InterPARES Project
International Research on Permanent Authentic Records in Electronic Systems
School of Library, Archival and Information Studies
The University of British Columbia
Suite 301 - 6190 Agronomy Road
Vancouver, B.C. Canada V6T 1Z3

Tel: 604-822-2694
Fax:604-822-1200
www.interpares.org

23 November 2007

Canadian Digital Information Strategy
Library and Archives Canada
395 Wellington Street
Ottawa, ON K1A 0N4
CDIS-SCIN@lac-bac.gc.ca

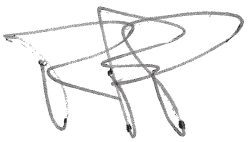
Dear colleagues,

The comments herein are respectfully submitted by the following individuals on behalf of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) Project: Luciana Duranti (University of British Columbia), Yvette Hackett (Library and Archives Canada) and Randy Preston (University of British Columbia), under the guidance of Professor Duranti, Director of the InterPARES Project.

InterPARES is a large, multi-year, multi-national research project of the archives community that is examining the long-term preservation of authentic records created and/or maintained in digital form. Since 1999, InterPARES researchers have been examining issues of authenticity, reliability and accuracy in the preservation of digital records, focusing more recently on records produced in complex digital environments in the course of artistic, scientific and e-government activities. Phase three of the research, InterPARES 3, which began in September of this year and will run until August 2012, is currently conducting a planned set of dozens of national and international test-bed case studies focused on addressing, in very practical terms, the digital preservation issues related to small and medium sized archives and archival units within organizations, and to other low-resources archival institutions.

Through its research, the InterPARES Project has identified opportunities for, and impediments to, long-term preservation of digital records. Because the overall vision, scope and challenges associated with Canadian Digital Information Strategy (CDIS) directly relate to our research activities and findings, we feel it is important that we offer our feedback on the draft strategy, and certainly appreciate having been given the opportunity to do so. For more information on the InterPARES Project, please see our Web site, http://www.interpares.org/.

Best regards,



Randy Preston
Project Coordinator, InterPARES Project
The University of British Columbia
Suite 301-6190 Agronomy Road
Vancouver, British Columbia V6T 1Z3 Canada
tel: +1(604) 822-2694 fax: +1(604) 822-1200
rpreston@interchange.ubc.ca
www.interpares.org

Canadian Digital Information Strategy
Draft Consultation Version
October 2007

InterPARES Project Comments
23 November 2007

Submitted on behalf of the InterPARES Project by:
Luciana Duranti (University of British Columbia), Yvette Hackett (Library and Archives Canada) and Randy Preston (University of British Columbia).

Introduction

In 2005, Library and Archives Canada (LAC) initiated discussions on issues in the digital field, with the goal of developing a Canadian Digital Information Strategy (CDIS) to address some of the critical issues in digital information production, preservation and access and to propose concrete actions to strengthen the Canadian digital information environment. To this end, LAC organized formal consultations with stakeholder organizations from a variety of sectors, culminating in a National Summit in 2006. Based on the results of this consultation process, a 24-member CDIS development committee produced a draft CDIS in October 2007. The Committee is now soliciting public feedback on this draft strategy. In particular, the Committee is asking the public to consider the following questions:

1. Do you agree with the overall vision, scope and challenges outlined in the strategy?

2. Are the objectives and actions set out in Part II the right ones? Which do you view as the most important or pressing?

3. What do you consider to be the critical next steps to advance the strategy? What role can you or your community play?

The findings of the first two phases of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) Project (1999-2007), suggest that some modifications to the current CDIS are warranted. The recommendations that follow draw on InterPARES research, which examines the requirements for, and impediments to, the creation, maintenance and long-term preservation of authentic, reliable and accurate records created, used and maintained in digital environments. While the focus of the CDIS is "information" rather than "records," we believe that the InterPARES findings are relevant to a critical review of the draft CDIS.

Question 1: Do you agree with the overall vision, scope and challenges outlined in the strategy?

As a Project whose ultimate goal is the development of strategies for the long-term preservation of authentic, reliable and accurate digital records, we applaud LAC's forward-thinking efforts to develop a national digital information strategy and, for the most part, concur with the overall vision, scope and challenges outlined in the strategy.

With respect to the repeated declaration in Part I that the "lines between information creator and consumer are blurring," we feel that this is an inaccurate assertion. Although digital technology in

general, and access to the Internet in particular, provides a ready means for people to create, distribute and/or access digital content, this does not make the creator, consumer or preserver roles any less "distinct." On the contrary, we feel that this so-called “blurring” of roles is an illusion that is being promulgated, in large part, by the fact that many consumers wilfully and/or unwittingly disrespect the rights of creators—they want free access, free use and free re-use of somebody else's material and are either pretending or erroneously assuming, based on, as acknowledged in the strategy (page 43), “uninformed assertions and interpretations of rights,” that the technology has somehow stripped creators of their right to fair compensation for their work (intellectual rights), or to protect it from unfettered and often distasteful re-use (moral rights).

For the most part, we believe that those individuals formally responsible for the management and preservation of digital information and records (i.e., information managers, records managers and archivists) understand and recognize quite well the distinction between creators and consumers. Thus, for example, while an interactive blog might appear at first glance to someone outside of the records management or archives professions to be a collection of essays from multiple creators, it is, in fact, a record within the fonds of the individual who created and controls the blog. This is no different than the existing practice of considering the letters of third parties received by an individual, and subsequently filed within the correspondence records of that individual, to constitute the records of that individual (i.e., the person who received and set the letters aside).

A similar inference about ‘blurred’ or ‘unclear’ roles in relation to preserver activities appears in Part II of the draft strategy on page 26, where it is noted that “stakeholders felt there was a need to clarify preservation roles amongst various stakeholder agencies." Quite apart from a genuine lack of clarity vis-à-vis the responsibilities of preservers, we feel that the current situation is more accurately characterized as the failure of many organizations, libraries, museums and archives to take on their already recognized and established management and preservation roles in the digital environment. While many preservers acknowledge their responsibilities for analogue materials, they, for whatever reason, seem all-too-willing to pretend or assume that the digital versions are somehow not their responsibility. Identifying open Web content as the equivalent of "published," or identifying records kept on an intranet as "archival" is not actually that difficult. Thus, we believe that if those responsible for the management and preservation of analogue materials actually stepped up and took care of the majority of digital materials for which the management and preservation responsibilities are (or should be) obvious, we could then better utilize our limited resources in a collaborative effort to categorize the remaining minority of materials for which management and preservation responsibilities are not so obvious.

Regarding the scope of the strategy, and more specifically the “key characteristics” of the digital information associated with each of the four sources identified in the table on page 10, we believe, based on the findings from the first two phases of the InterPARES Project, that the key characteristics presented in the aforementioned table should be augmented in the following two ways.

First, data quality should be listed as a key characteristic for the “Academic and research Community” source. Data quality is a critical characteristic used by scientists and other researchers to determine the usefulness and suitability of a set of data for use in their research activities, and is also an important factor in the decision of what scientific data archivists should select for long-term preservation. In fact, as the Science Focus case and general studies from InterPARES 2 demonstrate, scientists give primacy to data quality, which includes authenticity, normally articulated as provenance or lineage. Data quality is normally articulated in a dataset’s metadata; without metadata or data quality parameters, a scientist will not use, trust or rely on those data. Each scientific discipline differs in how it defines scientific data quality; however, most include some or most of the following data quality elements: positional accuracy;

attribute and thematic accuracy; completeness; semantic accuracy; and temporal information, reliability, lineage, logical consistency, and objectivity.

Second, reliability, accuracy and authenticity should be listed as a key characteristics for both the "Government and public sector" and "Business and corporate world" sources. In the management of digital records, reliability, accuracy and authenticity are three vital considerations for any organization that wishes to sustain its business competitiveness and to comply with legislative and regulatory requirements. The concept of reliability refers to the authority and trustworthiness of a record as a representation of the fact(s) it is about; that is, to its ability to stand for what it speaks of. In other words, reliability is the trustworthiness of a record's content. It can be inferred from two things: the degree of completeness of a record's documentary form and the degree of control exercised over the procedure (or workflow) in the course of which the record is generated. Reliability is then exclusively linked to a record's authorship and is the sole responsibility of the individual or organization that makes the record. Because, by definition, the content of a reliable record is trustworthy, and trustworthy content is, in turn, predicated on accurate data, it follows that a reliable record is also an accurate record.

An accurate record is one that contains correct, precise and exact data. Accuracy of a record may also indicate the absoluteness of the data it reports or its perfect or exclusive pertinence to the matter in question. The accuracy of a record is assumed when the record is created and used in the course of business processes to carry out business functions, based on the assumption that inaccurate records harm business interests. However, when records are transmitted across systems, refreshed, converted or migrated for continuous use, or the technology in which the record resides is upgraded, the data contained in the record must be verified to ensure their accuracy was not harmed by technical or human errors occurring in the transmission or transformation processes. The accuracy of the data must also be verified when records are created by importing data from other records systems. This verification of accuracy is the responsibility of the physical or juridical person receiving the data; however, such person is not responsible for the correctness of the data value, for which the sending person is accountable. Thus, the receiving person should issue a disclaimer regarding accuracy of records using other persons' data.

The concept of authenticity refers to the fact that a record is what it purports to be and has not been tampered with or otherwise corrupted. In other words, authenticity is the trustworthiness of a record as a record. An authentic record is as reliable and accurate as it was when first generated. Authenticity depends upon the record transmission and the manner of its maintenance and custody. Authenticity is maintained and verifiable by maintaining the identity and integrity of a record. The identity of a record is established and maintained by indicating at a minimum the names of the persons participating in the creation of the record (e.g., author, addressee); the action or matter to which the record pertains; the date(s) of compilation, filing or transmission; the record's documentary form; the record's digital presentation (or format); the relationship of the record to other records through a classification code or a naming convention; and the existence of attachments. The integrity of a record is established and maintained by identifying the responsibility for the record through time by naming the handling person or office(s) and the records keeper or the recordkeeping office, identifying access privileges and access restrictions and indicating any annotations or any modifications (technical or otherwise) made to the record by the persons having access to it.

Thus, record reliability is a quality that is established when a record is created and implies accuracy of the data contained in the record, while record accuracy and authenticity are qualities that are connected with the transmission and maintenance of the record. The latter are therefore the responsibility of both the records creator and any legitimate successor. Authenticity is protected and guaranteed through the

adoption of methods that ensure the record is not manipulated, altered, or otherwise falsified after its creation, either during its transmission or in the course of its handling and preservation, within the creator's recordkeeping system.

Question 2: Are the objectives and actions set out in Part II the right ones? Which do you view as the most important or pressing?

Insofar as the objectives and actions set forth in Part II are informed by the key assumptions (outlined on pages 12-13) of the strategy's 'framework for action,' we feel it is critical to augment these assumptions with reference to the importance of ensuring the quality and trustworthiness (i.e., the accuracy, authenticity and reliability) of the digital information assets in question. After all, what is the point of committing scarce resources to preserve and provide access to digital information "assets" whose accuracy, authenticity and/or reliability cannot be substantiated? Users must be able to either presume that the digital resources they are accessing and using are trustworthy—based on readily accessible and transparent documentation about the general creation, management and/or preservation activities of those responsible for oversight of the resources in question and the potential impacts of such activities on the trustworthiness of the resources under their custody—or, in cases where such a presumption of trustworthiness is weak or inadvisable, to verify the trustworthiness on the basis of a detailed examination of the resources themselves together with the known facts about the various contexts in which they have been created, maintained and preserved.

Regarding the sequence of the three "challenges" discussed in Part II (i.e., strengthen content, ensure preservation and maximize access and use), we feel that, whether intended or not, discussion of strengthening content as Challenge 1 may leave readers with the impression that digitization of Canada's existing analogue assets is (or should be) the "first" priority in the overall CDIS. While we acknowledge that the report is careful to be inclusive and that most of the recommendations are applicable to both born digital materials (including data) and digitized materials, we nevertheless believe—given the greater inherent 'fragility' of born digital materials in relation to the vast majority of analogue materials—that discussion of the issue of preservation should prevail over and precede discussion of the issue of digitization in the report. In the absence of established long-term preservation strategies for digital materials, "mass digitization on a national scale," as the report calls for on page 3, is a potential economic black hole should it later be discovered that the digital materials are incompatible with subsequently developed preservation strategies and thus require substantial reformatting or re-digitization, as well as a potentially significant liability risk in cases where it is later discovered that re-digitization is required for analogue materials that were destroyed following their initial digitization.

Regarding the report's discussion of digitization in Part II, particularly in relation to the explicit assertion earlier on in the report (page 6) that "[a]s a nation, we do not yet have the capacity to assure long-term access to our digital resources...[and that]...[i]ndeed, all digital information is at risk"—a sentiment that is repeated time and again throughout the report—we think it is confusing and, in all but relatively trivial cases, misleading to suggest that "digitization...can also serve a preservation purpose" (page 16). The example provided in the report, in which digitization is posited as a "preservation" strategy for deteriorating nitrate film, only serves to emphasize our point. In fact, given the inherent and notorious fragility of digital media, as well as the speed with which the hardware and software required to access the digital resources on such media pass into obsolescence, it seems incongruous to suggest, as a means of preservation, copying the unstable nitrate film to a digital format, and onto a digital medium, that are potentially every bit as unstable as, and definitely more short-lived than, the nitrate film itself, as

opposed to copying the images to a far more archivally permanent negative or photographic paper medium.

Question 3: What do you consider to be the critical next steps to advance the strategy? What role can you or your community play?

We believe that it is critical that institutions responsible for the creation and/or preservation of digital information assets move quickly to explicitly spell out their mission responsibilities for the types of digital assets that traditionally fall through the cracks, such as science databases. Libraries and archives should take responsibility for acquiring these types of databases in the context of the collections and fonds that they acquire.

Miscellaneous Issues, Comments and Recommendations

Regarding the characterization of information "as existing along a value chain with raw data at one end and packaged content...at the other" (page 9), we find the use of the term "value" in this context ambiguous at best. Presumably, a "value chain" means that one end of the information spectrum is more "valuable" than the other. If so, which end of the spectrum (raw data or derived, packaged content) does the CDIS consider to have more "value"? And on what criterion or criteria is this valuation assessment (supposed to be) made? And how does such an assessment account for context? For example, to those in the sciences, raw data have equal if not more "value" from a procedural point of view than any derived content data, since it is the raw data that enable and support the fundamental processes of scientific inquiry. Therefore, we recommend that the term "value chain" be changed to something less assumptive and more contextually neutral; perhaps to something like "process chain."

Regarding the summary discussion of the InterPARES Project that appears on page 29, we think it is important also to emphasize that, based on the findings of its numerous case studies, the Project has issued a Policy Framework[1] that includes creator and preserver principles designed to clarify and enhance the relationship between creators and preservers vis-à-vis their respective roles in the creation, use, management and preservation of digital records. Supplementing the Policy Framework are two sets of concrete guidelines for creation, management and preservation of digital materials, including one set for creators[2] and one set for preservers.[3]

Regarding Action 2.3.1 on page 30, which calls for "developing a collaborative Canadian digital preservation research agenda, including a planned set of digital preservation test bed projects," we would like to note that the InterPARES Project has been pursuing precisely this approach since 1999, and has, to date, completed 48 national and international test-bed case studies among creators and preservers in the arts, sciences and government sectors. Phase three of the research, InterPARES 3, which began in September of this year and will run until August 2012, is currently conducting a planned set of dozens of additional national and international test-bed case studies focused on addressing, in very practical terms, the digital preservation issues related to small and medium sized archives and archival units within organizations. We believe that the case study findings from all three phases of the InterPARES Project can be exemplary as typology for digital preservation research in Canada.

Regarding the "Highlighting Progress" sidebar discussion on page 30, we would like to clarify that the "Preservation of Digital Records" course cited is not the only course in the curriculum offered by the School of Library, Archival and Information Studies (SLAIS) at the University of British Columbia that

addresses the management of digital materials. In fact, discussion of issues surrounding the creation, management and appraisal of digital records are incorporated in all of the archival courses offered by SLAIS. As well, we think it is worth noting that the Library side of the curriculum also offers a number of courses that specifically address the creation, use and management of digital materials, including, for example, The Design of Electronic Text, Digital Image and Text Collections, and Digital Libraries.

Regarding Appendix II, please add the following InterPARES 3 Project citation and link to the InterPARES Project listings on page 60: InterPARES 3 Project http://www.interpares.org/ip3/ip3_index.cfm

[1] Available at: http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)policy_framework_document.pdf
[2] Available at: http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf
[3] Available at: http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)preserver_guidelines_booklet.pdf