object Description Schema (MODS) to ensure flexible and fast search and discovery.

File level descriptions within government holdings rely almost exclusively on the donor departments in the form of databases for record profiles. That metadata will become more standardized as the Government of Canada adopts the draft standard for Records Management Metadata.

A Metadata/Thesaurus Services group within the LAC/Documentary Heritage Collection Sector provides high quality metadata to partner government institutions to augment their own digital exhibits and programs. An example is the Canadian Military History Gateway (<http://www.cmhg.gc.ca/>). The Service employs such standards as Dublin Core and controlled vocabularies such as FAST, and this service is provided on a cost recovery basis.

#### **4.1.3. Access currently offered**

Access to the digital holdings of the Library and Archives Canada is offered through a combination of online catalogues, full text engines and specialized LAC websites.

##### ***AMICUS***

Published web resources are acquired, catalogued and made accessible using AMICUS, the LAC integrated library management system. AMICUS itself is offered through numerous presentation protocols such as AMICUS OPAC, AMICUS WEB and AMICUS Z39.50. Widely-used metadata standards such as AA2, MARC21, LCSH, CSH, RVM, Dewey, ISBN and ISSN are used. AMICUS offers access to the published digital holdings of LAC, but as well to the tens of thousands of digital holdings reported by Canadian libraries to the National Union Catalogue.

##### ***GenAMICUS***

Access to some digital collections is offered through portals of AMICUS, developed according to theme and genre. This allows the AMICUS infrastructure, search engines and metadata to be repurposed, and provides users with both integrated access through AMICUS and specialized access through the more specific LAC web product. Some of the major collections are:

- **Electronic Collection of the Library and Archives Canada:** This site includes more than 17,000 Canadian digital titles, more than 64,000 online serial issues and 200 websites published by both the commercial and the government publishing sectors. These digital resources are also accessible via full text searching within this application (<http://www.collectionscanada.ca/electroniccollection/>).

---

<sup>2</sup> Library and Archives Canada “Directions for Change” (to be published shortly)

- **Federal Publication Locator:** Provides access to bibliographic records of Canadian Federal Government Publications in the Library and Archives Canada (LAC) Catalogue, including electronic Canadian Federal Government Publications (<http://www.collectionscanada.ca/7/5/>).
- **Theses Canada Portal:** Provides access to the theses collection of the Library and Archives of Canada, established in 1965. Includes access via AMICUS records and full text searching for over 45,000 electronic theses. These electronic theses are also accessible via the LAC Open Archive Initiative (OAI) repository and are harvested by the Networked Digital Library for Theses and Dissertations (<http://www.collectionscanada.ca/thesescanada/>).
- **Virtual Gramophone:** A multimedia website devoted to the early days of recorded sound in Canada that includes 4000 fully-restored digital audio reproductions of selected 78 RPM recordings (<http://www.collectionscanada.ca/gramophone/>).

### **Archivianet**

Archivianet (<http://www.collectionscanada.ca/archivianet/>) is the online research and consultation tool providing access to the archival records of LAC through description and metadata uploaded from numerous internal databases. The metadata standards employed include Rules for Archival Description, internal thesauri and repurposed finding aids from private donors and federal government departments. Archivianet provides access to born electronic records stored offline and to the digitized online records of LAC. In general, the metadata for the analogue resource is repurposed for the copied digitized holding.

The LAC digital archival landscape is vast. Some important digital holdings accessible via Archivianet are:

- **Census Research Tools:** Includes Census of Canada, 1911 and 1901.
- **Government of Canada Files:** Includes access to 50,000 digital images including Cabinet Conclusions, Métis Scrip, Orders in Council (1867-1882) and War Diaries of the First World War.
- **Photographs:** The collection includes over 22 million photographs and online access is available to over 25,000 digital reproductions.
- **Maps, Plans and Charts research tools:** Includes Electoral Maps of Canada, Indian Reserves – Western Canada and Maps, Plans and Charts.

### *LAC Website*

The LAC website, arranged according to our collections, services and users, is a first point of contact for many LAC clients searching for digital information. Within each domain the user is offered access to a variety of digital resources including sites that involve many LAC partners. Examples include:

- **Images Canada:** Hosted by LAC, this site provides central search access to thousands of digital images held on the websites of 14 Canadian cultural institutions. Search is possible across all collections with a simple keyword (<http://www.imagescanada.ca/>).
- **Canadian Information by Subject:** This is an information service, developed and maintained by LAC, which provides organized links to information about Canada from over 8500 web sites around the world (<http://www.collectionscanada.ca/caninfo/>).
- **Dictionary of Canadian Biography Online:** Developed in partnership with Université Laval and University of Toronto, this rich resource provides detail about the people who

played an important role in the formation of Canada (<http://www.biographi.ca/index2.html>).

- **Canadian Genealogy Centre:** Provides a single point of access to genealogical digital information at LAC and across Canada (<http://www.collectionscanada.gc.ca/genealogy/>).

#### **4.1.4. Access: Near and Future Directions**

Over the coming months LAC will be implementing concrete changes which will demonstrate the new institution's focus on access. Central to this will be the availability of and provision of access to digital content.

##### ***AMICAN***

AMICAN is the new integrated holdings management system under development which will include a public search module offering Canadians and the world access to the combined published and archival holdings of LAC as well as to the holdings of Canadian libraries. AMICAN is the combination of AMICUS and MIKAN, LAC's integrated archival holdings management system. Access to digital holdings will be emphasized from the perspective of search design, indexes, search results and display. Currency will be enhanced as users will be offered access to an item's or collection's metadata as soon as the information is created. The metadata standards for digital holdings will continue to be those currently employed by AMICUS and MIKAN, however new standards are coming into play. The public search module will benefit in terms of speed and flexibility from an XML cache, created from MARC21 and local data elements using MODS 3.0 (Metadata Object Description Schema) and extensions.

##### ***Federated Search***

LAC is implementing federated search technology which will provide clients with a resource discovery tool that will be the first-stop access to LAC's rich collections and services, offering simultaneous searching of our comprehensive published and unpublished collections. The single search window of a federated search will identify and manage the results of targeted resources including LAC catalogues, union catalogues, digital collections, finding aids, nominal databases, images, site search (webpages), and partner databases.

##### ***Google Space***

LAC is currently developing policy which will govern future directions for exposing our rich digital content and metadata to Google search engines, thus increasing access to information for all Canadians. The first collection which will be crawled and indexed will be the Electronic Theses Collection of LAC, most likely using LAC E-Theses OAI metadata repository as the portal.

##### ***3-D Space***

LAC is exploring new learning spaces with the Portrait Gallery of Canada (<http://www.portraits.gc.ca/>) to connect Canadians to historic and contemporary exhibits through the virtual network of the Web.

#### **4.2. Access Issues, Challenges and Opportunities**

##### ***Search***

Traditionally, large quantities of digital material have been accessed through database formats using either metadata and/or full-text indexing of the digital resource itself. HTML pages, indexed through search engines such as Google, provide another means to access content online.

Issues include:

- the costs and time involved in the creation of metadata for Internet-based resources;

- enormous result sets and the effective clustering of results;
- the failure of search engines to search embedded Web metadata;
- the impenetrability of Web-based databases, unless marked up, to the popular Web search engines.
- the usefulness of written metadata when searching visual, or multi-media items and the development of non-textual search mechanisms.

### *Presentation*

After a digital collection is found, several questions then arise in how the digital resources should be presented. Some key issues that should be addressed include:

- Are different file formats and quality levels necessary to support different access devices, or can different presentation formats be rendered from a single high-quality source file?
- Should books and documents be made available in a page turning format that emulates the original media?
- Should artifacts be presented in a manner that allows a three-dimensional inspection online?
- Should digital surrogates of multiple media items be presented in a multiple media ways online?
- Should born digital content be presented in a way that emulates its original format? Or, should these files be presented in other formats?

### *Systems to manage access, rights and restrictions*

When providing access to digital materials, LAC must comply with Copyright, Access to Information and Privacy, and donor agreement provisions. At the present time, LAC supports only two levels of access to its online digital collections: either open access to all Internet users, or access restricted to on-site clients only in the LAC Reference Room. Publishers and donors must be consulted about what type of access they will allow, and LAC must develop systems and processes capable of handling in an automated way a variety of various access levels (for example, perhaps older serials issues are openly accessible but current issues are not).

### *Persistent naming*

Having a unique and persistent naming convention for digital resources is a key component of a reliable digital information infrastructure and is important for sustaining access to information. Without such persistence, Canadians who use the Internet to access government services and digital collections may be frustrated by the system's inability to find the digital object. The issue of persistent identification is a critical component of the anticipated Digital Content Management infrastructure, and a key goal is to develop a coherent and holistic method of providing unique identifiers for LAC online resources.

### *Bandwidth*

Storage space, cataloguing and retrieval are only the tip of the digital iceberg; transferring data at the required speed also becomes an important requirement. Secure and reliable networks, with sufficient bandwidth, are needed. As LAC makes more film content available through digitization, demand for the files will take up much of our available bandwidth, challenging our ability to deliver all other content and services.

### *Access to digital resources not held*

Should LAC offer links from our web site and/or access catalogues (e.g. AMICUS) to born digital materials held by other organizations? The Collection Development Committee will be looking at this during its policy work on digital collections in the Fall 2005

## 5. CONSIDERATIONS TOWARD A NATIONAL STRATEGY

*Note: This section simply reflects some of the observations noted by various contributors to this paper; it is not a comprehensive set of considerations.*

### 5.1. Overall

There will need to be a governance and accountability framework and infrastructure for national, provincial, and local/institutional responsibilities.

### 5.2. Digitization considerations

#### *National digitization strategy*

Planning on a national scale would be an enormous challenge. The mandate, mission, objectives and overall policy structure would need to be in place before tackling the grittier issues of selection, standards, technologies, logistics, content storage, content sharing, transfer protocols, preservation and access portals.

Roles and responsibilities would need to be clearly defined: who participates, who leads, who delivers, what are the timeframes. Sustainable funding is key to the success of the strategy as well as to the future of the LAC digital program.

Selection criteria will be a major policy consideration. Criteria may differ across the nation based on content specialization, cultural and geographic considerations, content formats, copyright, available technologies, user needs, preservation criteria and funding. What do users want and how will their perspective be engaged? Will the CDIS focus on “Canadiana” or will it include all digitized cultural content in this country? Will we try to minimize duplication and coordinate content or will each institution make its own decisions? A blending of both approaches may be needed.

Now may be the opportunity to identify the current **national inventory** of our collective digitized holdings. AMICUS (which includes canadiana.org), CAIN, and the CIDL inventory could be sources to begin an inventory. When thinking of national content one is reminded of the early Conspectus work, and the extraordinary resource drain this produced on participating libraries. We need to mitigate these huge resource implications for all participants.

This is a great opportunity to develop a real “national collection” of digital materials. If providing centralized access, would this be considered an extension of the LAC collection itself, or would it best reside in our union catalogues – AMICUS and CAIN?

#### *Production processes and standards*

Technical infrastructures, repositories and access tools will be prime considerations. Will we have a distributed national network of holdings with one or multiple access portals? Will there be a centralized content delivery repository? Will participants share preservation responsibilities locally or will preservation masters be sent to LAC? If we are sharing content, what technologies and standards will be used to deliver that content?

Large volumes of data require strong networks and bandwidth. LAC and other institutions in Canada have experience with protocols such as OAI for moving metadata and content, but high volume is an impediment for even this roadworthy protocol. The simplified method of transferring data via offline modes, such as hard disks, requires tremendous processing and hours of manual labour.

Do institutions have the technologies in place to manage their digitized holdings – networks, bandwidth, storage, data management software? Will all types of repositories, formats and access tools be welcome or will national standards influence local implementations?

LAC is in a strong position to share standards and procedures for handling originals and for digitization. Likewise, the Digital Media Preservation Technology area provides efficient technical services for long-term for digital preservation of all types of audio visual collections, and that expertise can be exported to other institutions. Roles can be complementary: while LAC is well positioned to share expertise in digitization techniques, scanner evaluation and format standards, Canadiana.org has strong OCR capabilities.

### *Metadata*

LAC is in a strong position to provide guidance in metadata standards applied within the context of digitization – in the areas of descriptive, technical and preservation metadata.

For the digitization of published materials, it is likely that most libraries would want to use the existing metadata for the original – AA2, MARC and the exchange and search protocols already in place. Persistent identifiers remain an issue across all institutions and at this point pose a pressing institutional problem that we must resolve at the local level. LAC corporate review of Rights Management metadata will strengthen the LAC leadership role nationally.

## **5.3. Digital preservation considerations**

In developing a strategy for digital preservation, technology is one aspect; however infrastructure, laws and policies, standards and practices, people, management and governance all have a role to play.

### *Federal role*

A broad national strategy for digital preservation implies coordination and action across levels of government, associations and interest groups, and private stakeholders: a considerable undertaking, especially in an area where experience and expertise is widely distributed. The role of the Government of Canada in a digital preservation strategy needs to be defined, but will likely be expressed in the broadest of terms, with a focus on results for Canadians.

Models for national strategies may exist in other sectors (e.g. health, emergency preparedness), but there are few examples to look at in the cultural sector. However, a recent example to consider may be the Historic Places Initiative (<http://www.pch.gc.ca/progs/ieh-hpi/>), where the Department of Canadian Heritage and Parks Canada are implementing a pan-Canadian program of consultations, support programs, incentives and legislation to protect historic places. Federal objectives<sup>3</sup> for this program may be applicable to the preservation of digital resources as well, or at least they illustrate the nature of consideration at the Federal level.

There is currently no formal federal strategy for preservation of non-digital heritage resources. What exists now is an approach to preserving movable cultural property whereby federal institutions maintain their own collections, and institutions and programs such the Canadian Conservation Institute, the Museums Assistance Program, preservation funding administered by the Canadian Council of Archives, and other G&C programs, provide assistance to qualifying non-federal organizations in specific areas of activity. Together, the mandates and objectives of these different organizations and programs constitute the GoC strategy.

---

<sup>3</sup> Canadians will have built a culture of heritage conservation; Governments will work together to achieve common goals; Canadians will have the basic tools to protect historic places; The Government of Canada will become a model custodian of historic places. From: ([http://www.pch.gc.ca/progs/ieh-hpi/index\\_e.cfm](http://www.pch.gc.ca/progs/ieh-hpi/index_e.cfm))

A Museums Policy, now under development at PCH, includes preservation as a priority, and archives and special collection libraries are included in its scope. Digital preservation is not specifically mentioned, but neither is it excluded from the scope of the policy. Proposed as part of the policy is a national preservation network of expertise and knowledge.

### *LAC role*

Considerations for LAC include: its role as a potential leader in the development and implementation of a broad strategy; coordination with federal partners with responsibilities in this area, such as the Treasury Board and other heritage institutions; its relationship with Canadian Heritage in coordinating actions on behalf of the GoC.

LAC has fostered partnerships and contact with other government departments in a number of ways, such as:

- Statistics Canada has been preserving its own databases for some time;
- Canadian Conservation Institute has developed a workshop on preservation of digital formats and organized a symposium on the preservation of electronic records in partnership with the NA, NL and the Canadian Heritage Information Network (CHIN) ([http://www.cci-icc.gc.ca/symposium2003/index\\_e.shtml](http://www.cci-icc.gc.ca/symposium2003/index_e.shtml));
- Other government departments transfer e-records to LAC;
- CHIN, a special operating agency of the Department of Canadian Heritage, has published recommendations for best practice in digital preservation for the museum community ([http://www.chin.gc.ca/English/Digital\\_Content/Digital\\_Preservation/](http://www.chin.gc.ca/English/Digital_Content/Digital_Preservation/));
- Canadian Heritage and Foreign Affairs coordinate contributions from the GoC to international initiatives such as the UNESCO Charter on the Preservation of Digital Heritage ([http://portal.unesco.org/ci/en/ev.php-URL\\_ID=13367&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=13367&URL_DO=DO_TOPIC&URL_SECTION=201.html));
- LAC is responding to the discussion guide on a New Museum Policy recently issued by Canadian Heritage;
- Development of the Depository Services Program for federal publications with PWGSC.

In addition, strong links with institutions in other countries, particularly those with developing national strategies, will be essential. Strategic participation in international projects, standards development, and research will assist LAC's leadership role in Canada.

Awareness and promotion of preservation as a critical activity is an essential role for a leader in this field to play. Contacts with traditional partners in the archives and library communities can be used to foster awareness, but it will be important also to establish and maintain contacts with other sectors implicated in the creation and preservation of digital information.

### *LAC readiness*

At the national level, LAC's partners have for some time been looking to LAC for guidance. However, recent work with Anne Kenney and Nancy McGovern of Cornell University revealed that LAC scores about 2 on a scale of 1-5 in terms of institutional readiness to ensure the long-term preservation of digital information. LAC needs to proceed on several fronts to be a credible leader in digital preservation in Canada. It needs to:

- 1) Acknowledge its own state of readiness, and create opportunities and models for collaboration, e.g., in the area of applied R&D such as testbed projects; technology "watches";
- 2) Establish its key roles first in areas where it has strength, such as acquisition and description, and some aspects of digital conversion.
- 3) Seek to aggressively enhance its own digital preservation program by:

- developing an LAC-specific set of metadata that is the minimum necessary needed to preserve digital content over time;
- developing specifications for a digital content management system;
- establishing common processes and a robust technical platform for preserving digital content;
- vesting clear accountability and roles/responsibilities for digital preservation across the organization and developing staff capacity to fulfill these roles;
- consolidating operations and developing a true program
- determining how to ensure financial and technological sustainability for preserving its digital assets
- and moving forward in its modernized legislative mandate to implement legal deposit.

### *Redundancy*

Given the continually evolving nature of both technology and organizations, and the relatively low state of development of any Canadian institution to assure digital preservation over the long term, the strategy should look to build in distributed redundancy of content. The LOCKSS (Lots of Copies Keeps Stuff Safe) approach may provide a model (<http://lockss.stanford.edu/>).

### *Sustainability*

Any strategy must address not just the needs of “start up” but for the sustained effort, henceforth, that digital preservation will require if we are going to preserve Canada’s digital documentary heritage into the future.

## **5.4. Access considerations**

### *Access for All*

It is important that we take into account the access needs of all Canadians. Several voluntary guidelines exist (such as the W3C [www.W3c.org](http://www.W3c.org)) that help promote universal access to the Web. What do disabled Canadians truly require? Using other guidelines as examples, such as the W3C or the American federal legislation Section 508 ([www.Section508.gov](http://www.Section508.gov)) a Canadian guideline should be produced that looks at the different types of resources available and the information needs of different communities.

### *Balancing creator rights with preservation and access*

The issue of copyright and the Internet has been brought to the fore with the Web’s unprecedented means of diffusing content effortlessly and internationally. Copyright vigilance must be maintained, and donor restrictions and other legislations (i.e. Access to Information and Privacy) must be taken into account. Not all content can be made available online; but in the digital realm, preservation usually implies the need to copy, and the purpose of preserving content is to support future access to it. Past experience has shown that it is an easier path to restrict access than to foster access, but a national digital strategy must seek to maximize the access that Canadians can have to their documentary heritage.

### *Central or Distributed Access*

To what extent will the CDIS support the notion of a single portal providing access to the country’s collective digital cultural heritage? If distributed, what configuration will be developed, how will it be managed?

### *Pervasive media production and usage – what merits retention?*

Discussions of a national strategy must acknowledge the new pervasiveness of digital media in Canadians daily lives. We both produce and consume much more digital content, much of it AV; what, of that vast body of content, merits selection and long-term preservation?

## CONCLUSION

As a new institution, LAC is transforming organizationally to a single institution, operationally to the new digital information reality, and philosophically to a more access-oriented, innovative knowledge institution. This paper has attempted to describe the range and scale of our digital activities, and to articulate some of the challenges that we face in the transition to digital.

The policy directions for the new institution are expressed in its *Directions for Change*. One of those directions, “Digital is mainstream” sums up well where we want to be going and why:

LAC must consciously, proactively seek to understand the impact of digital and adapt itself to the new reality. Digital acquisitions, digital preservation, digital collection management, digital content delivery, digital reference—all are part of that reality. While content is what users seek, online is where they increasingly seek it. Digital is core to our business, and will be pursued as an over-arching strategic priority.

Digital must be viewed as mainstream business because digital content is the expectation of the 21<sup>st</sup> century user. But also, the digital world is where information-related research and development, innovation, and funding is now concentrated. To adapt to the digital environment, Canadian libraries need guidance, well-conceived funding programs, and concerted effort. As part of its national role, LAC can make a vital difference in helping to frame and advance a collaborative national digital strategy for Canada.

## LAC Contributors

Pam Armstrong  
Jean-Eudes Bériault  
David Brown  
Alison Bullock  
Michael Eamon  
John Grace  
Susan Haigh  
Roselyn Lilleniit  
Paul McCormick  
René Paquet  
Jean-Stéphen Piché  
Sharon Reeves  
Andrew Rodger  
Brian Thurgood  
Bruce Walton