InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records

PART FOUR

METHODS OF APPRAISAL
AND PRESERVATION

Domain 3 Task Force Report

[including Appendices 13, 21, 21a, 21b and 21c]

by

Yvette Hackett, Library and Archives Canada

Status: Final (public)

Version: Electronic

Submission Date: February 2007

Publication Date: 2008

Project Unit: Domain 3 Task Force

URL: http://www.interpares.org/display_file.cfm?doc=ip2_book_part_4_domain3_task_force.pdf

How to Cite: Yvette Hackett, "Part Four—Methods of Appraisal and Preservation: Domain 3 Task Force Report," [electronic version] in *International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records*, Luciana Duranti and Randy Preston, eds. (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008). <http://www.interpares.org/display_file.cfm?doc=ip2_book_part_4_domain3_task_force.pdf>

Table of Contents

Introduction[1]

Of the wide range of research areas that the second InterPARES Project broached, appraisal and preservation issues are considered by many to represent the core of the archival profession. Appraisal is that key function where the archivist's decision about what to acquire, and consequently, what not to acquire, establishes what primary records will be available in the future to support legal actions, historical research, genealogy—matters of identity, culture, history and rights.

Preservation decisions will either: ensure the successful survival of selected records over the long-term; preserve them in a manner that strips them of credibility and subsequent usefulness; or destroy them through error or omission, thus unintentionally overturning the appraisal decision. For many archivists, appraisal and preservation issues have the greatest impact on day-to-day work.

Background and mandate

Despite the importance of both appraisal and preservation to archivists, Domain 3 seemed to suffer more difficulties than did the other domains in agreeing on the focus of its activities and establishing a work plan. As set out in the initial research proposal, the Domain 3 research unit would begin by working with two products developed by InterPARES 1. It would first merge the Appraisal Task Force's *Model of the Selection Function*[2] and the Preservation Task Force's *Model of the Preservation Function*[3] with the *Manage Archival Fonds* model produced by the UBC-MAS Project.[4] The resultant "Grand Unified Model," as it was initially called, would represent the complete life-cycle of a record from its initial generation by the creator to its long-term preservation and access while in the trusted custody of the preserver. Once completed, prototyping experiments could then be conducted to illustrate how the functions depicted in the model could be incorporated into software applications.

Organizationally, there were difficulties integrating the highly experienced modelers from InterPARES 1, who were essentially continuing work begun in 1999, with InterPARES 2's new recruits, many of whom were unfamiliar with both archival activities and with the intricacies of the modeling methodology adopted by the Project. This situation was resolved in June 2003, when the modeling activity was moved into its own research unit (Modeling Cross-domain). The subsequent decision in February 2004 to create a new, second model based on the concept of the records continuum rather than on the lifecycle further distanced the work being done on the models from the researchers in Domain 3.[5]

---

[1] The author acknowledges the general contribution of all members of Domain 3 in the preparation of this report. In particular, the author thanks Luciana Duranti, Ken Thibodeau and Randy Preston for their contributions to the text and their editorial guidance. Any errors of representation or omission are the responsibility of the author.

[2] See Appraisal Task Force (2001), "Appendix 4: A Model of the Selection Function," in *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, Luciana Duranti, ed. (San Miniato, Italy: Archilab, 2005), 239–252. Online reprint available at http://www.interpares.org/display_file.cfm?doc=ip1_aptf_model.pdf.

[3] See Preservation Task Force (2002), "Appendix 5: A Model of the Preservation Function," version 6.0, ibid., 253–292. Online reprint available at http://www.interpares.org/display_file.cfm?doc=ip1_ptf_model.pdf.

[4] See "Appendix B: Activity Models," in Luciana Duranti, Terry Eastwood and Heather MacNeil, *Preservation of the Integrity of Electronic Records* (Dordrecht: Kluwer Academic Publishers, 2002), 92–106. Online reprint available at http://www.interpares.org/UBCProject/a-0f.htm.

[5] See the Modeling Cross-domain Task Force Report. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_5_modeling_task_force.pdf.

The Domain 3 researchers also abandoned earlier plans to use InterPARES 1 case studies to validate the InterPARES 1 models prior to any system prototyping activity. The changes required in the original InterPARES 1 models to create the merged Chain of Preservation (COP) model meant the two original models were too out of date to provide useful input to the current work. In the case of the Business-driven Recordkeeping (BDR) model, no analysis could be undertaken until it was completed, which, given its late start, could only occur late in the Project.

Furthermore, as InterPARES 2 case studies were approved, it became obvious that InterPARES 1 case studies could not be used as a shortcut to get started on Domain 3 work while new case studies were being completed. The record-making environments being proposed for study were radically different from anything studied in InterPARES 1, even in the government focus. As with the modeling group, only a few members of Domain 3 had participated in InterPARES 1 and had extensive knowledge of the earlier case studies. It appeared that the learning curve would be daunting, while producing little insight relevant to the dynamic, interactive and experiential systems under study in InterPARES 2.

By the time the Midterm Progress Report was produced in the spring of 2004,[6] Domain 3 had little to nothing left of its original statement of work:

> ...the activities originally planned for the first two years of research of the Domain 1 and Domain 3 Task Forces have been reassigned to a new research unit, the Modeling Cross-domain Research Team....[7]

Of the planned outcomes—"prototypes of appraisal and preservation systems, activity models, and guidelines for records preservers"[8]—only the concept of the guidelines remained.

One final problem affected the researchers in Domain 3. Despite the large number of participants in the Project, archivists were in short supply. There were nine Working Groups,[9] joining either as three focus research units or as three domain research units and four cross-domain research units, requiring archival representation with experience from across the spectrum—from small to large archives, from corporate and governmental organizations, from independent institutions and those attached to large parent organizations, as well as archivists with experience with both private-sector and public-sector records, and the artistic, scientific and governmental sectors. Archivists were also required on each of the twenty-three approved case studies and eleven general studies. There were, in fact, not enough archivists to go around, and this problem was exacerbated by the difficulty that a number of the participating archival institutions seemed to have in maintaining consistent representation to the Project over the long term.

In the spring of 2005, with a number of case studies and general studies completed and with the Chain of Preservation model lacking only the narrative report, members of Domain 3 finally acknowledged that, in InterPARES 2 as in life, appraisal and preservation come last. It was now time for Domain 3 to begin its work.

The InterPARES 2 research proposal described the appraisal function, emphasizing the variations introduced when an appraisal addresses records in digital, rather than analogue, form:

> Appraisal assesses the continuing value of the records but it also assembles evidence for the presumption of their authenticity, and identifies the digital

[6] See http://www.interpares.org/ip2/ip2_midterm_progress.cfm.
[7] Luciana Duranti (2004), "InterPARES 2 Project Midterm Report to the Social Sciences and Humanities Research Council." MCRI Grant No. 412-2001-1003, 4 (unpublished).
[8] InterPARES 2 Project, Domain 3 Web page. Available at http://www.interpares.org/ip2/ip2_domain3.cfm.
[9] For a summary of the intellectual organization of the Project, see http://www.interpares.org/ip2/ip2_intellectual_organization.cfm.

> components or objects that need to be stored and reproduced to ensure the preservation of authentic records. Appraisal also establishes the feasibility of preserving a given body of electronic records in light of the existing and expected preservation capabilities of the preserver.[10]

This definition reflects the findings of the first InterPARES Project, which highlighted the need for three important shifts in traditional appraisal methodology when applied in a digital environment.

First, there is a need to overtly document the evidence in support of the authenticity of the records, given the ease with which digital records can be accidentally damaged or intentionally modified. With analogue records, the presumption of authenticity is frequently assumed, particularly when they are acquired directly from the creator, which confirms provenance and forms an unbroken chain of custody from creator to preserver.

Second, the definition cited above emphasizes the concept of "digital components," the various bits and pieces that must be identified and preserved to reproduce the complete record. In a vast majority of cases with analogue records, the content of the record and its structure cannot be altered or separated from the carrier—the components of the record are inextricably linked. Much of the context of the record's creation may also be overtly incorporated into this unalterable whole, primarily through the use of file classification numbers on the record and the presence or absence of elements of documentary form.

The third shift addresses the complex preservation alternatives that must now be assessed and costed before a preserver can realistically commit to the long-term preservation of digital records. The range of digital preservation strategies has moved far beyond the adoption of acid-free folders and boxes, although one must acknowledge that even with analogue records, "progress" has offered a steady supply of increasingly difficult formats requiring preservation—from acidic paper, to nitrate still and moving image negatives, to thermal fax paper, to magnetic audiotape.

As the InterPARES 2 research unit responsible for investigating the methods of appraisal and preservation, Domain 3 was tasked with investigating whether the concepts developed by both the Appraisal Task Force and the Preservation Task Force of InterPARES 1 would still apply in the digital environments examined by InterPARES 2. These dynamic, interactive and/or experiential environments could potentially produce records that have no obvious equivalent in the traditional analogue world.

The final report of the InterPARES 1 Appraisal Task Force offered a number of suggestions regarding the practice of appraisal in a digital environment.[11] Based on the analysis of case studies that focused primarily on databases and on document and records management systems, and based on the development of a model of the appraisal function, the Appraisal Task Force made the following conclusions:

- The appraisal of digital records is best conducted when the records are still active. The appraisal of digital records early in their lifecycle greatly improves the documentation available to the archivist about the operational role of the records in the creator's organization and provides technical information about how the application generates and

[10] Luciana Duranti (2001), "International Research on Permanent Authentic Records in Electronic Systems (InterPARES): Experiential, Interactive and Dynamic Records," SSHRC MCRI InterPARES 2 Project Proposal, 412-2001, 1.1-12. Available at http://www.interpares.org/display_file.cfm?doc=ip2_detailed_proposal.pdf.

[11] See Heather MacNeil et al., "Part One – Establishing and Maintaining Trust in Electronic Records: Authenticity Task Force Report," in Duranti, *Long-term Preservation*, op. cit., 19–65. Online reprint available at http://www.interpares.org/book/interpares_book_d_part1.pdf.

maintains records during the active and semiactive periods. Early identification of records with archival value should improve the chances that these records will not be destroyed accidentally or fall into unrecoverable technological obsolescence.

- The medium of records affects the process of appraisal but not the fundamental task of assigning value. Tasks such as the formal identification of indicators of authenticity,[12] the assessment of preservation strategies and the ongoing monitoring of appraisal decisions all represent new or expanded tasks in the appraisal process.
- Monitoring the appraisal decision to confirm the continued archival value of selected records is a necessary activity in the digital environment. The concept that appraisal decisions need to be re-visited at regular intervals is particularly applicable when appraisals are conducted, as suggested above, some time before the actual transfer of inactive records will occur, and to keep pace with the rapid rate of technological change.
- Information compiled during appraisal must be "packaged" and carried forward to assist with ongoing monitoring, transfer, processing, description, preservation and subsequent access. The automation of all aspects of archival work, including appraisal, will greatly facilitate this ongoing re-use, at subsequent phases, of information collected during the appraisal process.

Following a similar methodology, the InterPARES 1 Preservation Task Force, in its final report,[13] made the following conclusions:

- It is not possible to preserve a digital record: it is only possible to preserve the ability to reproduce the record. As with the findings related to appraisal, this statement emphasizes the concept of "components" when discussing digital records. This concept is not unknown among analogue technologies—examples of simpler forms of the concept include the negative and the print in photography, or the negative and positive image, the optical and/or magnetic soundtrack, the composite print and outtakes in moving images. Digital records offer the most complex version of the component system, requiring careful attention to multiple dependencies related to hardware, operating systems and application software.
- The intellectual and physical components of a digital record do not necessarily coincide; a digital component is distinct from an element of documentary form. For example, the content of a record may include both text contained in a word processing file and a table generated by spreadsheet software. Technically, the text file may only contain a link to the spreadsheet file, which in turn may depend on the spreadsheet software rather than word processing software to display it by recognizing and actualizing formatting information.
- The process of preservation must be thoroughly documented as a primary means for protecting and assessing authenticity over the long term. Since the process of preservation begins at creation, responsibility for this thorough documentation rests with both the creator and the preserver. In the past, the stability of most analogue record forms frequently allowed creators to ignore preservation concerns until the inactive records were transferred to the preserver.

[12] See Authenticity Task Force (2002), "Appendix 2: Requirements for Assessing and Maintaining the Authenticity of Electronic Records," ibid., 204–219. Online reprint available at http://www.interpares.org/book/interpares_book_k_app02.pdf.
[13] See Kenneth Thibodeau et al., "Part Three – Trusting to Time: Preserving Authentic Records in the Long Term: Preservation Task Force Report," ibid., 99–116. Online reprint available at http://www.interpares.org/book/interpares_book_f_part3.pdf.

When InterPARES 2 researchers developed the various research methodologies that would be used during the course of the Project, the overall thrust was to discover whether the case studies, modeling exercises or surveys planned for InterPARES 2 would uncover any theories or practices that disagreed with the existing recommendations of InterPARES 1. Overall, the findings of InterPARES 1, based on large databases and records management applications, had fit smoothly into existing archival knowledge and practice. Essentially, these databases and applications were found to produce digital manifestations of record types that were well-established in the paper-based recordkeeping environment, such as case files and textual records. Would the study of newer forms of digital records by InterPARES 2 researchers suggest the need for new appraisal criteria, or the addition of steps to current appraisal practices, or the development of new preservation strategies or practices?

Research team

The following is a list of researchers and research assistants who participated in the Domain 3 Task Force throughout the Project.[14]

Chairs and Co-chairs:

Yvette Hackett and Sally Hubbard	Jun 2005 - Dec 2006 (Co-chairs)
Hans Hofman and Sally Hubbard	Feb 2004 - Jun 2005 (Co-chairs)
Ken Thibodeau	Jan 2002 - Dec 2003 (Chair)

Researchers:

Howard Besser	New York University, USA—Working Group 3.1
Ann Butler	New York University, USA—Working Group 3.1
Kevin Glick	Yale University, USA—Working Group 3.2
Elaine Goh	National Archives of Singapore—Working Group 3.3
Yvette Hackett	Library and Archives Canada—Working Group 3.1
Babak Hamidzadeh	Library of Congress, USA—Working Group 3.2
P.C. Hariharan	Systems Engineering & Security, Inc., USA—Working Group 3.2
Ken Hawkins	National Archives and Records Administration, USA—Working Group 3.3
Hans Hofman	National Archives of the Netherlands—Working Group 3.3
Sally Hubbard	Getty Institute, USA—Working Group 3.1
Mary Ide	WGBH, USA—Working Group 3.1
Randal Luckow	Turner Broadcasting, USA—Working Group 3.1
Richard Marciano	San Diego Supercomputer Center, USA—Working Group 3.3
Evelyn McLellan	Insurance Corporation of British Columbia, Canada—Working Group 3.3
Reagan Moore	San Diego Supercomputer Center, USA—Working Group 3.2
Isabella Orefice	Associazione Nazionale Archivistica Italiana—Working Group 3.1
Jim Suderman	Archives of Ontario, Canada—Working Group 3.3
Ken Thibodeau	National Archives and Records Administration, USA—Working Group 3.3

[14] Researcher membership in Domain 3 changed substantially over the five years of the Project. Among those who were interested in appraisal and preservation issues but were unable to participate for the full length of the Project are: Filip Boudrez, City Archives of Antwerp/the DAVID Project, Belgium; Michèle Cloonan, Simmons College, USA; Margaret Hutchison, British Columbia Archives, Canada; Glenn Isaac, British Columbia Archives, Canada; Rick Kopak, The University of British Columbia, Canada; Rich Lysakowski, CENSA, USA; Jean-Stéphen Piché, Library and Archives Canada; Shelby Sanett, Amigos Library Services, Inc., USA; Lynne Tibbitt, British Columbia Archives, Canada; and Bill Underwood, Georgia Tech Research Institute, USA.

James Turner	Université de Montréal, Canada—Working Group 3.1

Research Assistants:

Tom Anderson	The University of British Columbia, Canada
Patsy Baudoin	Simmons College, USA
Carolyn Casenas	The University of British Columbia, Canada
Natalie Catto	The University of British Columbia, Canada
Alan Doyle	The University of British Columbia, Canada
Adam Farrell	The University of British Columbia, Canada
Fiorella Foscarini	The University of British Columbia, Canada
Peggy Heger	The University of British Columbia, Canada
Tracey Krause	The University of British Columbia, Canada
Karen Langley	The University of British Columbia, Canada
Catherine Miller	The University of British Columbia, Canada
Luke Meagher	The University of British Columbia, Canada
Jennifer Meehan	The University of British Columbia, Canada
Shaunna Moore	The University of British Columbia, Canada
Elisheba Muturi	The University of British Columbia, Canada
Carolyn Petrie	The University of British Columbia, Canada
Brian Trembath	The University of British Columbia, Canada
Sherry Xie	The University of British Columbia, Canada

Research questions

The four research questions that were to be answered over the course of the Project were identified in the original Project proposal, as follows:

1. How do the appraisal concepts, methods and models developed by InterPARES 1 for the administrative and legal records created in databases and document management systems apply to the appraisal of the records of artistic, scientific and governmental activities resulting from the use of the technologies examined by InterPARES 2?
2. How do the preservation concepts, methods and models developed by InterPARES 1 for the administrative and legal records created in databases and document management systems apply to the preservation of the records of artistic, scientific and governmental activities resulting from the use of the technologies examined by InterPARES 2?
3. What preservation paradigms can be applied across activities and technologies? What preservation paradigms are required for specific types of records resulting from each activity?
4. What metadata are necessary to support appraisal and preservation of authentic digital records resulting from each activity?

As noted earlier, the model-based analysis originally envisioned for questions 1 and 2 were subsequently transferred to the Modeling Cross-domain. However, the Domain 3 researchers would still be able to recognize any shifts in how appraisal and/or preservation activities might need to be conducted in the artistic, scientific and governmental environments being examined in the InterPARES 2 case studies, as opposed to the more traditional bureaucratic environment targeted in InterPARES 1. Of particular concern here was determining whether appraisal or preservation procedures would require any additional adjustments or deviations to accommodate digital technologies beyond those already identified by InterPARES 1.

The paradigmatic shift encountered in the InterPARES 1 research resulted in a shift in the language of digital preservation, from the concept of the preservation of "the record" to an acknowledgement of the importance of identifying and preserving "digital components" and their relationships.[15] Of particular concern here was determining whether the use of interactive, experiential or dynamic systems would require a similar shift in archivists' understanding of how to carry out archival activities.

Finally, it was determined that the answer to the fourth, metadata question should flow naturally from archivists' review of the case studies. An inability to appraise the case studies' records or to determine an appropriate preservation strategy for them would strongly suggest missing metadata, and should actually identify the type of metadata required.

Research Methodology

The concept of appraisal, as practised by archivists, is largely unknown among records creators. The identification of the value of their records as transitory, short term or long term is based on business rules and practices and, in some professions, legal requirements. But these values are generally assigned to records without reference to issues of reliability, authenticity or the value of the records from an historical, cultural or social perspective or the future interests of any potential third-party researchers. Therefore, the initial meetings of the Domain 3 researchers concentrated on the development of methodologies appropriate to the study of the creators' maintenance activities, despite the differences between record maintenance strategies and the time frames or concerns involved in archival preservation.

The first research initiative to be considered by the Domain 3 researchers was the development of a bibliography on digital preservation. This was eventually rejected, since there are many excellent bibliographies already available, and they are well documented by the National Library of Australia's Preserving Access to Digital Information (PADI) site.[16]

Case studies

> *The Domain 3 task force will use the knowledge gained in the course of the case studies...*[17]

The Domain 3 researchers considered the usefulness of initiating one or more case studies designed to answer the research questions. This would have involved studying organizations with appraisal and preservation either as their core activity, or with internal units tasked with appraisal and preservation on behalf of the larger organization. The archivists in Domain 3 were well aware of the fact that, internationally, very few archival units had developed practices and procedures in these areas and, of those that had, most had already been targeted by other research units within the Project, be they a focus, domain or cross-domain.

The Domain 3 researchers confirmed the value of participating in already approved InterPARES 2 case studies and surveys and of analyzing those results from an appraisal and preservation perspective, rather than launching competing projects.

[15] See Thibodeau et al., "Preservation Task Force Report," op. cit., 6–7.
[16] The subsection of the PADI site devoted to bibliographies contains resources about preserving access to digital information (see http://www.nla.gov.au/padi/format/bib.html). Recently, the indicator "Historical" was added to identify less current material.
[17] InterPARES 2 Project, Domain 3 Web page. Available at http://www.interpares.org/ip2/ip2_domain3.cfm.

Summaries of the case studies are available on the InterPARES 2 Web site[18] and in the Focus Task Force Report.

Case study interview questions

An early activity of the Domain 3 researchers was to design the following question for inclusion in the suggested list of case study questions:[19]

> *39. Do you have a standard procedure when it is time to preserve [your documents/work] for the long-term?*

It was hoped that this question would elicit information about any practices implemented or lessons learned about preservation by creators handling even fairly recent digital objects. The limited perspective of many of those who work in the information technology industry guarantees that obsolescence and interoperability problems can be experienced within extremely short time periods.

A number of other questions were also reviewed, since it was possible that case study respondents would provide information about various aspects of their preservation activities in any number of the suggested questions. Question 34 looked for any efforts at standardization in the work-flow that could represent the traditional concept of the "normal course of business." A broad question, it also allowed the identification of any standards adopted by the records creator:

> *34. Have you had to make rules, or adopt standards to help you in your work? Do you find you have to update them regularly?*

Although primarily a question for the Policy Cross-domain, question 46 addressed the fact that the identification of legal or ethical issues related to the work could have an important influence on the preservation strategies appropriate for the digital objects of a particular case study:

> *46. Do any legal or ethical issues arise from your electronic work?*

Questions 9 and 10 directly addressed the concerns of the Description Cross-domain. The existence of identity and integrity metadata describing a digital object is a *benchmark requirement* for authenticity.[20] Therefore, the descriptive practices of the records creators involved in the case study were an important aspect of their successful preservation:

> *9. Did you create or adopt a standard list of information which you try to record about each file, or work?*
>
> *10. Where did you get it? Do you know if others use the same one?*

Since the process of preservation of digital records really begins at the moment of creation, any number of the suggested questions could also elicit information of interest to Domain 3's research concerns. As case studies were completed, the results of the interviews were reviewed, both from the perspective of the case study questions and the Domain 3 research questions.

[18] See http://www.interpares.org/ip2/ip2_case_studies.cfm.

[19] Domain 3 (2003), "List of possible case study questions that the researchers may ask the subjects of their case studies to acquire the information necessary to answer the 23 questions." Available at http://www.interpares.org/display_file.cfm?doc=ip2_possible_questions_for_interviewees.pdf.

[20] Specifically, Benchmark Requirement A.1 - Expression of Record Attributes and Linkage to Record. See Authenticity Task Force Report, "Appendix 2," op. cit., 210–211. See Appendix 21a for an abridged version of the benchmark requirements.

Domain 3 Template for Case Studies Analysis

During the February 2005 InterPARES workshop in Vancouver, the Domain 3 researchers proposed developing a template for analyzing the Domain 3-related data being generated by the case studies. The template included nine questions related to the appraisal and preservation issues being investigated by Domain 3.[21] The questions were designed to extract information contextualized to fit the template from each of the case studies. Although the information the Domain 3 researchers were looking for about appraisal and preservation activities was largely embedded within sections E (narrative answers to the core research questions) and F (narrative answers to the applicable domain and cross-domain research questions) of the final case study reports,[22] supplementary information could come from interim case study reports or from other available documentation, such as interview transcripts or the creators' Web sites. In some cases, additional information could be inferred, based on researchers' knowledge of certain fields. In addition to the nine questions, the template included instructions and guidelines for answering the questions, along with two appendices that provided detailed descriptions and definitions of key concepts, as well as immediate access to the relevant InterPARES 1 benchmark and baseline requirements and commentaries.

After being presented and discussed by InterPARES 2 researchers at the Chicago workshop in September 2005, a test "walkthrough" of the template was conducted with several case studies at both the Chicago workshop and the International Team meeting in Venice in December 2005. Following revisions and final approval of the template at the February 2006 workshop in Vancouver, the template was applied to the remaining case studies by a team of University of British Columbia research assistants under the direction of the Domain 3 researchers. Once completed, the templates for each case study were forwarded to the appropriate case study principal investigator(s) for review, comment and, as necessary, revision to ensure that the templates had drawn out the relevant information from the case studies. Once validated, the completed templates were analyzed by researchers who represented a wide range of disciplines.

At the end of this process, the Domain 3 Task Force had a tool that assisted the research in several ways. First, it augmented the validation of the case studies that was being undertaken by the Domain 1 Task Force, by highlighting certain gaps in the existing reports and serving as an informal methodological guide for writing the narrative answers to the Domain 3 questions in the final reports of the case studies that were still underway. Second, this tool served as an index to appraisal- and preservation-related information in the case studies, which InterPARES researchers could use to: (a) associate with the analyses done by other research units, including the diplomatic analyses, case study characterizations and modeling walkthroughs; and (b) return to the source material as needed to clear up a given question. Third, by characterizing all of the case studies in relation to the Domain 3 questions, it was possible to identify activities that would affect the ability of archivists to subsequently appraise and preserve the records of the creators in each of the case studies. Finally, this tool provided a more formalized and analytically rigorous compendium of case study-related information upon which to draw while developing the *Preserver Guidelines*.[23]

[21] Appendix 13.
[22] See Appendix 9. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_09.pdf.
[23] See Appendix 21. The Guidelines also are available in booklet form at http://www.interpares.org/ip2/display_file.cfm?doc=ip2(pub)preserver_guidelines_booklet.pdf.

General studies

In addition to the information extracted from the case studies, the Domain 3 researchers determined that a number of the Project's general studies provided supplementary information that could be used to help address the Domain's research questions and further inform development of the Guidelines for Preservers. These studies and their relevance to Domain 3 are summarized as follows.

General study 01: Persistent Archives Based on Data Grids

The first of the preservation-oriented general studies to receive approval in InterPARES 2, entitled *Persistent Archives Based on Data Grids*, was conducted by Reagan W. Moore of the San Diego Supercomputer Center. Beginning from the premise that preservation environments for digital records are successful when they can separate the digital record from any dependence on the original creating infrastructure, the study investigated the potential use of data grid technology, which supports the management of distributed records. The final report, *Building Preservation Environments with Data Grid Technology*,[24] examined the minimum capabilities required to preserve records, focusing on selected digital holdings of the U.S. National Archives and Records Administration and a number of other data grid implementations around the world.

General study 04: Survey of Recordkeeping Practices of Composers

This is one of three general studies that was derived from the case studies and that involved surveying a group of records creators who were also the subject of a case study.[25] As the Focus Task Force Report explains:

> ...while case study 13 was delving deeply into the work of one composer, Keith Hamel, and the technical details of one specific composition, *Obsessed Again...*, the Focus 1 researchers understood that this single case was not necessarily representative of the full range of adoption and use of digital technologies among composers...[26]

This 2003 survey confirmed that most composers using digital technology work with off-the-shelf commercial software, and almost half of the respondents (forty-seven percent, or seventy-five individuals) have already lost files they considered valuable, through either hardware or software obsolescence.[27] The majority of composers keep digital files primarily for re-use rather than from any concern about long-term preservation or posterity, and thus their primary concern is accuracy, which is clearly related to their need for continued access to the files rather than to concerns about authenticity. Composers work alone; therefore, there is limited access to their digital files by others. This finding also identified the appraisal and preservation issues surrounding the difference between the score and the performance, which are discussed extensively in the report of the task force on Records Reliability, Accuracy and Authenticity (i.e., Domain 2). The only "preservation" measures that the surveyed composers consistently took to

[24] Reagan W. Moore (2004), "InterPARES 2 Project - General Study 01 Final Report: Building Preservation Environments with Data Grid Technology." Available at http://www.interpares.org/display_file.cfm?doc=gs01_final_report.pdf.
[25] General studies 07 (Survey of Recordkeeping Practices of Photographers using Digital Technology) and 09 (Digital Recordkeeping Practices of GIS Archaeologists), discussed in the following two sections, were the other two studies. All three of these general study surveys—dealing with composers, photographers and archaeologists—contained questions relating to preservation practices.
[26] Focus Task Force Report, 2. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_1_focus_task_force.pdf.
[27] Michael Longton (2004), "InterPARES 2 Project - General Study 04 Final Report: Survey of Recordkeeping Practices of Composers." Available at http://www.interpares.org/display_file.cfm?doc=gs04_final_report.pdf.

protect their files were either to produce a backup copy on an external medium or else to upgrade file formats to more current versions.

General study 07: Survey of Recordkeeping Practices of Photographers Working with Digital Materials

This 2004 survey was completed by a larger number of practising photographers (*n*=402) than that of the composers (*n*=161) and revealed a rather different situation:

> In response to the pressures of faster turnaround times, creative innovation and remote transmission, professional photographers have universally embraced the transition from analogue to digital photography. The majority of photographers identified their practice as completely digital, allocating the use of analogue film to the occasional personal project. Even amongst those who identified their practice as a hybrid of digital and analogue, the bulk of their images were born digital - only a small percentage of analogue images were made and most of these were eventually digitized.[28]

Like composers, photographers were most interested in maintaining their digital images to support re-use and reference, although the business aspects of their profession, such as journalism or forensic work, had made them more sensitive to issues such as authenticity and copyright. Their work environment also made many of them aware of transmission problems with digital files, as well as hardware, software and storage issues. In addition to refreshing digital media and upgrading file formats, the significant number of photographers who responded to the survey routinely record metadata, practice quality control, implement security procedures and maintain multiple versions of their digital images, including in-camera file formats and working drafts and versions.

General study 09: Survey of the Digital Recordkeeping Practices of GIS Archaeologists Worldwide

This 2004 survey of GIS use by archaeologists examined one field of scientific endeavour. Unlike the previous two surveys, this survey could compare its results against a 1998 survey investigating similar issues.[29] The InterPARES survey revealed that a significant increase in the awareness of archaeologists about digital issues had occurred during the intervening six years since the 1998 survey:

> A key indicator of this increased awareness is the growing sense of frustration expressed by many participants over the current lack of suitable long-term preservation repositories available to archaeologists, as well as over the continuing absence of any concerted, profession-wide response to these particular issues and concerns.[30]

On the other hand, the survey also revealed that many respondents engaged in idiosyncratic and ad hoc file creation, management, preservation and/or documentation practices. Lack of training in GIS software was cited as one problem, as was a lack of knowledge about how to

[28] Jessica Bushey and Marta Braun (2006), "InterPARES 2 Project - General Study 07 Final Report: Survey of Recordkeeping Practices of Photographers using Digital Technology," 30. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs07_final_report.pdf.

[29] Khalid Gourad. "Geographic Information Systems in Archaeology: A Survey" (unpublished Master's thesis, Hunter College of the City University of New York, Department of Anthropology, 1999), 75 pp. Both the thesis and the survey are available at: http://khalid.gourad.com/thesis/.

[30] Randy Preston (2006), "InterPARES 2 Project - General Study 09 Final Report: Digital Recordkeeping Practices of GIS Archaeologists Worldwide: Results of a Web-based Survey," 3. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs09_final_report.pdf.

provide adequate documentation within the bounds of available project funding. The specialized nature and complexity of GIS software was reminiscent of the composers' situation, while the problem of ownership of the digital files (the archaeologist or the client/funding source) resembled the photographers' problems with employers, freelancing and commissions.

General study 10: Preservation Practices of Scientific Data Portals

This general study involved a researcher survey of thirty-two data portals from a variety of scientific disciplines designed to reflect the heterogeneity of scientific research. Structured information was collected about each portal's service, its host institution, income sources, access fees and metadata standards. In addition, information about preservation practices, as well as statements related to the data quality, accuracy and/or reliability of each portal's data were collected. The purpose of the study was to compile structured information about the standards and protocols in place at science data portals (a.k.a., archives, repositories, catalogues, etc.) that would, in turn, provide better insight into how the key InterPARES 2 issues of accuracy, reliability and authenticity are interpreted and understood in the sciences and how this understanding is used to underwrite confidence in data accuracy, reliability and authenticity as practiced and implemented in the natural and physical sciences.

General study 11: Selecting Digital File Formats for Long-term Preservation

This general study, which focused on the criteria used by various preserver organizations to select digital file formats for preservation, was originally developed in the Policy Cross-domain. The initial interest was in studying policy statements related to preservation activities, but the research was transferred to Domain 3 when it began to focus on the more specialized question of criteria for format selection. The final report's seven recommendations address issues ranging from vague or misleading terminology, to the need to clearly distinguish between various file types, to compression, among others.[31]

General study 05: An Examination of the Processes to Preserve and Manage Electronic Records

This general study, which was never completed, would have constituted the third part of an extended research project initiated during InterPARES 1. Begun in 2000, the initial Survey of Preservation Practices and Plans investigated the current state of digital preservation practices and plans at thirteen institutions, projects or programs in North America, Australia and Europe.[32] The findings, reported in June 2001, revealed:

> ...a number of preservation techniques were in use but that none of them could be considered meeting archival requirements for authenticity. The study also revealed that while developing technological processes to preserve authentic electronic records, almost every institution had deferred costing digital preservation processes and implementing digital preservation policies.[33]

[31] Evelyn Peters McLellan (2006), "InterPARES 2 Project - General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation," 16–17. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_english.pdf. French language version available at http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_french.pdf.

[32] Michèle V. Cloonan and Shelby Sanett (2002), "Where We Are Now: Obliquity and Squint?" *The American Archivist* 65(1): 70–106. The June 2001 draft of the report, entitled "Survey of Preservation Practices and Plans," is available as a draft appendix to the Preservation Task Force's final report at http://www.interpares.org/display_file.cfm?doc=ip1_survey_of_preservation_practices-plans.pdf.

[33] Michèle V. Cloonan and Shelby Sanett (2004), "InterPARES 2 Project - General Study 05 Final Report, Round 2: The Preservation of Authentic Electronic Records: *Ad hoc*, Inconsistent, or Strategic?" 3. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs05_r2_final_report.pdf.

The second phase of the survey was undertaken between August 2001 and February 2003, thus bridging the transition from InterPARES 1 to InterPARES 2. The authors re-surveyed eight of the thirteen organizations that participated in the first round, and conducted eighteen in-depth interviews with selected key informants. This process provided a more detailed understanding of the preservation strategies being explored and measured the progress that had been made in the interim. They found that the programs' successes were uneven, moving ahead in some areas while lagging behind in others:

> For example, institutions are beginning to think about cost issues and models, but have been slow to develop digital preservation policies and plans. As one of our survey respondents in round 2 observed about his own institution, "As long as there is no plan, the risk will be that preservation will be ad hoc, inconsistent, and not imbedded in the organization."[34]

The third survey—An Examination of the Processes to Preserve and Manage Electronic Records: Round Three at The National Archives of Australia and WGBH—was approved by InterPARES 2 in February 2004. The intent was to conduct in-depth studies of two organizations that had been surveyed in the two previous rounds. One particular focus would be policy development and planning, which had been identified as lacking in the previous report. The second area of emphasis would be the specific preservation strategies adopted by the two organizations. Unfortunately, issues of distance and availability—of both the researchers and the subjects of the research—eventually posed insurmountable problems and the final phase of the project could not be carried out.

Research Findings

Analysis of the case and general study data suggests that too many records creators are still neglecting the long-term preservation of their digital files, whether they be static or dynamic, evidential or experiential, historically significant or interactive. The range and breadth of the many case and general studies does mean that almost every variation in approach and attitude was documented at least once. However, some distinctions can be made between individuals and small organizations on the one hand and large corporations and government institutions on the other. The InterPARES 2 case studies done by the government focus provided similar findings to those done in InterPARES 1, demonstrating a greater knowledge and awareness of appropriate practices for records creation and maintenance, whether of digital or paper-based records. Unlike individuals and small organizations, the larger institutions also tended to have analogue recordkeeping systems in place, which the institutions could fall back on by identifying the printed paper copy as the authoritative record and moving it into the existing analogue records management system. One exception to this observation occurred with Web sites, which both large and small organizations frequently treated almost as ephemera—insofar as organizations made little or no effort to set aside and preserve fixed versions of their Web sites—although this was balanced by those organizations that viewed Web sites as either a type of "recordkeeping system"[35] or as a type of legacy site on which to preserve their work.[36]

[34] Ibid., 3–4.
[35] See, for example: Henry Daniel and Cara Payne (2004), "InterPARES 2 Project - Case Study 02 Final Report: Performance Artist Stelarc," 7–13. Available at http://www.interpares.org/display_file.cfm?doc=cs02_final_report.pdf.
[36] See, for example: Martine Cardin (2004), "InterPARES 2 Project - Case Study 01 Final Report: Arbo Cyber, théâtre (?)." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs01_final_report_english.pdf. Original French language version available at http://www.interpares.org/display_file.cfm?doc=ip2_cs01_final_report_french.pdf.

With individuals and small organizations, in far too many instances, there were no preservation practices to study and little evidence that even basic maintenance strategies were being pursued with any consistency.[37] In many cases, the maintenance procedures that were in place were more the result of a happy accident than of an intentional act of preservation. The most popular strategy was undoubtedly "redundancy," or making a copy somewhere other than on the creator's computer or server hard drive. For large organizations, regular backups run by information technology staff were and are now standard practice.[38] For smaller organizations and individual users, some form of backup was frequently cited.[39] One improvement in the redundancy strategy is the disappearance of the old DOS-based concept of "backup" formats, which were highly sensitive to upgrades in the operating system. Steady decreases in the cost of storage means more copies are basic copies; that is, copies saved in the file formats in which the records were originally created, or in which they were last used and saved, thus (theoretically) making them more immediately human-readable in the creator's usual desktop environment. Directory structures and file names are immediately accessible without the need to apply an extra "Undo" command to re-open the—often compressed—backup or protection copy. Where very large files are concerned, use of compression may reintroduce the extra layer of software needed to re-access the backup copy of the records.

These redundant copies were also, in some cases, stored in off-site locations. Perhaps the most comprehensive example of this approach is found in case study 21 (Electronic Filing System (EFS) from the Supreme Court in Singapore), where it is noted:

> There is both a daily and weekly backup of data, which are kept in an off-site location to ensure the full restoration of data in the event that the system fails. For example, every time a record is filed by the law firm, the records are stored on three disks. One disk is stored permanently in the jukebox to facilitate online access to information. The second disk is taken to an off-site storage at the end of the week and the third disk is sent off-site once the disk is full.[40]

In some cases, redundant copies were distributed between the creator's home and office,[41] neither of which offered stable environmental conditions such as temperature, relative humidity

[37] Eight basic maintenance strategies are described in an appendix to the *Preserver Guidelines* for developed by Domain 3 researchers and appended to this report (see Appendix 21c). In brief, these strategies are: (1) clear allocation of responsibilities; (2) provision of the appropriate technical infrastructure; (3) system maintenance, support and replacement; (4) transfer of data to new storage media on a regular basis; (5) adherence to appropriate conditions for storage media; (6) redundancy and geographic location; (7) system security; and (8) disaster planning.

[38] See, for example, in the arts focus, James Turner, et al. (2004), "InterPARES 2 Project - Case Study 09(3) Final Report: Digital Moving Images - Commercial Film Studio," 8–10. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs09-3_final_report.pdf; in the science focus, Tracey P. Lauriault and Yvette Hackett (2005), "InterPARES 2 Project - Case Study 06 Final Report: Cybercartographic Atlas of Antarctica," 17. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs06_final_report.zip; and, in the government focus, Elaine Goh (2005), "InterPARES 2 Project - Case Study 21 Final Report: The Electronic Filing System (EFS) of the Supreme Court of Singapore," 32. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs21_final_report.pdf.

[39] See, for example, reliance on backup to: (1) CD-ROM or DVD in Daniel and Payne, "Case Study 02 Final Report," op. cit., 13, and Bart Ballaux (2005), "InterPARES 2 Project - Case Study 26 Final Report: MOST Satellite Mission - Preservation of Space Telescope Data," 7. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs26_final_report.pdf; and (2) separate computer(s) in Sydney Fels and Seth Dalby (2004), "InterPARES 2 Project - Case Study 15 Final Report: *Waking Dream*," 4. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs15_final_report.pdf. See also Ballaux, "Case Study 26 Final Report," op. cit.

[40] Goh, "Case Study 21 Final Report," op. cit. See also Ballaux, "Case Study 26 Final Report," op. cit., for an example of the use of off-site backup procedures by a creator in the science focus.

[41] See, for example, the strategy used by the creator in case study 14, in which copies of the GIS data and related records are "maintained," in an ad hoc manner, "on either the creator's personal computer, the organization's server and/or copied onto CD-ROMs" (Richard Pearce-Moses, Erin O'Meara and Randy Preston (2004), "InterPARES 2 Project - Case Study 14 Final Report:

or clean air. Moreover, although this type of "informal" distributed storage strategy does offer increased protection against vandalism, theft or fire, it may be of little use in larger-scale disasters such as earthquakes or floods.[42] Aside from the limited data recovery functionality provided by the relatively basic, and often ad hoc, practices noted above involving *selective backups* of data or files,[43] formal and explicit disaster planning was almost non-existent among all the creators examined.[44]

A number of the case studies and surveys also confirmed what the Domain 3 researchers already knew based largely on their own records creation and maintenance practices. For example, most creators learn only the minimum required about a technology to get it working and perform basic functions. Few users become experts or learn to appreciate the implications for long-term records preservation of, for example, configuration choices, or add-on programs generated by third-parties, or relevant functionality within their chosen software environment.[45]

The Domain 3 researchers observed that, especially in the arts, hardware components, particularly peripherals, can play an essential role in the accurate and authentic reproduction of, for example, a work of art.[46] However, hardware dependencies can be almost impossible to overcome, especially if responsibility for preservation is transferred to a trusted custodian who lacks access to the required hardware and/or the ability or resources to maintain such hardware. Acknowledgement of this fact led the Domain 2 researchers to recommend that, whenever possible, records creators should strive to "eliminate dependence on hardware by transferring hardware functionalities to software (i.e., use a software application to simulate the actions of a

Archaeological Records in a Geographical Information System: Research in the American Southwest," 28. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs14_final_report.pdf.

[42] Of course, the same concern may also apply in more "formal" distributed storage arrangements, such as the one noted above in case study 21, depending on the nature of the off-site storage facilities used and their geographic proximity to the on-site location.

[43] This is in contrast to *comprehensive system backups* in which the operating system, all software applications and all data/files are backed up so that the *entire system* can be restored in the event of a large-scale system failure (see *Creator Guidelines*, guideline 8.A, in Appendix 20). Two notable exceptions to the otherwise exclusive reliance on selective backup strategies were observed in the case studies; one in the science focus, the other in the government focus. In case study 26 (MOST Satellite Mission), it is stated that, "[s]ince the MOST researchers mostly work with custom-made software, one of the issues related to their work is the preservation of the software. Backups are made of the various programs that are used in the project. Moreover, if anything is changed (i.e., added) to one of the programs, the old version of the software is always preserved. In this way, the researchers are always capable of recreating the results that they created previously" (Ballaux, "Case Study 26 Final Report," op. cit., 7–8). In case study 12 (Antarctic Treaty Searchable Database), it is noted that "the digital-record entities in the *Antarctic Treaty Searchable Database* have been maintained on servers with backup copies on additional hard-drives as well as on webCDservers ... which replicate the full functionality and contents of the website.... In one instance, the webCDserver was used to restore the website for the *Antarctic Treaty Searchable Database*" (Paul Arthur Berkman et al. (2005), "InterPARES 2 Project - Case Study 12 Final Report: Antarctic Treaty Searchable Database," 39. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs12_final_report.pdf). Although reference to the backed up entities in the foregoing quote is limited to "the digital-record entities," it seems apparent that, in fact, copies of the operating system and the *Digital Integration System* software application are also included on the backup webCDservers, which the creator elsewhere describes as containing "fully executable copies of the websites" (Ibid., 18).

[44] Case study 09(3) (Commercial Film Studio) provides the only explicit reference to disaster planning (see Turner et al., "Case Study 09(3) Final Report," op. cit., 15 (see answer to question 3.2).

[45] An example of this is provided in general study 09 (Digital Recordkeeping Practices of GIS Archaeologists), where it is noted that "several of the [survey] participants admitted feeling somewhat overwhelmed by the analytical potential of their GISs, due, primarily, to a lack of formal GIS training." As the report cautions, "[i]f true, this could potentially have ramifications with regard to the broader issue of the long-term preservation of digital GIS records. In fact, it is possible that the more alienation, disconnection and/or intimidation one feels toward one's GIS projects, the less likely one may be to feel the sense of obligation, initiative and/or competence necessary to effectively address the project's long-term preservation requirements" (Preston, "General Study 09 Final Report," op. cit., 32).

[46] See, for example, the discussion in case study 13 (*Obsessed Again...*) of an attempt by the Domain 2 researchers to resurrect an electroacoustic work that had already fallen victim to technological obsolescence. A more concise summary of this particular activity is also provided in the section titled "A strategy for preventing technological obsolescence of an artistic work" of the Domain 2 Task Force Report. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_3_domain2_task_force.pdf.

piece of hardware," [since] "[t]his provides a more stable way to retain the function when the hardware becomes obsolete."[47]

Another observation noted by the researchers is that a number of creators are simply ignoring preservation issues, although here too the reasons are varied. In some cases, creators (especially those in the arts focus, but also in the science focus) noted that it would be easier and less expensive to re-create some types of content in digital form at a later date than it would be to attempt to carry forward and convert or provide an emulation platform for highly proprietary material. This view is best epitomized in the following excerpt from the arts focus case study 09(3) (Commercial Film Studio):

> Since important hardware and software changes have usually occurred since the artwork was created, in the environment we studied it is deemed more economical to re-create artwork if it is to be reused than to engage in a process of migrating or otherwise upgrading it in case it will be reused. This approach limits the amount of time, energy and money spent on long-term preservation of digital entities and avoids the need to implement preservation strategies that respond to problems of hardware and software obsolescence.[48]

A similar sentiment was expressed by the creator in the science focus case study 14 (Archaeological Records in a Geographical Information System), who, in response to a question about the intermediary files (e.g., cost surface algorithms) he creates and their relationship to his overall GIS research, stated:

> I guess I don't ascribe a lot of value to those intermediary documents for the most part, because they're easily recreated ... Whenever I'm trying to teach somebody about GIS, people will say, here's the data I have and here's what I want to get to. And the first [thing] that I always [say] – my first caveat that I offer them is, if you ask 10 different GIS people, they'll tell you 10 different ways to get there. Each one of those different ways will produce different intermediary files, and so I don't think, for the most part, that those things have much value. Like I said, they're easy to reproduce, if I did want to, I could always make another slope file, if I lose one or something.[49]

In other cases, some artists do not want their works preserved over the long term if doing so might compromise certain characteristics that they consider essential to the essence of their artworks, such as the ephemerality and variability of works that change each time they are reproduced. An example of this is provided in case study 15 (*Waking Dream*), which examines a multimedia performance art piece. One of the performers and co-creators of the work, Sachiyo Takahashi, "views *Waking Dream* as a performance art piece defined by her role as performer," a view that she believes "precludes performances of the piece without her involvement [as one of the performers]."[50] Based on this interpretation, it would be impossible to "preserve" the ability to reproduce the work or, at least, an authentic reproduction of the work beyond the performer's death.

These observations led InterPARES researchers to further explore and refine their understanding of the distinction between digital documents that are either works or that *document* specific performance events, versus digital objects that either *enable* subsequent

[47] See *Creator Guidelines*, guideline 9, in Appendix 20. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_20.pdf.
[48] Turner et al., "Case Study 09(3) Final Report," op. cit., 4.
[49] Pearce-Moses et al., "Case Study 14 Final Report," op. cit., 31.
[50] Fels and Dalby, "Case Study 15 Final Report," op. cit., 8.

reproduction of a work or performance—a concept more closely analogous to a "negative" in photography—or provide *instructions* about executing a work or performance—a concept more closely analogous to a "score" in music—with the added fact that these *enabling* and *instructive records* will actually play an active or an instructive role in the re-instantiation of the work or performance. In contrast to instructive records, which "are intended to be read by humans and, therefore, are materialized by being reproduced from stored digital components into a human-readable form,"[51] a key characteristic of enabling records, whether in the arts[52] or in the manufacturing sector where other examples were found,[53] is that they contain instructions intended to be read by a machine and thus

> …achieve their purpose in the digital form in which they are stored[64] and, conversely, cannot achieve that effect if transformed into human-readable format. Moreover, as long as they remain active, enabling records must be maintained in the systems in which they were created—or in systems with identical functionality. Otherwise, they will not produce or enable the interactions, experiences, performances or other processes they were intended to generate.[65 54]

Still other creators feel they can either delay addressing preservation issues—or ignore them altogether—because they believe that the information technology industry will come up with whatever solution will be needed, whenever they need it. "Trusting the vendor" may work in cases where a technology has been widely adopted and there is an equivalent product on the market to which consumers could move. On the other hand, it is instructive to remember how many physical media since the 8-track tape have been discontinued by manufacturers, even though obsolescence of such media meant, for example, that consumers had to migrate their entire music or video collections to the new technology.

This willingness to "trust the vendor" (or the technology industry in general) seems to be the assumption driving the adoption of digital signature technology despite the well-known problems with carrying this "authenticity" solution forward over time.[55] This, in fact, appears to be the

[51] Luciana Duranti and Kenneth Thibodeau (2006), "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science* 6(1): 60. (Note: a reprint of this article is included in Appendix 2).

[52] For example, the custom written program code used to control the functionality of a remote control device used by a performer in *Waking Dream* (see Fels and Dalby, "Case Study 15 Final Report," 2 (technological context)), and the various custom written software patches used in *Obsessed Again...*, including a software synthesizer patch of coded instructions that controls the sounds that the computer causes to be played on the synthesizer in response to the notes played by the bassoonist (see J. Scott Amort (2004), "InterPARES 2 Project - Case Study 13 Final Report: *Obsessed Again...*," 3 (see answer to question 4). Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs13_final_report.pdf.

[53] For example, the Pro-Engineer CAD system file used to create the original CAD files in case study 19 (see Kenneth Hawkins (2006), "InterPARES 2 Project - Case Study 19 Final Report: Preservation and Authentication of Electronic Engineering and Manufacturing Records," 6. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs19_final_report.pdf.), meets the definition of enabling record in that it stands as a fixed set of instructions only readable by a machine and that supports subsequent actions and provides instructions to engineering robots, etc. The addition of semantic or actionable metadata to the subsequent iterations of the same fixed information provides the basis for additional subsequent actions; namely, the "interrogation" of the knowledge-enhanced formats by reasoning engines.

[54] Duranti and Thibodeau, "The Concept of Record," op. cit. Note: footnote references in the quote are from the original text, and are not reproduced here. As Duranti and Thibodeau further clarify, "[e]xamples of "enabling" records include software patches that enable a musical instrument to interact with a computer, software in online marketing sites that interprets data about a visitor's actions on the site to determine what elements of content should be presented next to that visitor, and software agents that enable interacting business applications to execute transactions autonomously" (Ibid., 59).

[55] As is emphasized in the *Creator Guidelines*, guideline 6 (see Appendix 20, op. cit.), *technology-dependent* authentication techniques, such as digital signatures, are subject to obsolescence themselves. In fact, by virtue of their purpose and inherent functionality, digital signatures cannot, at present, be migrated to new or updated software applications together with the documents to which they are attached. Moreover, the life of digital signatures and other authentication technologies may be much shorter than the length of time that even a temporary document not requiring migration may need to be maintained, because authentication technology is changing rapidly. A number of different preservation strategies (see Appendix 21c, Part B) would

situation in case study 18 (Computerization of Alsace-Moselle's Land Registry), where a specifically dedicated administrative body, GILFAM (Groupement pour l'Informatisation du Livre Foncier d'Alsace-Moselle), created in 1994 and charged with overseeing computerization (and digitization) of France's land registry

> …has the legal responsibility to provide continued access to the land registry in a fashion that preserves its evidential value, in conformance with Article 1316-1 of the Civil Code, regardless of technological change, [yet] has not considered the problem of maintaining the digital entities, except through the mechanisms afforded by system vendors when upgrading the database management system. In addition, it has not considered the problem of maintaining the evidential value of digital signatures through technological evolution.[56]

Still other creators feel confident about their ability to "self-monitor" the evolution of technology and standards affecting the long-term preservation of records and to adapt their preservation systems and practices accordingly in time to keep pace with the changes. A good example of this approach is provided in case study 21, where the creator (the Supreme Court of Singapore) describes its current, apparently ad hoc, approach to preservation as consisting of an attempt "to foresee new standards/technology before the current technology that has been deployed becomes fully obsolete," so as to provide "sufficient time to migrate to the new technology/standards without loss of data."[57] Although monitoring the technological environment and taking steps against hardware and software obsolescence are key recommendations provided in the *Creator Guidelines*,[58] the activities must be supported by an integrated framework of policies and procedures to be effective. Unfortunately, this does not appear to be the situation in the case of the Supreme Court of Singapore, since it is acknowledged later in the report that although "[t]he court recognizes that there is a need to address the long-term preservation of electronic records … there is currently no strategy in place *because the court views the system to be fairly current*."[59] The reason provided by the creator in this case for the delay in developing a long-term preservation strategy demonstrates a general concern identified by the Domain 3 researchers in relation to the preservation practices of many of the creators they studied. In fact, beyond a lack of foresight, it shows, more seriously, a fundamental misunderstanding of the temporal scope of the problem at hand and a lack of awareness of the fact that effective long-term preservation begins at creation.

The Domain 3 researchers observed that the technology used by innovators and early adopters, regardless of the focus area in which they belonged, was proprietary[60] and frequently customized. In many cases, the point of the work of these types of creators is to explore, test and

---

eventually require the conversion of a digitally-signed record to a new logical format. Following this process, it is unlikely that the document and its signature could be re-validated by the external digital signature infrastructure since the check-sum will have changed. Consequently, creators and preservers are encouraged to rely on *technology-independent* authentication techniques, grounded in effective administrative policies and procedures for records creation, maintenance and preservation, whenever possible.

[56] Jean-François Blanchette, François Banat-Berger and Geneviève Shepherd (2004), "InterPARES 2 Project - Case Study 18 Final Report: Computerization of Alsace-Moselle's Land Registry," 27. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs18_final_report.pdf.

[57] Goh, "Case Study 21 Final Report," op. cit., 33.

[58] See Appendix 20, guideline 9, op. cit.

[59] Goh, "Case Study 21 Final Report," op. cit., 38 (emphasis added).

[60] Proprietary either in the sense of commercially available software, the source code of which is privately owned and controlled, or in the sense of custom-built software, the source code of which is developed "in-house" and is not made publicly available. The custom-built software used to process satellite data in case study 26 (MOST Satellite Mission) is a good example of the latter.

push the limits of the available technology, be it hardware or software. This environment cannot wait for internationally agreed-on standards or "open" formats.[61] It requires instead a robust interoperability,[62] at least until such time as widespread adoption of the technology generates some standards, should this ever occur. The Policy Cross-domain, in its final report, noted that "steps taken to ensure interoperability across systems performed many of the same purposes as preservation."[63]

Many scientists view the publication of an article in a journal as a means of preservation, instead of attempting to preserve any of the digital antecedents that led to the published item. This belief was expressed, for example, by many of the survey participants in general study 09 (Digital Recordkeeping Practices of GIS Archaeologists) in response to a question asking them to identify what they considered to be the most important elements or outputs of their GIS projects to preserve for future use or reference, and why. As the study's final report notes, "many participants assume that publication alone constitutes sufficient long-term preservation of their research."[64] In light of this finding, the report goes on to caution that

> …by focusing on the preservation of final reports (and related types of documentation, such as published articles), at the expense of more comprehensive on-going project documentation, archaeologists run the risk of not preserving the types of records required to ensure that the key preservation goals identified by survey participants—i.e., project/data reuse, verification, replication and accountability—can in fact be met.[65]

Government employees may also feel this way about reports that document a study, which they favour for preservation to the underlying data and research that informed the study. This is clearly inadequate, since loss of the data in a manipulable form prevents verification of results, potentially impedes successful duplication of the data-gathering experiment if it is replicable or destroys a particular point in a time series dataset. Moreover, this "publication" argument has lost much of its relevance with the move to electronic journals and reports, where it is clear that publication, even in digital form, cannot replace the original data. The new digital platform for e-journal preservation has created an environment where the related data may well be preserved and can be linked to the publication.

Several case study interviews show that, for many creators of digital objects, there is already a digital equivalent to the concept of trust in the long-term survival of traditional publications: "putting it on the Web." This is, however, a misguided view since there is no organization even attempting to preserve the entire World Wide Web and the millions of changes made to its content every second.[66] In fact, these two proposed preservation strategies are manifestations of

---

[61] For a discussion of the use and understanding of the concept of "file format" in the context of archival institutions, including the terms "standard file format" and "open file format" (also known as "open-source" and "non-proprietary" file format), see McLellan, "General Study 11 Final Report," op. cit.

[62] The InterPARES 2 Terminology Database defines "interoperability" as "the ability of one application/system to communicate or work with another" (http://www.interpares.org/ip2/ip2_terminology_db.cfm).

[63] Policy Cross-domain Task Force Report, 11. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_7_policy_task_force.pdf.

[64] Preston, "General Study 09 Final Report," op. cit., 4.

[65] Ibid., 91.

[66] See, for example, case study 03 (*HorizonZero*/ZeroHorizon Online Magazine), where the creator notes that, in response to the question *What preservation strategies and/or methods are implemented and how*? (research question 19a), "[m]aintenance of the Web site will be contracted for a period of ten years, though much of the interaction (chat rooms, message boards) will be disabled" (Brent Lee (2004), "InterPARES 2 Project - Case Study 03 Final Report: *HorizonZero*/Zero Horizon Online Magazine and Media Database," 9. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs03_final_report.pdf). Likewise, in case study 02 (Performance Artist Stelarc), in response to the same research question, the creator states that "[e]ach performance

the most widely held belief about digital preservation that the case studies uncovered; namely, that "it must be somebody else's problem."[67] This belief can target the specific—somebody else in the organization is in charge—or the general—for example, a state institution or some other research and resource-rich entity or organization must be doing something.

Another variation on this "it must be somebody else's problem" approach is to *purposely* transfer responsibility for long-term maintenance and preservation to somebody else. An example of this approach is provided in case study 21, where, in relation to concerns of the Supreme Court of Singapore over its ability to continue to manage the public key infrastructure (PKI) that it uses to issue digital certificates to court solicitors, the Court is considering outsourcing management of its PKI activities to a licensed certification authority so that it can be "insulated from managing technology obsolescence."[68]

Somewhat related to the "it must be somebody else's problem" pretext is a paralysis towards preservation due ostensibly to concerns about cost but that, in fact, often belie an underlying malaise or lack of will, especially among organizations, to address preservation issues and/or to provide access to preservation resources for preservation-conscious individuals within an organization's sphere of operations. Concerns regarding this latter issue were most notably expressed by scientists decrying the conspicuous lack of suitable repositories for preserving scientific data. For example, as noted earlier, there was a palpable sense of frustration (even, in some cases as the report states, desperation) expressed by many of the participants in the general study 09 (Digital Recordkeeping Practices of GIS Archaeologists) survey over the continuing lack of suitable long-term preservation repositories available for digital archaeological data and records.[69] A slightly different perspective on this situation was expressed by the creator in case study 06 (Cybercartographic Atlas of Antarctica), who noted that "[d]ata creators have few preservation incentives beyond meeting scientific and professional requirements and the peer review process,"[70] suggesting that, in many cases, individual researchers are equally culpable when it comes to the current lack of will to pursue effective, long-term preservation of scientific research records and data. As offered by the creator in case study 06, one suggestion to help address this situation is to press "[r]esearch funding agencies … to include preservation as part of their award structures and also … provide institutional support and … require policy frameworks."[71] Finally, part of the blame for the continuing reluctance of both individuals and organizations, particularly in the sciences, to acknowledge their role in the preservation lifecycle rests in the very nature of the materials requiring preservation, both in terms of their sheer quantity and their internal and external complexity. In fact, aggregations of data, and their associated documents and records, are often so large, complex (e.g., incorporating integrated multimedia, interactivity, etc.) and distributed (both within and among multiple creators), that it

generates its own documentation, each of the drawings, photos and videos made are converted to appropriate digital formats when needed, and are then posted to the Web site" (Daniel and Payne, "Case Study 02 Final Report," op. cit., 14).

[67] A telling example of this is provided in general study 09, which characterized the rationale "as a "middle man mindset," in which the [survey] participant (especially those working as consultants) considers long-term preservation to be solely their client's problem" (Preston, "General Study 09 Final Report," op. cit., 54). As one of the study's survey participants matter-of-factly stated, "we provide the data to our clients and what they do with it is generally up to them," (Ibid.) seemingly oblivious to the fact that, in the digital environment, preservation begins with the creator and extends through to the preserver and is, therefore, a continuous and *distributed* responsibility.

[68] Goh, "Case Study 21 Final Report," op. cit., 33.

[69] Preston, "General Study 09 Final Report," op. cit., 90. See also specific participant comments regarding this issue, summarized on p. 54 in the Final Report.

[70] Lauriault and Hackett, "Case Study 06 Final Report," op. cit., 31.

[71] Ibid.

is difficult to find organizations that are willing, let alone able, to take on the responsibility for authentic, long-term preservation.

This exact situation appears to be exacerbating the preservation efforts of the creator in case study 06, who is working with the Carleton University Library "to attempt to archive the CAA [Cartographic Atlas of Antarctica], as it exists at the end of the project, as per SSHRC [Social Sciences and Humanities Research Council of Canada] requirements," since "at the moment, there are no institutions capable of archiving SSHRC supported project data and results."[72] Moreover, although it is expected that the Scientific Committee on Antarctic Research (SCAR) will eventually take over responsibility for the long-term maintenance and future development of the CAA, the creator acknowledges that this is by no means certain, given SCAR's limited human, financial and institutional resources, as well its limited technical capacity.[73]

Another variation on the publication/put-it-on-the-Web preservation strategies sees digitization as a solution. This view is generally offered either by creators: (a) who have a physical storage space problem, (b) who have trouble finding specific items among their analogue records, (c) who feel compelled (and in some cases are required by law) to oblige increasing public demand for remote electronic access to digital records and other resources[74] or (d) who see digitization as a means to effect preservation through increased access to, and diffusion of, the digitized resources.[75] For many, technology is touted as the solution to each of these concerns or issues. These assertions, generally, come from creators who are unfamiliar with the technical difficulties and the recurring costs involved in the long-term preservation of digital objects.

Where authenticity is concerned, most records creators saw little need to actively protect it in the digital environment. Reasons were varied among the individual and small group case studies and included the lack of monetary value inherent in the object;[76] the small number of practitioners in a field where everybody knew everybody and would recognize the work;[77] or the desire to make the content freely available for re-use by others, sometimes in an artistic context,[78] sometimes in a scientific context.[79] For the most part, individuals and small groups in

[72] Ibid., 18, 29.
[73] Ibid., 18.
[74] (a), (b) and (c) were all cited as motivating reasons behind the digitization of land registers discussed in case study 18 (see Blanchette et al., "Case Study 18 Final Report," op. cit.).
[75] This concern is very aptly demonstrated by the informant in case study 01 (Arbo Cyber, théâtre (?)), who, while acknowledging that the theatre "group is more concerned with digitization as a means of diffusion," nevertheless "see digitization as a better means of preservation than the traditional recordkeeping system" (Cardin, "Case Study 01 Final Report," op. cit., 48..
[76] For example, as is noted in case study 09(3), "[a]nalogue material is kept largely because of its commercial value. Individual animation cels or other artwork can be sold in galleries or at auction, but digital files have no value or only ephemeral value for such purposes. Ironically, digital files can take on value if they are touched by famous people, for example a celebrity may have written on the label of removable media even though he/she never used the computer with which the file was created" (Turner et al., "Case Study 09(3) Final Report," op. cit., 5). On the other hand, as is also stated later in the same report, "[o]ther marketing and promotional uses include re-use of digital entities for interviews with animators, party events, awards, and so on, such as value-added material on DVD versions of the studio's films" (Ibid., 8).
[77] Such as the work of performance artist Stelarc, who, for example, believes that the authenticity of the digital objects emanating from his performances "is assured primarily because of his own unique position at the centre of the entire process and the unique nature of the performance events" (Daniel and Payne, "Case Study 02 Final Report," op. cit., 12). This concept of authenticity relates closely to the concept of "personal" authenticity identified by the Domain 2 Task Force, in which "authenticity denotes the degree to which an artwork manifests the individuality and essence of its creator ... [such that] the artist is the artwork, unmediated by any records." Thus, "[t]he prevalence of this notion of authenticity explains why many artists do not concern themselves with explicitly marking the identity of their works; to them it is inconceivable that anyone else either could or would produce art like theirs. Anything an artist makes (or directs the making of) is authentic, by this definition" (Domain 2 Task Force Report, op. cit., 27).
[78] The most explicit example of this is found in case study 01, where it is stated that "[t]he concept of re-use is so important to the [Arbo Cyber, théâtre (?)] group, that the informants felt that even after the dissolution of the group, artists should be able to later

the arts focus seemed content with the understanding that they are the arbiters of authenticity during their lifetime. In fact, for example, in the interactive multimedia installation of case study 10, direct stewardship over the digital objects in question by their individual creators was believed by all subjects in the study to be the only reliable means of maintaining and assuring the authenticity of those objects, such that the authenticity of the objects could no longer be assured after the death of their creator.[80]

In contrast, those in organizations tended to trust that somebody else was responsible for implementing necessary measures to ensure authenticity and effect preservation—in some case studies this was true, while in others it was not. There were also examples, as mentioned earlier, of continued use of paper-based record systems, especially in environments where legal issues related to evidential value—especially in relation to intellectual property rights—might emerge at a later date.[81]

Where questions of safeguarding the identity and integrity of the records are concerned; for example, with questions of copyright and intellectual property, the findings of InterPARES 1 related to the importance of consistently and properly recorded elements of metadata was again a primary finding.[82] An interesting variation on the role of metadata was introduced in some of the arts-based case studies. As with authenticity, individual creators could only control how their works might be re-presented during their lifetime. For an accurate (and authentic) understanding of their intent for re-presentations of their works beyond that time frame, it would be necessary for the parameters of that intent to be recorded and carried forward, most probably in metadata.[83] When asked to identify, from among a list of six options, which method they believed was the best approach for representing the identity of musical works that lack a musical score (in the traditional sense), nearly one quarter (24%) of those who participated in the general study 04 survey of the recordkeeping practices of composers indicated that the identity of such works is best represented by a audio or video recording of the performance,[84] suggesting that recording a performance that fulfilled the composer's vision for the work would also provide a method to fix the composer's intent.

---

use records in future individual or other group projects. In this sense, the informants treat the *Ludosynthèse* not only as a testament to past performances, but also as a source of information for future use. It is creating memory" (Cardin, "Case Study 01 Final Report," op. cit., 23). In fact, one of the primary purposes of the *Ludosynthèse*—the group's interactive and dynamic Web site—is to allow "spectator-users to develop or recreate performances in Arbo style using digital media" (Ibid., 3).

[79] In a sense, the Antarctic Treaty Searchable Database, which "has been designed to facilitate knowledge discovery about the policies and strategies that promote *"international cooperation*" and *the "use of Antarctica for peaceful purposes only"* [...] *"in the interest of all mankind,"* as promoted in the *Preamble* of the *1959 Antarctic Treaty*" (Berkman et al., "Case Study 12 Final Report," op. cit., 9, 34 (emphasis as in original)), also falls into this category, in as much as the database serves as a portal to public-domain policy documents whose authenticity, the database's creator insists, "can only be assured by the government agencies that issue the records" (Ibid., 29).

[80] Sally Hubbard (2006), "InterPARES 2 Project - Case Study 10 Final Report: *The Danube Exodus*," 8. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs10_final_report.pdf.

[81] See, for example, case study 03—which analyzed the recordkeeping activities of an online arts magazine (*HorizonZero*)—where it is noted that e-mail, contracts and other legal documents are kept in analogue form in a paper-based filing system apart from the creator's digital "recordkeeping" system where all the other documents and digital objects created by *Horizon*Zero are kept (Lee, "Case Study 03 Final Report," op. cit., 4, 6).

[82] See case study 06 (Cybercartographic Atlas of Antarctica) for an excellent example of this (Lauriault and Hackett, "Case Study 06 Final Report," op. cit).

[83] The necessity of such a process was clearly demonstrated in the aforementioned attempt to resurrect and stage an authentic performance of an electroacoustic work, the partial failure of which led the Domain 2 researchers to conclude that "[c]reators, while they are still living, are the best arbiters of the authenticity of performances. So it behooves them to describe their works in technologically independent (and authentically preservable) ways that will allow authentic performance in the future" (Domain 2 Task Force Report, op. cit., 37).

[84] Longton, "General Study 04 Final Report," op. cit., 4 (see question 11).

Perhaps the most promising and forward-thinking preservation strategy encountered by the Domain 3 researchers was observed in science focus case study 08 (NASA). Here the creator claims to have implemented preservation procedures that support *persistent object preservation* (POP), which is "a technique to ensure electronic records remain accessible by making them self-describing in a way that is independent of specific hardware and software."[85] In NASA's case, this strategy relies, in part, on the use of Object Description Language (ODL)[86] to "create labels (data descriptions) for data files and other objects such as software and documents."[87] Under this strategy, technology obsolescence requires only the migration of the interpreter for the file structure description language, rather than the creation of access software for each of the file structures.[88] The fact that, in the fifteen-plus years that the PDS has been operational, "it has not been necessary to update (convert or migrate) any of the data products to other data formats,"[89] suggests that this strategy may indeed offer a viable solution for at least one aspect of the long-term preservation of digital materials.[90]

Addressing the Research Questions

Question 1

> *How do the appraisal concepts, methods and models developed by InterPARES 1 for the administrative and legal records created in databases and document management systems apply to the appraisal of the records of artistic, scientific and governmental activities resulting from the use of the technology examined by InterPARES 2?*[91]

The InterPARES 2 case studies revealed more acute examples of problems that have always challenged the preserver's ability to appraise records. Lack of metadata, idiosyncratic arrangements and fragmented storage arrangements are just a few examples. More unusual, perhaps, is the reminder that digital records, especially non-textual ones, may still have specific hardware dependencies, a situation now largely absent from the text-based recordkeeping environments of government institutions and corporations that were studied in InterPARES 1.

[85] Richard Pearce-Moses, "Persistent Object Preservation," in *A Glossary of Archival and Records Terminology* (Chicago: The Society of American Archivists, 2005). Available at http://www.archivists.org/glossary/index.asp. The use of "self-describing formats" (i.e., POP or tagging) is one of the twelve basic preservation strategies (B1.1) described in an appendix to the *Preserver Guidelines* developed by Domain 3 researchers (see Appendix 21). See also "Digital Records Maintenance and Preservation Strategies," in Appendix 21c.

[86] ODL is a language created by NASA that is used to encode data labels for its Planetary Data System (PDS) and other data systems (see National Aeronautics and Space Administration, "Chapter 12. Object Description Language Specification and Usage," in *Planetary Data System Standards Reference*, version 3.7, March 20, 2006. JPL D-7669, Part 2 (Pasadena, CA: Jet Propulsion Laboratory, California Institute of Technology, 2006). Available at http://pds.nasa.gov/documents/sr/.

[87] William Underwood (2005), "InterPARES 2 Project - Case Study 08 Final Report: Mars Global Surveyor Data Records in the Planetary Data System," 15. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs08_final_report.pdf.

[88] Ibid., 29.

[89] Ibid.

[90] It is worth noting that, based on the repeated and consistent successes of tests of the POP strategy by the Distributed Object Computation Testbed (DOCT), an interagency collaboration between the U.S. Department of Defense's Advanced Research Projects Agency, the U.S. Patent and Trademark Office and the U.S. National Archives and Records Administration (NARA), NARA regards the POP "approach as the most promising one ever suggested for preserving digital information in general, and electronic records in particular" (Kenneth Thibodeau (2001), "Building the Archives of the Future," *D-Lib Magazine* 7(2). Available at http://www.dlib.org/dlib/february01/thibodeau/02thibodeau.html), and has in fact chosen POP as the preservation method on which its soon-to-be-launched Electronic Records Archives (ERA) will rely (see http://www.archives.gov/era/about/).

[91] InterPARES 2 Project, Domain 3 Web page. Available at http://www.interpares.org/ip2/ip2_domain3.cfm.

The absence of selection criteria was also noted in relation to the types of records being studied in a number of case studies, but the development of selection criteria is not one of the preserver's functions seriously affected by digital technology, beyond the added importance of functionality for certain types of digital records, such as those contained in database applications.

The InterPARES 1 case studies also clearly demonstrated the need, during appraisal, to confirm whether the systems under review contained or could produce records. This situation remains essentially true for the interactive, experiential and dynamic applications studied in InterPARES 2, although the process may prove to be more complex, given the larger number of components involved and given their increasingly complex relationships.

The study of interactive, experiential and dynamic cases in InterPARES 2 considerably enriched the articulation of the nature of digital records that was provided by InterPARES 1. The implications of the findings of the InterPARES 2 case studies on the definition of a record are discussed extensively in the article *The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES.*[92] One of the major findings of InterPARES 1 was that "it is not possible to preserve a digital record: it is only possible to preserve the ability to reproduce the record."[93] That finding was based on the idea that it is not possible for a stored record written in binary code form (a form that is illegible to humans) to be capable of serving as a record. The findings of the InterPARES 2 Project, informed by its expanded scope to interactive, experiential and dynamic electronic systems, combined with a re-emphasis on the core definition of "record" and supported by precedents dating back centuries, demonstrate that this finding is valid only for a subset of digital records; namely those that, to be effective, must be manifested to a person.

InterPARES 2 case studies in the arts and in the sciences (specifically manufacturing) identified *enabling records* kept in, or associated with, computer applications or systems that are never intended to output a human-readable document. The computer instructions and specifications that enable performance of digital music works and that drive the computer-assisted manufacture of physical parts are obvious examples. Their purpose is not (re)production of a human-readable document, but (re)performance or (re)production with the mediated assistance of a computer. The definition of a record imposes no a priori constraints on the content or form of a record. The determination of whether a digital document[94] is a record depends on whether it contains and presents the necessary elements and attributes of a record, such as whether it participates in an action; has an archival bond; has an author, writer and addressee, etc.[95] Any digital document that satisfies these requirements is a record, even though it may be utterly impossible for a human to understand the record in its digital form.

[92] Duranti and Thibodeau, "The Concept of Record," op. cit., especially 46–60.
[93] Thibodeau et al., "Preservation Task Force Report," op. cit., 106.
[94] The InterPARES 2 Terminology Database defines "document" as "An indivisible unit of information constituted by a message affixed to a medium (recorded) in a stable syntactic manner. A document has fixed form and stable content," while "digital document" is defined as "A digital component, or group of digital components, that is saved [i.e., affixed to non-volatile storage on a digital medium] and is treated and managed as a document." A digital object is, in turn, defined as "A discrete aggregation of one or more bitstreams and the metadata about the properties of the object and, if applicable, methods of performing operations on the object."
[95] For a detailed discussion of all the necessary elements and attributes of a record, see the InterPARES 1 *Template for Analysis*, which is included as Appendix 1 in Duranti (ed.), *The Long-term Preservation of Authentic Electronic Records*, op. cit., 192–203. Online reprint available at http://www.interpares.org/book/interpares_book_j_app01.pdf.

Digital records are composed of one or more objects, which InterPARES 1 researchers named "digital components."[96] When "stored" (i.e., digitally encoded and placed in a storage system on digital media), these digital components necessarily have a different external form than a digital record in a form that is comprehensible to a person (i.e., as a manifested record). Further analysis in InterPARES 2 disclosed that, if kept and managed as a record, the stored digital component(s) of a manifested digital record collectively constitute a stored digital record. Thus, InterPARES 2 distinguished two sub-classes of digital records: "stored digital record," the encoding of a digital record within a system;[97] and "manifested digital record," a stored digital record that is visualized in a form suitable for presentation either to a person (i.e., in human-readable form) or to a computer system (i.e., in machine language). In practical terms, a stored digital record is what is kept, while a manifested digital record output from the stored digital record on a display screen or other output device is a copy that is reproduced on demand for the purpose of communicating information to persons or computer systems, rather than as a record that is kept. Thus, in such instances, a stored digital record is qualified as a record because it is intended to be used, and is capable of being used, to reproduce a manifested digital record.[98]

Analysis of the InterPARES 2 case studies showed that there is not necessarily a manifested digital record corresponding to a stored digital record. A stored digital record (or any of its digital components) may be used with—or, as in the case of enabling records, may even control—user inputs and interactions and variable data from other sources to produce changing manifestations that cannot be records precisely because their content and/or form is not fixed. Thus, a stored digital record is qualified as a record because it satisfies the definition of a record, regardless of whether it participates in the production of other records or provides a means for reproducing a manifested digital record.

InterPARES 2 refined the InterPARES 1 definition of "digital component" to one that is simpler, less ambiguous and easier to apply: "An aggregation of digitally encoded data composed of one or more bitstreams, including any metadata necessary to order, structure or manifest its content and form, requiring a given preservation action."[99] There are three types of data that comprise stored digital documents and records: (1) content data, which are about the acts, facts or data that the document or record conveys (i.e., its content); (2) form data, which enable the system to reproduce the manifested digital document or record in the correct form from the stored digital document or record; and (3) composition data, which identify the elements of stored form and content data belonging to the document or record and map them to the different elements of structure defined by the form data. A digital component may comprise

[96] Defined in the InterPARES 1 Glossary as "A digital object that is part of an electronic record, or of a reproduced electronic record, or that contains one or more electronic records, or reproduced electronic records, and that requires specific methods for preservation" (http://www.interpares.org/book/interpares_book_q_gloss.pdf).

[97] Although the literal InterPARES 2 definition of stored digital record is "A stored digital document that is treated and managed as a record," when taking into account the fact that a digital document is composed of a digital object, the effective definition becomes "A digital object, placed in a storage system on a digital medium, that is managed as a record, and which includes information about the properties of the object and may also include methods of performing operations on or with the object." Although perhaps of more immediate relevance to the task of preservation than appraisal, it is critical to recognize that a stored digital record may not correspond to a single physical file stored on a digital medium, but rather to an amalgamation of discrete digital component(s), or subsets of components, that are logically linked to one another at the file system level through the digitally encoded information contained in a file header block, and at the application level through digitally encoded content data, form data, composition data and related rules. Thus, the digital component(s) of a stored digital record may be a subset of a stored file, a set of stored files, or various elements of data extracted from different files.

[98] In most but not all cases, enabling the reproduction of the manifested digital record is, in fact, the primary purpose of keeping the stored digital record, while the manifested record is reproduced to communicate information to persons or other systems (Duranti and Thibodeau, "The Concept of Record," op. cit., 51).

[99] InterPARES 2 Terminology Database. Available at http://www.interpares.org/ip2/ip2_terminology_db.cfm.

one or more of these types of data. Another important type of digital component is the digitally encoded rules that govern the reproduction of the content and form of the manifested digital document or record by determining its extrinsic elements of form, as well as any allowable content and form variations.

InterPARES 2 analysis also proposed an addition to the established typology of records (dispositive, probative, supporting or narrative). The new type, *prospective record*, does not relate to a past action or state of affairs, but determines the form and/or content of records, actions or states subsequently created. Prospective records are subdivided into either *instructive* or *enabling* records. An instructive record is one that provides instructions, intended to be read by humans, on the creation of records in the course of some activity. It can exist either in hard copy or as a manifested digital record. Conversely, an enabling record is one that is used, in the digital form in which it is stored, by a computer system in performing an action, interaction or process and that cannot achieve its purpose if transformed into human-readable format. Thus, to be considered records, enabling records must be "properly maintained and managed [in their stored digital form] as intellectually interrelated parts of records aggregations."[100]

In this light, the findings of InterPARES 1 need to be clarified. As has already been discussed, the digital component, or set of digital components, used to reproduce an authentic copy of a human-readable record may itself be a record; namely, a stored digital record. However, it is important to understand that a stored digital record cannot be used for the purpose for which the manifested digital record is reproduced. It is at least one step removed from that purpose. In fact, many steps may be required to process the digital component(s) to output the manifested digital record in its requisite human-readable form. Thus, one might say that the purpose of the stored digital record is not to achieve the dispositive, supportive or narrative purpose of the (human-readable) record, but to enable production of authentic copies of that record. However, although that view is valid, it is overly narrow. When, for example, a student registers online to take a course, provided the university's registration system is a trustworthy recordkeeping system, the data from the student's registration record that is created from that transaction—which is kept as a stored digital record in the recordkeeping system—can be used, in combination with comparable data extracted from the stored digital registration records of other students, to produce a new record—a course roster—without going through the intermediate step either of reproducing the individual registration records in human-readable form (i.e., the manifested records), or even of reproducing manifested digital records of the individual registrations for use by the computer registration system. All that is needed is to extract the relevant data elements (i.e., course and section enrollment) from the relevant stored digital records (i.e., the registration records for each student) to produce the roster. Other data elements, such as when each student completed the course enrollment or whether the student satisfied prerequisites for the course, need not be used. Thus, the digital encodings of the enrollment records are used in a way that exactly parallels the use of paper records. When used in this manner, the *stored* digital versions of the registration records, as opposed to their *manifested* digital versions, should be recognized as the records that share an archival bond with the course roster record and should be properly managed and appraised as such.

Thus, the case studies revealed the existence of several new types of records, some with analogue equivalents and others that can exist only in a digital environment. These digital records are created within record systems with increasingly complex groups of digital components, multiplying the number of relationships that must be identified, documented,

[100] Duranti and Thibodeau, "The Concept of Record," op. cit., 59.

managed and preserved. There are systems that cannot display their stored digital records in human-readable form. There are also systems that can display the content of a manifested digital record in a number of documentary forms (such as spreadsheet data), although each available form is in fact fixed and stable. There are systems where non-human-readable records will participate in the production of digital content that never becomes fixed and is therefore not a record. And there are systems that appear to produce variable content, although careful analysis confirms that this apparently changing content is delimited by fixed rules, and therefore is, in fact, fixed and stable. Termed "bounded variability" within the computer and information sciences field, the concept describes an environment where "changes to the form and/or content of a digital record ... are limited and controlled by fixed rules [and a stable store of content data, form data and composition data], so that the same query, request or interaction always generates the same result."[101] It is important to emphasize that this is an aspect of digital records, with implications for appraisal, that goes beyond the conclusion reached in InterPARES 1 (and re-affirmed in InterPARES 2) that, although the physical integrity of a record—such as the proper number of bit strings—may be compromised, as long as the record still retains all of its essential attributes such that the message it is meant to communicate to achieve its purpose is unaltered, it can be considered "essentially complete and uncorrupted."[102]

In fact, through identification of the concept of bounded variability, InterPARES 2 has expanded the circumstances under which variations in a record's form and content may be acceptable beyond basic considerations of unintended, accidental or incidental alterations to the record's physical integrity to include consideration of a creator's intentions vis-à-vis the use of variable record presentation elements. With respect to variations in documentary form, it is important to recognize that:

> In many interactive, experiential and dynamic documents, authors or writers[57] intentionally use specific possibilities which digital technology offers for variability in the form in which information is presented. In such cases, the form is 'fixed' in that the design allows certain aspects of form to vary and not others. Documentary forms that include variable elements do not violate the requirements for fixed form, any more than analog audio and motion video recordings, which present temporal variations in sound and imagery. Such variability in presentation intended by the author should be seen as part of the extrinsic elements of the documentary form….With electronic records, then, the 'fixed' form consists of those aspects of form which the author or the writer intended or could control.[103]

Likewise, a creator may invoke this type of intentional "bounded variability" in the content of a record without compromising the record's integrity. For example, as is the case with certain online sales catalogues, interactive digital environments enable a creator to use documentary forms that permit variable selection of content and variable sequencing of that selection, such as in the display of subsets of the content of such catalogues in response to specific user input.[104] As long as the *stored digital record* is controlled by fixed rules such that the same query, request

[101] Excerpted from the definition for "bounded variability" from the InterPARES 2 Terminology Database, available at http://www.interpares.org/ip2/ip2_terminology_db.cfm. For a more in-depth discussion of *bounded variability*, see Duranti and Thibodeau, "The Concept of Record," op. cit., 47–49.
[102] MacNeil et al., "Appraisal Task Force Report," op. cit., 47.
[103] Duranti and Thibodeau, "The Concept of Record," op. cit., 47–48. Note: footnote reference in the quote is from the original text, and is not reproduced here.
[104] Ibid., 49.

or interaction always generates the same documentary form and content selection in the *manifested digital document*,[105] such cases can satisfy the requirement for fixed content.[106]

Another difference that must be taken into account during appraisals of records in interactive, experiential and dynamic digital environments is that the use of technology may change the activities of the records creator. In turn, such changes are likely to cascade into changes in records creation and recordkeeping activities. An example of the relevance of this to appraisal, and, in particular, the need to periodically re-examine a creator's document and records creation activities (and supporting technologies) after the initial appraisal, can be found in case study 24 (VanMap), where, although the creator's existing GIS system is not designed to create records, there are tentative plans to modify its architecture so that it can. This also highlights the notion that, in contrast to the emphasis in InterPARES 1 on monitoring only records that had been appraised for permanent retention, the InterPARES 2 research clearly demonstrates the need to expand the monitoring function to include those data and records that earlier appraisals decided were not worthy of preservation, as well as any systems and/or activities that earlier appraisals determined did not generate records. However, even when the records (or the systems generating them) do not themselves change, putting the records online may give rise to significant changes in their archival bond, as the ease of access often results in records being used in many more activities than previously. Given that archival appraisal should select records for preservation based on knowledge of the creator's processes and of the entire archival fonds of that creator, appraisal in the digital environment should be seen not as a one-time evaluation of a static body of records, but as an activity that needs to span the life of the records and the activities of their creators from creation to transfer to the archives for permanent preservation.[107] This, of course, greatly increases the complexity of the appraisal process, especially in relation to the more typical "one-off" appraisals of traditional analogue records.

A final difference between the process of appraisal envisioned by InterPARES 1 and the findings of InterPARES 2 relates to the stronger differentiation that now exists, in a digital environment, between retrospective records and prospective records, particularly those that play an enabling role in the (re)production of another record. Aside from the greater technical difficulty of preserving this function of reproduction, the preserver must decide whether this type of record fits within the acquisition mandate of its institution or program. Enabling records are not a completely foreign concept within past preservation practices: scores provide the basis for future performances; photographic and moving image negatives produce new prints; digital components interact to produce authentic copies of records. With respect to enabling records,

[105] Defined in the InterPARES 2 Terminology Database as "A digital document that is visualized or rendered from a stored digital component and/or stored digital component(s) in a form suitable for presentation either to a person (i.e., in human-readable form) or to a computer system (i.e., in machine language)."

[106] It is also important to emphasize that in cases such as this, where the manifested digital document appears to be a record, it is actually the stored digital record—which encompasses the sales catalogue's entire "palette of possibilities" (Duranti and Thibodeau, "The Concept of Record," op. cit., 32) vis-à-vis its record form and content, not just the subset presented to the user in response to a query—that is the record that is kept and used for future reference. Duranti and Thibodeau identify these types of manifested digital documents that appear to be records, but which are not actually represented internally by a stored digital record that corresponds exactly to the apparent manifested digital record, as pseudo-documents or "pseudocs" (Ibid., 27).

[107] This process is precisely encapsulated by activity A4.2.4 (Monitor Appraisal Decisions) in the Chain of Preservation model, which the model defines as "To keep track of appraisal decisions in relation to subsequent developments within the creator's and/or preserver's activities that might make it necessary to adjust or redo an appraisal, such as substantial changes to: (1) appraised records and/or their context, (2) the creator's organizational mandate and responsibilities, (3) the creator's record-making or recordkeeping activities or systems, (4) the preserver's records preservation activities or systems and/or (5) the preserver's organizational mandate and responsibilities" (Appendix 14, 56. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_14.pdf).

however, preservation would appear to be a "prospective" commitment to the record, rather than a retrospective one.

Questions 2 and 3

> *How do the preservation concepts, methods and models developed by InterPARES 1 for the administrative and legal records created in databases and document management systems apply to the preservation of the records of artistic, scientific and governmental activities resulting from the use of the technologies examined by InterPARES 2?*
>
> *What preservation paradigms can be applied across activities and technologies? What preservation paradigms are required for specific types of records resulting from each activity?*[108]

As was the case with the appraisal question, the case studies revealed many of the familiar problems that have always challenged the preserver's ability to safeguard records over the long term. Poor records creation practices, poor organization of the records, poor maintenance practices, an absence of long-term planning, instability of the record's physical carrier, poor storage conditions and poor documentation are just a few well-known examples from the analogue era that have been carried forward. Given the increasingly complex and, in many instances, unique nature of the digital records associated with the interactive, experiential and dynamic applications examined in InterPARES 2, it perhaps goes without saying that many of the same preservation problems that have carried over from the analogue realm are, in many instances, far more acute in the digital realm.

The InterPARES 2 research, particularly in regard to the conceptualization of records in interactive, experiential and dynamic electronic systems, has resulted in a number of paradigmatic shifts impacting digital preservation that broaden, rather than contradict, those encountered in the InterPARES 1 research. First, whereas "[o]ne of the most important findings of InterPARES 1 was the recognition and articulation of the difference between the form in which an electronic document is manifested to a person and the form in which it is stored digitally," InterPARES 2 has enriched this distinction to suggest that individual digital components, or aggregations of digital components, might themselves constitute a record or a set of records, depending on how they are instantiated in the system and how they are used by the creator.

As discussed earlier, this finding led InterPARES 2 researchers to distinguish two fundamental sub-classes of digital records; stored digital records and manifested digital records. Although a stored digital record, composed of one or more stored digital components, and the manifested digital record reproduced from those components are related, they are nevertheless distinct and, as noted earlier in the school-registration example, achieve their respective intended purposes in fundamentally different ways: the former is used in its binary code form for "presentation" to other electronic applications or systems, while the latter is translated from its stored binary code form into a form suitable for presentation to a person. This fact has important ramifications for preservation in that the specific preservation requirements for both types of records may also be distinct—informed, in part, by the way that the records are used to achieve

[108] InterPARES 2 Project, Domain 3 Web page. Available at http://www.interpares.org/ip2/ip2_domain3.cfm.

their respective purposes. Indeed, a record, such as an enabling prospective record—or, depending on the circumstances, perhaps also a more "traditional" retrospective record—that need only be "presented" and used in binary code form to achieve its purpose will likely require different preservation measures than either a retrospective or an instructive prospective record that to achieve its purpose must be reproduced in its manifested digital form for presentation to a person.

A second paradigmatic shift of the InterPARES 2 research impacting digital preservation relates, as discussed earlier, to the stronger differentiation that now exists, in a digital environment, between retrospective and prospective records, and particularly those that play an enabling role in the (re)production of another record. Enabling prospective records, in particular, offer unique preservation challenges because of the strict requirement that they be maintained in the systems in which they were created—or in systems with identical functionality—to be able to preserve their ability to produce or enable the interactions, experiences, performances or other processes they were intended to generate. Although the bitstream encoding of the stored digital components of retrospective and instructive prospective records that are kept to reproduce manifested digital records can, in many cases, be converted from one format to another (e.g., .doc to .pdf) without compromising the ability of the manifested digital record to achieve its purpose,[109] the bitstream of an enabling record must be preserved in its original form for the record to achieve its purpose.

A third paradigmatic shift relates to the concept of bounded variability and its impact on our understanding of manifested digital records. In particular, this new concept enriches the concept of the manifested record to encompass any and all types of variability of form and content that are specific to the record (i.e., represent the author's intent) and that are governed by *fixed* rules or instructions.[110]

These realizations lead, in turn, to concerns about whether this more complex and broadened range of records can be preserved according to known preservation strategies within existing preservation paradigms. As discussed earlier in the "Findings" section, the Project's case studies, which focused on records creation and maintenance, offered few concrete examples of preservation scenarios, aside from digitization and "putting it on the Web." The above discussion of variations in the form, or function, or even behaviour of digital records suggests that preservation is already possible within the known parameters of today's preservation strategies. These records clearly include more components, which exist in increasingly complex relationships. The accurate and authentic reproduction of records generated by interactive, experiential and dynamic systems will require increasingly sophisticated metadata (and/or metadata management) to document either the intent(s) of the creator or the limitations to that intent imposed by inadequacies in the available technology.[111] New types of hardware dependencies will have to be overcome. New methods to confirm the successful identification and preservation of digital records will have to be defined, especially in cases where the successful reproduction of content and documentary form in a human-readable format (i.e., manifested digital record) is no longer the goal of the preservation process.

Among other things, these issues will require that systems intended to preserve digital records be capable of the following:

[109] See Duranti and Thibodeau, "The Concept of Record," op. cit., 20.
[110] Ibid., 51.
[111] For a comprehensive analysis of the requirements and real-life context for metadata that relate to the establishment of reliability and authenticity, as well as the long-term preservation and potential re-usability of digital records, see the Description Cross-domain Task Force Report. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_6_description_task_force.pdf.

1. Identifying and locating all the digital components of each record in the system.
2. Managing each digital component based on its specific preservation requirements.
3. Identifying and managing as records those digital components that are used as records, including digital components that are records themselves (e.g., enabling records) and/or are components of one or more records.
4. Identifying and managing both the hardware and the software required for processing all digital components in the formats required for them to fulfil their function(s) as records and/or as components of stored and/or manifested digital records.
5. Applying the appropriate software, and, as required, hardware, to each component to reproduce the manifested digital record from the stored digital record and/or process the stored digital record in cases where the stored digital record is also an enabling record.

Question 4

> *What metadata are necessary to support appraisal and preservation of authentic digital records resulting from each activity?* [112]

Although it was hoped that the answer to this final Domain 3 question would flow naturally from archivists' review of the case studies, it was subsequently realized that this would not occur for two main reasons. First, as noted earlier, the focus of the case studies was on creators, nearly all of whom were neither involved in, nor interested in, appraisal of their records, and many of whom also were not involved in long-term preservation of their records. Second, despite initial hopes, it was not possible, because of the relatively small number of archivists involved in the Project, to appraise the case studies' records, nor, consequently, to determine an appropriate preservation strategy for them. Thus, very few data were collected that could be used by the Domain 3 researchers to address this question. However, the reports of both the Description Cross-domain and the Modeling Cross-domain task forces provide detailed analyses and discussions of the metadata necessary to support appraisal and preservation of authentic digital records. In fact, this final Domain 3 research question was "adopted" by the Modeling Cross-domain and integrated into its work during development of the Chain of Preservation model. Likewise, as stated in the Description Cross-domain Task Force Report, the premise underlying its work "is that detailed trustworthy metadata are key to ensuring the creation of reliable, and preservation of authentic, records and other entities in electronic systems…. [which] argues for an end-to-end metadata management regime that addresses which metadata need to be created and/or carried forward in time, for what purposes, by whom, and how they are to be preserved and validated."[113] To this end, the Description Cross-domain, together with the Modeling Cross-domain, developed a *metadata specification model*. The metadata specification model, which aligns closely with the Reference Model for an Open Archival Information System (OAIS),[114] is intended to be used, in conjunction with the Chain of Preservation model, to provide the basis for developing specifications for automated tools that can be used to systematically assist with the creation, capture, management and preservation of essential metadata for active and preserved

[112] InterPARES 2 Project, Domain 3 Web page, op. cit.
[113] Description Cross-domain Task Force Report, op. cit., 30.
[114] The OAIS model describes the key roles, responsibilities and functions of a digital repository, including Ingest, "Archival" Storage, Data Management, Preservation Planning, Administration and Access. See http://public.ccsds.org/publications/archive/650x0b1.pdf.

records as well as with the identification of which metadata need to be manually created and also which can be summarized and discarded at certain points.

Toward Guidelines for Preserving Authentic Digital Records

Much of the InterPARES 2 research revealed the critical absence of tools necessary to ensure the preservation of digital records. This finding was repeated across the arts, science and government focuses and affected individual creators, small groups, funded projects, government departments and large corporations. Each InterPARES research unit—the focuses, domains and cross-domains—identified missing policies, strategies, principles and guidelines to assist records creators and preservers during the ongoing transition from analogue to a digital recordkeeping environment. It fell to Domain 3 to produce the *Preserver Guidelines*,[115] a document designed to accompany the *Principles for Records Preservers*[116] developed by the Policy Cross-domain. Because the case studies focused on records creators, little was found in the way of preservation "best practices" being undertaken by the participants. Instead, the *Preserver Guidelines* reflect two perspectives:

- Actions that would have to be undertaken to avoid some of the situations encountered in the more problematic case studies
- Actions that would have to be undertaken to address the appraisal and preservation concerns identified in the InterPARES 2 research

[115] See Appendix 21.
[116] See the *Policy Framework* in Appendix 19. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_19.pdf.

Appendix 13

Domain 3 Template for Case Study Analysis

Domain 3
Case Study Analysis

Case Study
Title
Organization

1. What types of entities does the diplomatic analysis identify in this case study? (i.e. records, publications, data, etc.)[1]
    1a. If there are no records, should there be records? If not, why not?
    1b. If there should be records, what kinds of records should be created to satisfy the creator's needs (as defined by an archivist)?
    1c. What characteristics of records (as defined by an archivist)[2] are missing yet necessary to preserve these entities?
        - completed as part of an action
        - involving a communication among 3 juridical or physical persons (e.g. author, writer, addressee), or over time
        - a fixed documentary form
        - a stable content
        - an archival bond with other records either inside or outside the system
        - an identifiable context
2. Are the entities reliable? If not, why not? (Give evidence from both the diplomatic analysis and the case study report.)
3. Are the entities accurate? If not, why not? (Give evidence from both the diplomatic analysis and the case study report.)
4. To what degree can the entities be presumed to be authentic, and why?[3] (The answer to this question requires providing the evidence for all benchmark requirements that have been fulfilled and also reaching a cumulative presumption of authenticity. The higher the number of satisfied requirements, and the greater the degree to which requirement is satisfied, the stronger the presumption of authenticity.)

    Benchmark Requirements Supporting the Production of Authentic Copies of Electronic Records (these apply to the creator):
        1. Capture of identity and integrity metadata
        2. Enforcement of access privileges

[1] If multiple entity types are identified, answer questions 1a onward for each type of entity selected for analysis. See Appendix [1a], section 1 for a discussion of General vs. Special Diplomatics and section 2 for a number of definitions relevant to Question 1.
[2] See Appendix [1a], section 3 for a more complete definition of the characteristics of a record.
[3] See Appendix [1b] for a more extensive discussion of the elements of the Benchmark and Baseline Requirements Supporting the Presumption of Authenticity of Electronic Records, from InterPARES 1.

3. Protection against loss and corruption
4. Protection against media and technology obsolescence
5. Established documentary forms
6. Ability to authenticate records
7. Procedures in place to identify the authoritative record
8. Procedures in place to properly document removal and transfer of records from the creator's originating system

Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records (these apply to the preserver):

1. Controls over records transfer, maintenance, and reproduction
2. Documentation of reproduction process and its effects
3. Archival description

5. For what purpose(s) are the entities to be preserved?[4]
6. Has the feasibility of preservation been explored?
   6a. If yes, what elements and components need to be preserved?
7. Which preservation strategies[5] might most usefully be applied, and what are their strengths and weaknesses, including costs and degree of technical difficulty?
   7a. Which alternative preservation strategies[6] might be applied? What are their strengths and weaknesses, including costs and degree of technical difficulty?
8. What additional information does the preserver need to know to facilitate appraisal and preservation?
   8a. If required information is missing, where should it come from and how should it be made manifest?
9. Are there any policies in place that affect preservation?
   9a. Are there any policies in place that present obstacles to preservation?
   9b. Are there any policies that would need to be put in place to facilitate appraisal and preservation?

[4] If multiple purposes are identified, answer questions #3 onwards for each purpose.
[5] For a list of the many existing preservation strategies, see the *Domain 3 Preservation Strategies* document. Draft 3, dated 2006-02-22 is located at ip2(d3)_preservation_strategies.pdf. Confirm that you have the most up-to-date version of the document. Identify any of the maintenance strategies from Section A that are not undertaken or are prevented by the existing recordkeeping system. Then apply any one or combination of more than one strategy from Section B that might be applied by the preserver.
[6] For a list of the many existing preservation strategies, see the *Domain 3 Preservation Strategies* document. Draft 3, dated 2006-02-22 is located at ip2(d3)_preservation_strategies.pdf. Confirm that you have the most up-to-date version of the document. Identify any of the maintenance strategies from Section A that are not undertaken or are prevented by the existing recordkeeping system. Then apply any one or combination of more than one strategy from Section B that might be applied by the preserver.

Appendix [Definitions]

1. General vs. Special Diplomatics

The limitations of the diplomatic model of a record as it is elaborated in the *Template for Analysis* are attributable mainly to the fact that the model was built on the premises of general diplomatics. *General diplomatics* seeks to decontextualize records, to eliminate their particularities, variations and anomalies in the interest of identifying the common, shared elements of records that cut across juridical, provenancial, and technological boundaries. Given the complexity and variety of electronic systems, it might make more sense to adopt and adapt the approach of *special diplomatics*, which, traditionally, has focused on the records of individual chanceries and specific juridical systems. In such an approach, one would begin with an analysis of the various features of the systems themselves and the broader record-keeping environment in their own terms, with all their particularities, variations, and anomalies; and, on the basis of that analysis, begin to build a more general framework.[7]

2. Current IP2 Definitions

Record[8]
[Archival Science] - *n.,* A document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference.

Reliable record[9]
[Archival Science] - *n.,* A record capable of standing for the facts to which it attests.

Reliability refers to the trustworthiness of a record as a statement of fact. It exists when a record can stand for the fact it is about, and is established by examining the completeness of the record's form and the amount of control exercised on the process of its creation. The records forms generated using new information technologies make increasingly difficult to determine when a record is complete and whether the controls established on its creation are either sufficient or effective for anyone to be able to assume its reliability.[10]

Accuracy refers to the truthfulness of the content of the record and can only be established through content analysis. With administrative and legal records, it is usually inferred on the basis of the degree of the records' reliability and is only verified when such degree is very low. The volatility of the digital medium, the ease of change, editing, and the difficulty of version control, all make it harder to presume accuracy on the traditional [basis].[11]

[7] From the Final Report of the Authenticity Task Force, p. 24, at http://www.interpares.org/display_file.cfm?doc=ip1_atf_report.pdf
[8] From the InterPARES Glossary, at http://www.interpares.org/ip2/ip2_terminology_db.cfm.
[9] Ibid.
[10] From the detailed proposal, p. 12, at http://www.interpares.org/display_file.cfm?doc=ip2_detailed_proposal.pdf.
[11] Ibid.

Authenticity refers to the trustworthiness of a record as a record. An authentic record is one that is what it purports to be and has not been tampered with or otherwise corrupted. Authenticity is established by assessing the identity and the integrity of the record.[12]

3. Characteristics of a Record:[13]

- a **fixed documentary form**, which means that:

  - the binary content of the record, including indicators of its documentary form, are stored in a manner that ensures it remains complete and unaltered

  - technology has been maintained and procedures defined and enforced to ensure that the content is presented or rendered with the same documentary form it had when it was set aside

- a **stable content**

- an **archival bond** with other records either inside or outside the system

- an **identifiable context**, which means that it participates in or supports an action, either procedurally or as part of the decision-making process (meaning its creation may be mandatory or discretionary), and at least three persons (author, writer, and addressee) are involved in its creation (although these three conceptual persons may in fact be only one physical or juridical person).

Appendix [Benchmark and Baseline Requirements]

[Note: This section has been omitted. Abridged versions of the InterPARES 1 Benchmark and Baseline Requirements are instead provided in Appendices 22a and 22b, respectively.

[12] Ibid.
[13] From the Final Report of the Authenticity Task Force, InterPARES 1, p. 6, at http://www.interpares.org/display_file.cfm?doc=ip1_atf_report.pdf.

Appendix 21

PRESERVER GUIDELINES
Preserving Digital Records: Guidelines for Organizations[1]

Introduction

These guidelines have been developed to provide concrete advice to various groups that are responsible for the long-term preservation of digital records. They are not intended to be comprehensive but to highlight a number of areas that are particularly important to the preservation of authentic digital records and which experience has shown to be often overlooked in the rush to accept digital records into archival repositories.

As is widely recognized, digital records must be carefully managed throughout their entire existence to ensure that they are accessible and readable over time with their form, content and relationships intact to the extent necessary for their continuing trustworthiness as records. It is also widely recognized that management of digital records must proceed from a comprehensive understanding of all phases or stages of records' existence, from the time they are generated, through their maintenance by their creator, and during their appraisal, disposition and long-term preservation as authentic memorials of the actions and matters of which they are a part. From the perspective of long-term preservation, all the activities to manage records throughout their existence are linked, as in a chain, and interdependent. If a link in the chain fails, the chain cannot do its job. If certain activities and actions are not undertaken on records, their integrity (that is, their reliability and authenticity) and preservation are imperilled.

These guidelines focus on the preservation link in the chain of preservation and are organized according to the sequence of preservation activities presented in the InterPARES Chain of Preservation (COP) model,[2] which charts the many sequential steps in the creation, maintenance and preservation of authentic records. The alphanumeric number in parentheses following each section title in these Guidelines is a cross reference to the applicable preservation activity presented in the COP model.

The guidelines have been tailored to address the preservation needs of organizations or programs whose records must be retained and consulted for long periods and those of archival institutions that take on the responsibility for the long-term preservation of the records of others and for their continuing accessibility to the public they serve. In both these cases, human and financial resources as well as in-house technical expertise are frequently limited.

Institutions, organizations and programs with preservation responsibilities should also consult the *Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records* (a.k.a., Policy Framework)[3] developed by the InterPARES 2 Policy Cross-domain, which complement these Guidelines. Many of the recommendations of these Guidelines may also be applicable to the preservation of digital objects other than records, such as documents, publications or data.

[1] These Guidelines have also been issued in an illustrated booklet form that is freely available at http://www.interpares.org/display_file.cfm?doc=ip2(pub)preserver_guidelines_booklet.pdf.
[2] Available at http://www.interpares.org/ip2/ip2_models.cfm.
[3] Available at http://www.interpares.org/public_documents/ip2(pub)policy_framework_document.pdf.

1. Manage Chain of Preservation

This aspect involves determining framework requirements, and designing, implementing and maintaining a chain of preservation framework. A *Chain of Preservation Framework* includes all the elements of policy, strategy, methodology and so on.

1.1. Establish scope and objectives

Preservers must define the scope and objectives of their digital preservation program. In the arts, for example, they may wish to preserve the recording of the performance(s) of a work, or they may choose to undertake the more complex preservation of the components of a work of art that support its reproduction or re-performance. In the sciences, preservers may wish to preserve only the final report of the results of an experiment, or hold raw data, normalized data and/or aggregated data to document the methodology used and the result obtained, as well as to ensure the availability of the data for future uses. Preservers should also consider who the eventual users of the archives will be. Technically sophisticated users generally require less assistance in accessing even technologically complex digital materials, while the general public might require extremely user-friendly access mechanisms and materials transformed into a few simple, but widely available, formats. The scope of the preservation program will help define which preservation strategies (see Section 4 and Appendix 21c, Section B) a preserver might need to support.

In defining the digital preservation program, preservers should build on previous efforts. To develop appropriate policies and strategies, preservers should consult the InterPARES 2 Policy Framework for guidance applicable at organizational, sectoral, national, international and supranational levels. For the functions of the preservation program, preservers should consult the ISO Open Archival Information System (OAIS) standard[4] and should follow the InterPARES 2 Chain of Preservation model for an adaptation of the OAIS standard specifically intended for digital records. Plans should also reflect the *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, a revised and expanded version of the *Audit Checklist for Certifying Digital Repositories* originally developed by the NARA/RLG Digital Repository Task Force.[5]

1.2. Acquire resources

Digital preservation requires substantial resources in funding, technological capabilities and expertise. An organization responsible for digital preservation has several options, including: a) acquire new resources, b) reallocate existing resources and/or c) leverage other resources.

Regardless of the option(s) chosen, a fundamental requirement is that resources must be sustainable. One-time resources, such as grants, may be appropriate for specific finite tasks, such as establishing the preservation program or processing a given body of records, but a reliable source of sustained resources is a *sine qua non* for any preservation program.

Acquiring new financial resources will require a sound plan for the program and a matching communications plan to convince funding sources and stakeholders that preservers are likely to consult that the program should be funded. A viable strategy for a new program may be to start

[4] International Organization for Standardization, ISO 14721: 2003 - Space data and information transfer systems—Open archival information system—Reference model.

[5] See Online Computer Library Center, Center for Research Libraries (2007), "Trustworthy Repositories Audit & Certification: Criteria and Checklist," v. 1.0, February 2007. Available at http://www.crl.edu/PDF/trac.pdf.

small and plan on short-term successes to convince funding sources to incrementally increase resources for the program. An incremental strategy should evaluate whether funding sources are more likely to be influenced by short-term success in basic program objectives or in areas of more particular concern to the funding sources and stakeholders. For example, funders and stakeholders may be more swayed by demonstrations of technological capabilities than by a sound and comprehensive plan for appraising digital records.

For most organizations, reallocating resources to digital preservation is likely to entail painful choices. As with seeking new funds, an incremental approach may be best. Furthermore, ongoing adjustments can be made to the plan, based on the experience gained during each phase of implementation. If the digital preservation program is to be established in a larger institution, it would be helpful to address digital preservation as part of the overall strategic plan rather than as a special initiative.

Even when a preserver successfully acquires new resources or is able to reallocate existing resources to digital preservation, it is unlikely it will have sufficient resources to address all the challenges. Therefore, preservers should capitalize on opportunities for leveraging outside resources. There are a variety of paths for doing this. For example, rather than trying to hire technical experts on a permanent basis or training staff in all requisite technical knowledge and skills, preservers might engage outside experts on a consultative or task basis. They should not exclude options to contract for both basic and ad hoc services. On a basic level, preservers should evaluate the possibility of using a computer service provider rather than acquiring a dedicated preservation system. Ad hoc options include engaging specialized companies for tasks such as re-copying from obsolete digital media or converting rare formats. Another option is to participate—actively or passively—in open-source communities developing technologies needed for digital preservation (e.g., FEDORA,[6] Global Registry of Digital Formats[7]).

Finally, preservers in an organization lacking the required resources to support a digital preservation program should investigate the possibility of establishing collaborative partnerships or consortia to develop and finance a program that meets a minimum acceptable standard.

1.3. Focus on digital records

Preservers must ensure that digital preservation resources are primarily deployed to protect authoritative copies[8] of digital records, rather than to preserve digitized copies of surviving analogue records. The rationale for this is that most analogue records will survive without digitization, whereas digital records will be lost without a digital preservation program.

1.4. Offer advice

Because the chain of preservation of digital records begins at creation, preservers should provide advice on digital records creation and maintenance. Depending on the mandate of the preserver, this may be quite specifically targeted to, for example, employees in the preserver's organization or, as in the case of national archives, other government institutions. In other cases, the advice may be disseminated widely to special interest groups or to the general public, with

[6] See http://hul.harvard.edu/formatregistry/.
[7] See http://www.fedora.info/.
[8] Authoritative copy is defined as "The instantiation of a record that is considered by the creator to be its official record and is usually subject to procedural controls that are not required for other instantiations" (InterPARES 2 Terminology Database. Available at http://www.interpares.org/ip2/ip2_terminology_db.cfm).

the purpose of reaching the person(s) or organization(s) whose records fall under the mandate of the preserver.

1.5. Set a good example

Preservers must establish, within their own organization, a record-making and a recordkeeping environment such that their own control records produced in the course of their preservation function will be created and maintained in a way that satisfies the InterPARES 1 *Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records*.[9] Not only is this an essential requirement for any organization undertaking long-term preservation, but the development of this type of in-house environment will also provide:

- hands-on training to archivists in the technologies they are championing to records creators;
- an invaluable "user's eye view" of actual recordkeeping solutions and how they really work in a day-to-day operational environment;
- a testbed where upgrades and innovations can be introduced and evaluated; and
- a working prototype that can be used in demonstrations.

1.6. Develop procedures

Preservers must establish controls over records transfer, maintenance and reproduction, including the procedures and system(s) used to transfer records to their own organization or program within the organization; maintain them; and reproduce them in a way that satisfies the InterPARES 1 *Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records*.[10] These procedures must embody adequate and effective controls to guarantee the records' identity[11] and integrity,[12] and specifically that:

- unbroken custody of the records is maintained;
- security and control procedures are implemented and monitored;
- the content of the records and the required annotations and elements of documentary form remain unchanged after reproduction.

1.7. Implement maintenance strategies

Although much attention is paid to the development of complex long-term preservation strategies, such strategies are inapplicable if the records for which they are to be used are not properly maintained and protected in the recordkeeping and/or record preservation systems that contain them. A complete version of the eight primary maintenance strategies is available in Appendix 21c, Section A. Briefly, they include:

A1. Clear allocation of responsibilities

[9] See Authenticity Task Force (2002), "Appendix 2: Requirements for Assessing and Maintaining the Authenticity of Electronic Records," in *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, Luciana Duranti, ed. (San Miniato, Italy: Archilab, 2005), 204-219. Online reprint available at http://www.interpares.org/book/interpares_book_k_app02.pdf. See Appendix 21a for an abridged version.

[10] Ibid. See Appendix 21b for an abridged version.

[11] Identity is defined as "The whole of the characteristics of a document or a record that uniquely identify it and distinguish it from any other document or record. With integrity, a component of authenticity" (InterPARES 2 Terminology Database, op. cit.).

[12] Integrity is defined as "The quality of being complete and unaltered in all essential respects. With identity, a component of authenticity" (Ibid.).

A2. Provision of appropriate technical infrastructure
A3. Implementation of a plan for system maintenance, support and replacement
A4. Implementation of a plan for the transfer of records to new storage media on a regular basis
A5. Adherence to appropriate storage and handling conditions for storage media
A6. Redundancy and regular backup of the digital objects
A7. Establishment of system security
A8. Disaster planning

2. Appraise Records for Permanent Preservation (A4.2)

In cases where, as recommended in the InterPARES 2 Chain of Preservation model, retention scheduling is employed, decisions on the disposition of records will regularly be made as part of the management of a recordkeeping system. In some cases, appraisals may be conducted when it is determined that records in a longstanding system need to reach a disposition. Eight important aspects of the appraisal process are discussed below.

2.1. Appraise early

Given the technical difficulties involved in the preservation of digital records, the identification of what records need to be preserved for the long term should be carried out at the earliest possible opportunity. Performing appraisal, establishing transfer methods and even identifying potential preservation strategies with the records creator will improve the likelihood of success. This process may also provide the preserver with an opportunity to offer records creation and maintenance advice (see Section 1.4).

Professional preservers, such as archivists, are frequently encouraged to participate in the actual design of computer applications being developed by organizations with which they have a donor-preserver relationship. This approach will help integrate appropriate recordkeeping and preservation practices. Preservers who have joined system design teams have learned that it is an enormously time-consuming practice that requires a far more detailed understanding of the organization's internal workflows and procedures than an archivist normally acquires during an appraisal. Furthermore, system specifications are rarely an accurate depiction of the application that will eventually be implemented. An appraisal will still have to be conducted once the system is operational and is meeting organizational requirements. It may be more reasonable for archivists to contribute to system design as part of the advice function discussed in Section 1.4. Sharing high level strategies, principles and guidelines developed by the archival profession may prove to be a more realistic goal.[13]

[13] Many aspects relating to the creation of effective digital preservation programs have been studied in recent years. Among the Web sites containing useful information or examples are: the InterPARES Project at http://www.interpares.org; Model Requirements for the Management of Electronic Records (MoReq) at http://www.cornwell.co.uk/edrm/moreq.asp; the Metadata Encoding and Transmission Standard (METS) at http://www.loc.gov/standards/mets/; the Electronic Records from Office Systems (EROS) at the National Archives of the United Kingdom at http://www.nationalarchives.gov.uk/electronicrecords/advice/guidelines.htm; and the Australian DIRKS (Designing and Implementing Recordkeeping Systems) manual at http://www.records.nsw.gov.au/recordkeeping/dirks-manual_4226.asp.

2.2. Locate multiple owners

In cases where the intellectual components of a digital object have multiple owners, these owners must be identified during the appraisal process to assess the ramifications of this situation for long-term preservation. This can occur, for example, where institutions at various levels of government contribute, and share access to, data resources. Another example is illustrated by Web sites that access and use resources located outside their span of control. Although access agreements are frequently negotiated in these circumstances, they rarely include provisions for long-term preservation of all significant digital components.

2.3. Assess authenticity

The assessment of authenticity has always formed part of the traditional archival appraisal process. In the first instance, it has relied on confirming the existence of an unbroken chain of custody from the time of the records' creation to their transfer to the archival entity responsible for their long-term preservation. Periods when records were not subject to some form of protective measures by the records creator or by a successor institution with a vested interest in maintaining the accuracy and completeness of the records can cast significant doubt on the authenticity of the records.

The assessment of authenticity has also depended on the archivist's knowledge of recordkeeping practices, both historically and in relation to the record types and administrative procedures of a specific creator. The general framework for this assessment was originally codified in diplomatics.[14] A third, less frequently used method to confirm the identity and integrity of records is based on comparison. Records within a fonds are compared to copies forwarded to and held by external sources in the normal course of the creator's business.

Records created and maintained using digital technology present additional difficulties, and archivists have not yet developed standard practices to assess authenticity in this environment. Issues revolve around the fact that digital objects are easily duplicated, distributed, re-named, re-formatted or converted, as well as to the ease with which they can be falsified without leaving a trace. The following examples illustrate the extent of the loss to archivists, historians, lawyers and others who require authentic records in their work:

- The physical support on which digital documents are stored has largely lost its significance in confirming the date of a record or its place of manufacture. Anyone with access to functioning, obsolete equipment and storage media has the capability to copy digital files to, for example, 9-track tape or 5-1/4" diskettes.
- The date stamp on any digital file can be modified by adjusting the system clock.
- Few institutions understood what their employees would do once entrusted with word processing software. Standard document forms, such as memos and correspondence on letterhead, disappeared under the onslaught of new, individualized record forms, which rapidly included personalized colour, graphics and even sound effects, as well as the attribution of new meaning to capitalization, colour and the development of emoticons. The degree of erosion of standard records creation practices varied enormously across types and sizes of corporate and government organizations.

[14] See discussion of diplomatics in Luciana Duranti and Kenneth Thibodeau (2006), "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science* 6(1): 15-21.

- The introduction of e-mail networks allowed records to travel by many new routes among staff, rather than according to the well-established distribution routes of traditional office procedures.
- The severe reductions in records management personnel in most organizations, fuelled by an assumption that digital objects somehow did not need to be managed, played havoc with the holdings of the Records Office, which largely stopped receiving records created and transmitted in digital form.

When appraising records created in a digital environment, the assessment of the authenticity of records must become a more overt, visible process performed and documented by the preserver. Unbroken chain of custody, knowledge of recordkeeping practices, and verification may still offer some assurances of authenticity. To these must now be added the verification of compliance with each of the benchmark requirements for authenticity listed in Section 2.4.

2.4. Document the assessment of authenticity

The appraisal report should document the controls put in place by the creator to guarantee the identity and integrity of the records and thus the presumption of their authenticity. These controls include each of the benchmark requirements supporting the presumption of authenticity.[15] Briefly, these include:

A.1 Expression of Record Attributes and Linkage to Record (e.g., identity and integrity metadata)
A.2 Access Privileges
A.3 Protective Procedures against Loss and Corruption of Records
A.4 Protective Procedures against Media Deterioration and Technological Change
A.5 Establishment of Documentary Forms
A.6 Authentication of Records
A.7 Identification of Authoritative Record
A.8 Removal and Transfer of Relevant Documentation

2.5. Monitor records identified for long-term preservation

Once the appraisal is completed, the records identified for preservation must be monitored at regular intervals until such time as they will be transferred to the preserver. Monitoring involves confirming with the records creator that nothing has changed with regard to how classes of records identified for transfer are being created or maintained or, if changes have occurred, that they have not affected the nature and attributes of the records, their value, their authenticity or the feasibility of their preservation.

Many changes within an organization can affect the ongoing survival of digital records. The possibility that records will be destroyed in an instant is much higher than for traditional records. This danger is somewhat offset by the tendency to duplicate material in an uncontrolled fashion. Unfortunately, if the production of copies is uncontrolled, it is unlikely that anyone will realize when the last copy of a record is destroyed.

The simplest scenario may involve a system upgrade either to the hardware or to the software, which will affect the archives' ability to accept the records. An upgrade could also

[15] See Appendix 21a.

result in even minor system re-design that could remove the ability to separate temporary records from those that must be removed for transfer to the preserver.

A second scenario can involve changes in an organization's mandate or functions. This can easily lead to changes in how computer applications are used, and the nature and amount of data that they contain. People responsible for system re-design may not be aware of the requirement for transfer of the existing records to the designated preserver before the system can be modified. Without intervention, even documentation about the original application and backup tapes will move inexorably toward a scheduled destruction date.

Finally, the widespread collapse of proper records management practices in most organizations means that records are poorly identified and incorrectly stored in unsecured locations. Managers, and even records managers, may not understand the details of the technical infrastructure, while IT staff may be unfamiliar with either the history of an organization or the relative importance of older records in various data stores. Hard drives may be wiped, user accounts and all the files they contain may be deleted, tapes and discs may be recycled or destroyed, and obsolete playback technology may be disposed of to meet day-to-day operational requirements of speed and efficiency, with no understanding of the impact of such actions on an organization's records or on pre-existing transfer agreements designed to ensure their long-term preservation.

2.6. Update appraisals

Appraisals also need to be updated at regular intervals, though less frequently than records identified for transfer need to be monitored. Information gathered during a monitoring visit may provide the first indication that a new appraisal is required. Change within organizations and within their record-making and recordkeeping systems is inevitable. Organizational mandates and responsibilities may change, as well as the way those responsibilities are carried out, and data accumulated in existing systems may be put to new uses, which might increase their long-term value. At the simplest level, systems that did not initially contain records may be upgraded to do so. This is particularly true during this period of "hybrid" recordkeeping systems, where paper-based record systems continue to co-exist with the early stages of digital information, document or record systems.

2.7. Identify all digital components[16]

Paper records kept in traditional recordkeeping systems generally offer a tightly-wrapped package, where the content of the record is firmly attached to its paper support and the record itself is contextually filed with the related records. This seamless system began to break down with the introduction of technology when, for example, photographic negatives had to be processed to produce prints and moving images resulted from multiple layers of sound and images, combined and re-combined to produce the final composite print that is screened in cinemas or broadcast on television.

Digital technology has further dismantled the record into a series of components. To successfully extract digital records from the system in which they were created, or even from a secondary maintenance system, the preserver must ensure that all essential digital components

[16] A digital component is defined as "A discrete aggregation of one or more bitstreams and the metadata necessary to order, structure or manifest its content and form, requiring a given preservation action" (InterPARES 2 Terminology Database, op. cit.).

are identified and that implicit relationships are made explicit in the metadata before the whole construct is transferred. One of the most common examples of a digital component is the library of fonts, any number of which can be selected by the creator to be used in the presentation of a word-processed document. In Windows, these are stored in '.dll' (or dynamic link library) files. For the preserver to be able to reproduce this record to reflect the creator's original intentions, both the digital component containing the text and the digital component containing the font must have been preserved, as well as the link between them established in such a way that the software attempting to display the content of the text file can find the appropriate font library.[17]

2.8. Determine the feasibility of preservation

Although not part of the assessment of the value of the records, the appraisal process must be completed by a careful investigation of the technical preservation requirements for preservation. Different preservation strategies (see Appendix 21c, Section B) can vary widely in cost and can produce very different results. A textual record stripped of all its formatting may be acceptable in a situation where the preserver is interested in carrying forward only the content of the record. However, where meaning is conveyed by the documentary form and the display characteristics of the record, a more complex preservation solution will be required.

A determination of the feasibility of preservation is essential if the preserving body is to clearly understand the cost of the acquisition and preservation to which it is committing. This is not a new activity; it is simply the extension to the digital realm of the identification of the resources needed to preserve, for example, paper records that are mouldy or moving image reels that are badly shrunken. The current state of digital preservation does mean, however, that preservation costs must be viewed as recurrent. Re-copying holdings from one physical carrier to another will be required as often as the selected format becomes obsolete. Conversion of file formats will be required when logical obsolescence threatens to make the content unreadable. In addition, the digital records considered for long-term preservation may require measures far too complex for the technological environment and the knowledge resources of the preserving organization, and this might imply a postponement of the transfer.

3. Acquire Selected Records for Permanent Preservation (A4.3)

The activity of the preserver acquiring selected records, and all the activities of preservation that follow from that, have as their goal the continued authenticity and accessibility of those records that are selected for continuing preservation. This movement of records from the creator's (or legitimate successor's) custody to the preserver's custody is a critical juncture in the chain of preservation and must be done with great care to ensure that nothing goes awry in the transfer process.

3.1. Develop shared plan for transfer

A successful transfer from the current custodian of the records (be it original creator or legitimate successor) to the organization or program taking on responsibility for long-term

[17] A more detailed description of the "digital component," with additional examples illustrating the concept, is available in Preservation Task Force (2001), "Appendix 6: How to Preserve Authentic Electronic Records," in Duranti, *Long-term Preservation*, op. cit., 293–328. Online reprint available at http://www.interpares.org/book/interpares_book_o_app06.pdf.

preservation requires a plan agreed upon by both parties. Re-accessing obsolete systems or extracting inactive records from operational systems will definitely involve human resource costs for copying time and, potentially, for programming time. Special hardware and software may also be required. The logical and physical (or virtual) formats used for the transfer must be agreeable to both parties. As a general rule, the transfer plan should be developed when the technical feasibility of acquisition and preservation are undertaken. If the two parties cannot agree on a transfer process, the appraisal decision may have to be re-visited. Again, in this period of hybrid recordkeeping, paper and microfilm-based options may still exist. Alternatively, the preserver might encourage the records creator to adopt upgrades to the record system that will allow for easier regular transfers.

3.2. Enforce standardized procedures

The controls over the transfer of digital records from the creator's to the preserver's custody must include:

- establishing, implementing, and monitoring procedures for registering the records transfer;
- verifying the authority for transfer;
- examining the records to determine whether they correspond to the records that are designated for transfer; and
- accessioning the records.

As part of the transfer process, the authenticity of the creator's records, which was assessed as part of the appraisal process, should be verified. This includes verifying that the metadata relating to the records' identity and integrity have been transferred together with the related records and are linked to them, and that the records are accompanied by any relevant documentation of the technical and administrative environment in which they were created and maintained.

3.3. Keep the oldest available logical format

The logical format[18] in which the records were originally created, or in which they are held by the creator at the time of transfer, should, whenever feasible, be maintained by the preserver, in addition to any preservation or reference copies generated after the transfer. Should selected preservation strategies, such as a specific conversion path, fail over time, continued custody of the initial logical format will allow the preserver to essentially re-start the preservation process with the most authoritative copy of the records, by applying a different preservation strategy to the records. Over the long periods during which preservers hold records, experience may show that other preservation strategies are more stable over time or can more easily be carried forward over the long-term. Alternately, new methods of preservation may have been developed following the acquisition and initial processing of the records.

[18] Logical format is defined as "The organized arrangement of data on electronic media that ensures file and data control structures are recognizable and recoverable by the host computer operating system" (InterPARES 2 Terminology Database, op. cit.). Two common logical formats for files and directories are ISO 9660 for CD-ROMs, and Universal Disk Format (UDF) for DVDs.

3.4. Avoid duplicates

Because of the ease of replication of digital records, the preserver must put in place procedures to ensure that digital records from a specific series are transferred by a specific creator to the preserver only once. Accurate identity information is an important first step in avoiding duplication of effort by the creator and the preserver. Also, if reference copies are provided by the preserver to the creator after the transfer of the records, they should be clearly identified and marked as such to prevent accidental re-transfer.

3.5. Document all processing

Initial processes applied during and immediately after transfer may or may not be related to preservation per se. Confirming the identity of the transferred material, checking for viruses and confirming completeness of files tend to leave the transferred file unchanged. File conversion, renaming digital objects and encapsulating files are more intrusive activities. In both cases, preservers must document all processing of digital records and the effects of processing while records are in their custody (see Appendix 21b, Requirement B.2). This documentation should include information such as:

- why certain processes were applied to the records;
- what records were processed;
- the date when the process was performed;
- the names of persons performing and documenting the various steps of the process(es);
- the impact of the process performed on the records' form, content, accessibility and use; and
- the description of any damage, loss or other problems encountered as a result of the processing, including any effect on the elements expressing the records' identity and integrity.

Should the preserver produce copies of the acquired records, it is important to remember that, as discussed in Section 1.5, these copies should be produced in an environment that satisfies the relevant requirements[19] from the InterPARES 1 Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records.

4. Preserve Accessioned Records (A4.4)

The designated records preserver is the entity responsible for taking physical and legal custody of, and preserving (i.e., protecting and ensuring continuous access to), a creator's records. Be it an outside organization or an in-house unit, the role of the designated preserver should be that of a trusted custodian[20] for a creator's records. The authentic copies of the creator's records are kept by the trusted custodian in a *trusted preservation system* (see Appendix 21c), which should include in its design a description and a retrieval system. This trusted preservation system must also have in place rules and procedures for the ongoing production of authentic copies as the existing system becomes obsolete and the technology is upgraded.

[19] Requirement A.5 (Establishment of Documentary Forms), where the creator establishes the documentary form of the record, would usually not apply to the preserver, except if the original documentary form of the record has been lost and the preserver must specify a substitute to permit access.

[20] A trusted custodian is defined in the InterPARES 2 Terminology Database as "A preserver who can demonstrate that it has no reason to alter the preserved records or allow others to alter them and is capable of implementing all of the requirements for the authentic preservation of records" (InterPARES 2 Terminology Database, op. cit.).

4.1. Describe the records

The information about the records and their contexts collected during the appraisal and processing stages should form part of the archival description of the fonds or series in which the records belong (see Appendix 21b, Requirement B.3). This should also include information about intellectual property rights or privacy concerns.

The archival description of the fonds or series containing the digital records should include—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the digital records of the creator have undergone since they were first created. The description should also include an overview of the transfer and preservation processes based on the documentation discussed in Section 3.5 and the explanation of the relationships among digital components discussed in Section 2.7.

4.2. Identify legal ramifications of preservation actions

When a preservation strategy is selected, its legal implications should be reviewed. For example, format conversion out of a proprietary environment could involve the preserver in illegal actions. In the United States, the *Digital Millennium Copyright Act* has made it a criminal offence to produce tools that can circumvent copyright protection measures. Internationally, the World Intellectual Property Organization Copyright Treaty (WIPO WCT) contains provisions that include copyright protection for software as well as digital works and that introduce criminal penalties for infringement, which ranges from unauthorized copying of material placed on a Web site to the removal or alteration of rights management controls from digital works. Most software packages also include some type of similar restrictions, which users must agree to during the installation process.

4.3. Confirm the effectiveness of the selected preservation strategy

As discussed in Section 2.8, there are now a number of preservation strategies available. Ideally, the selected preservation strategy should be tested on the records prior to the formal transfer to the preserver, to ensure that it will perform as expected. Realistically, most preserving organizations or programs can only fund this type of testing on an exceptional basis. Just as traditional conservators carefully test proposed treatments before applying them wholesale to analogue records, digital preservers must be constantly alert to the impact that each preservation process may have on the records and ensure that it is the appropriate choice for preserving authentic records. Flaws in application software and variations in the functionality of versions over time can result in unexpected consequences when applied to a new group of records.

Part of this process includes a constant awareness of the need to track the presence and the performance of all digital components. A change in one component may have unexpected results on a second component, or it may affect how the relationship functions between any two essential components of the record or affect these components' ability to interact. A different relationship that could be affected is that which exists among the members of a related group of records, such as a dossier or series, and the presentation of that aggregate in the correct order (e.g., alphabetical, chronological or hierarchical). If the original order has been lost, corrective measures will have to be taken.

4.4. Maintain proper storage

It is a widely accepted archival preservation principle that maintaining an appropriate and consistent storage environment (temperature and relative humidity) for the material being stored is the most cost-effective contribution to the long-term preservation of records. Manufacturers of magnetic or optical storage media generally offer advice on optimum storage conditions. The environment must be monitored constantly and the readings checked on a regular basis. This recommendation is one of the eight mandatory maintenance strategies outlined in Section 1.7 and discussed in Appendix 21c, Section A.

5. Output Records (A4.5)

As noted earlier, continued accessibility (i.e., use) is an integral part of the archival process. Consequently, providing access to preserved records is an essential component in the chain of preservation. It should be managed by the preserver with the same sense of responsibility and degree of technical and professional competence imparted to records appraisal, acquisition/ transfer, description and storage.

5.1. Explain how the reference copies were made

The relationship between the records acquired from the creator and any copies produced by the preserver must be clearly described and readily accessible to users (see Appendix 21b, Requirement B.2.b). This should also include documenting how the reproduction process control measures that are in place were established and implemented and how they are monitored to ensure that the content of the reproduced records is not changed in the course of reproduction. Copies of records in the preserver's preservation system may not be designated authentic if the preserver has made them for purposes other than preservation; for example, a copy from which personal identifiers are removed may be made for access purposes.

Documenting the records reproduction process and its effects is an essential means of demonstrating that the reproduction process is transparent (i.e., free from pretence or deceit). Such transparency is necessary to the effective fulfillment of the preserver's role as a trusted custodian of the records. It also provides users of the records with a critical tool for assessing and interpreting the records by demonstrating the continuing authenticity of the records and by providing a complete history of the records, of which the history of reproduction is an essential part.

5.2. Explain the technical requirements for access

As mentioned in Section 1.1, different preservers provide reference services to different types of users. This will affect the reference formats and mechanisms adopted by the preserving organization or program, with simpler methods required for members of the general public who may not even own a computer or who may own a fairly simple machine with a few standard pieces of software. To meet the needs of these users, the preserver may have to undertake additional processing or create specialized tools to assist the researchers. More technologically adept users, such as statisticians doing data analysis or forensic accountants conducting fraud investigations, are more likely to apply their own software tools to copies of the records.

Conclusion

This document has outlined a series of guidelines for institutions, organizations and programs with preservation responsibilities for digital records that can be presumed to be authentic and accurate while in the custody of the preserver. For individual preservers and small preservation organizations, the burden may seem great, but the alternative—loss of records or the emergence of corrupt and inauthentic records—would be an even greater problem in the long run. Small organizations will benefit by making a clear designation of the individual or individuals responsible for overseeing the preservation of the organization's digital records. Bear in mind, however, that not all recommendations presented in this document need to be implemented in each circumstance; each preserver should be able to select and adopt the measures that address its particular problems in the specific context in which it operates. There may also be cases in which additional measures are necessary because of legislative or regulatory requirements specific to the preserver's administrative or cultural jurisdiction. In such cases, consultation with legal experts may be required. Individuals, offices and small organizations responsible for preservation should not hesitate to contact such experts for advice on any issues relating to the preservation of the digital records in their custody and under their control.

Appendix 21a

Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records[1]

Preamble

The benchmark requirements are the conditions that serve as a basis for the preserver's assessment of the authenticity of the creator's electronic records. Satisfaction of these benchmark requirements will enable the preserver to infer a record's authenticity on the basis of the manner in which the records have been created, handled and maintained by the creator.

Within the benchmark requirements, Requirement A.1 identifies the core information about an electronic record—the immediate context of its creation and the manner in which it has been handled and maintained—that establishes the record's identity and lays a foundation for demonstrating its integrity. Requirements A.2–A.8 identify the kinds of procedural controls over the record's creation, handling and maintenance that support a presumption of the record's integrity.

Benchmark Requirements (Requirement Set A)

To support a presumption of authenticity the preserver must obtain evidence that:

REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record	the value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records.

A.1.a Identity of the record:

A.1.a.i Names of the persons concurring in the formation of the record, that is:

- name of author[2]
- name of writer[3] (if different from the author)
- name of originator[4] (if different from name of author or writer)
- name of addressee[5]

A.1.a.ii Name of action or matter

A.1.a.iii Date(s) of creation and transmission, that is:

- chronological date[6]

[1] Excerpted from: Authenticity Task Force (2002), "Appendix 2: Requirements for Assessing and Maintaining the Authenticity of Electronic Records," in *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, Luciana Duranti, ed. (San Miniato, Italy: Archilab, 2005), 204-219. Online reprint available at http://www.interpares.org/book/interpares_book_k_app02.pdf.

[2] The name of the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.

[3] The name of the physical or juridical person having the authority and capacity to articulate the content of the record.

[4] The name of the physical or juridical person assigned the electronic address in which the record has been generated and/or sent.

[5] The name of the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

- received date[7]
- archival date[8]
- transmission date(s)[9]

A.1.a.iv Expression of archival bond[10] (e.g., classification code, file identifier)

A.1.a.v Indication of attachments

A.1.b Integrity of the record:

A.1.b.i Name of handling office[11]

A.1.b.ii Name of office of primary responsibility[12] (if different from handling office)

A.1.b.iii Indication of types of annotations added to the record[13]

A.1.b.iv Indication of technical modifications;[14]

REQUIREMENT A.2: Access Privileges	the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records;

REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records	the creator has established and effectively implemented procedures to prevent, discover, and correct loss or corruption of records;

REQUIREMENT A.4: Protective Procedures: Media and Technology	the creator has established and effectively implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change;

[6] The date, and possibly the time, of compilation of a record included in the record by the author or the electronic system on the author's behalf.
[7] The date, and possibly the time, when a record is received by the addressee.
[8] The date, and possibly the time, when a record is officially incorporated into the creator's records.
[9] The date and time when a record leaves the space in which it was generated.
[10] The archival bond is the relationship that links each record, incrementally, to the previous and subsequent ones and to all those [that] participate in the same activity. It is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).
[11] The office (or officer) formally competent for carrying out the action to which the record relates or for the matter to which the record pertains.
[12] The office (or officer) given the formal competence for maintaining the authoritative record, that is, the record considered by the creator to be its official record.
[13] Annotations are additions made to a record after it has been completed. Therefore, they are not considered elements of the record's documentary form.
[14] Technical modifications are any changes in the digital components of the record as defined by the Preservation Task Force. Such modifications would include any changes in the way any elements of the record are digitally encoded and changes in the methods (software) applied to reproduce the record from the stored digital components; that is, any changes that might raise questions as to whether the reproduced record is the same as it would have been before the technical modification. The indication of modifications might refer to additional documentation external to the record that explains in more detail the nature of those modifications.

REQUIREMENT A.5: Establishment of Documentary Forms	the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator;

REQUIREMENT A.6: Authentication of Records	if authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication;

REQUIREMENT A.7: Identification of Authoritative Record	if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative;

REQUIREMENT A.8: Removal and Transfer of Relevant Documentation	if there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

Commentary on the Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records

The assessment of the authenticity of the creator's records takes place as part of the appraisal process. That process and the role of the benchmark requirements within it are described in more detail in the "Appraisal Task Force Report." This assessment should be verified when the records are transferred to the preserver's custody.

A.1 Expression of Record Attributes and Linkage to Record

The presumption of a record's authenticity is strengthened by knowledge of certain basic facts about it. The attributes identified in this requirement embody those facts. The requirement that the attributes be expressed explicitly and linked inextricably[15] to the record during its life, and carried forward with it over time and space, reflects the task force's belief that such expression and linkage provide a strong foundation on which to establish a record's identity and demonstrate its integrity. The case studies undertaken as part of the work of the task force revealed very little consistency in the way the attributes that specifically establish the identity of a record are captured and expressed from one electronic system to another. In certain systems, some attributes were explicitly mentioned on the face of the record; in others they could be found in a wide range of metadata linked to the record or they were simply implicit in one or more of the record's contexts. In many cases, certain attributes (e.g., the expression of the archival bond) were not captured at all. The task force's concern is that, in the absence of a precise and explicit statement of the basic facts concerning a record's identity and integrity, it will be necessary for

[15] For the purposes of this requirement, *inextricable* means incapable of being disentangled or untied, and *link* means a connecting structure.

the preserver to acquire enormous, and otherwise unnecessary, quantities of data and documentation simply to establish those facts.

The link between the record and the attributes listed in Requirement A.1 is viewed by the task force as a *conceptual* rather than a *physical* one, and the requirement could be satisfied in different ways, depending on the nature of the electronic system in which the record resides. For example, in electronic records management systems, this requirement is usually met through the creation of a record profile.[16] In other types of systems, the requirement could be fulfilled through a topic map. A topic map expresses the characteristics (i.e., *topics*) of subjects (e.g., records or record attributes) and the relationships between and among them.

When a record is exported from the live system, migrated in a system update, or transferred to the preserver, the attributes should be linked to the record and available to the user. When pulling together the data prior to export, the creator should also ensure that the data captured are the right data. For example, in the case of distribution lists, the creator must ensure that if the recipients specified on "List A" were changed at some point in the active life of records, the accurate "List A: Version 1" is exported with the records associated with the first version, and that the second version is sent forward with those records sent to recipients on "List A: Version 2."

A.2 Access Privileges

Defining access privileges means assigning responsibility for the creation, modification, annotation, relocation, and destruction of records on the basis of competence, which is the authority and capacity to carry out an administrative action. Implementing access privileges means conferring exclusive capability to exercise such responsibility. In electronic systems, access privileges are usually articulated in tables of user profiles. Effective implementation of access privileges involves the monitoring of access through an audit trail that records every interaction that an officer has with each record (with the possible exception of viewing the record). If the access privileges are not embedded within the electronic system but are based on an external security system (such as the exclusive assignment of keys to a location), the effective implementation of access privileges will involve monitoring the security system.

A.3 Protective Procedures: Loss and Corruption of Records

Procedures to protect records against loss or corruption include: prescribing regular back-up copies of records and their attributes; maintaining a system back-up that includes system programs, operating system files, etc.; maintaining an audit trail of additions and changes to records since the last periodic back-up; ensuring that, following any system failure, the back-up and recovery procedures will automatically guarantee that all complete updates (records and any control information such as indexes required to access the records) contained in the audit trail are reflected in the rebuilt files and also guarantee that any incomplete operation is backed up. The capability should be provided to rebuild forward from any back-up copy, using the back-up copy and all subsequent audit trails.

A.4 Protective Procedures: Media and Technology

Procedures to counteract media fragility and technological obsolescence include: planning upgrades to the organization's technology base; ensuring the ability to retrieve, access, and use stored records when components of the electronic system are changed; refreshing the records by

[16] If the attribute values contained in the profile are also expressed independently as entries in a register of all records made or received by the creator, then, in addition to establishing the identity and supporting the inference of the integrity of the record, they would corroborate such identity and strengthen the inference of integrity.

regularly moving them from one storage medium to another; and migrating records from an obsolescent technology to a new technology.

A.5 Establishment of Documentary Forms

The documentary form of a record may be determined in connection to a specific administrative procedure, or in connection to a specific phase(s) within a procedure. The documentary form may be prescribed by business process and work-flow control technology, where each step in an administrative procedure is identified by specific record forms. If a creator customizes a specific application, such as an electronic mail application, to carry certain fields, the customized form becomes, by default, the required documentary form. It is understood that the creator, acting either on the basis of its own needs or the requirements of the juridical system, not an individual officer, establishes the required documentary form(s) of records.

When the creator establishes the documentary form in connection to a procedure, or to specific phases of a procedure, it is understood that this includes the determination of the intrinsic and extrinsic elements of form[17] that will allow for the maintenance of the authenticity of the record. Because, generally speaking, that determination will vary from one form of a record to another, and from one creator to another, it is not possible to predetermine or generalize the relevance of specific intrinsic and extrinsic elements of documentary form in relation to authenticity.

A.6 Authentication of Records

In common usage, to authenticate means to prove or serve to prove the authenticity of something. More specifically, the term implies establishing genuineness by adducing legal or official documents or expert opinion. For the purposes of the benchmark requirements, authentication is understood to be a declaration of a record's authenticity at a specific point in time by a juridical person entrusted with the authority to make such declaration. It takes the form of an authoritative statement (which may be in the form of words or symbols) that is added to or inserted in the record attesting that the record is authentic.[18] The requirement may be met by linking the authentication of specific types of records to business procedures and assigning responsibility to a specific office or officer for authentication.

The authentication of copies differs from the validation of the process of reproduction of the digital components of the records. The latter process occurs every time the records of the creator are moved from one medium to another or migrated from one technology to another.

A.7 Identification of Authoritative Record

An authoritative record is a record that is considered by the creator to be its official record and is usually subject to procedural controls that are not required for other copies. The identification of authoritative records corresponds to the designation of an office of primary responsibility as one of the components of a record retention schedule. The Office of Primary Responsibility is the office given the formal competence for maintaining the authoritative (that is, official) records belonging to a given class within an integrated classification scheme and

[17] The extrinsic and intrinsic elements of form are defined and explained in the InterPARES 1 *Template for Analysis* (see Authenticity Task Force (2000), "Appendix 1: Template for Analysis," in Duranti, *Long-term Preservation*, op. cit., 192–203. Online reprint available at http://www.interpares.org/book/interpares_book_j_app01.pdf).

[18] The meaning of authentication as it is used by the Authenticity Task Force in this report is broader than its meaning in public key infrastructure (PKI) applications. In such applications, authentication is restricted to proving identity and public key ownership over a communication network.

retention schedule. The purpose of designating an office of primary responsibility for each class of record is to reduce duplication and to designate accountability for records.

It is understood that in certain circumstances there may be multiple authoritative copies of records, depending on the purpose for which the record is created.

A.8 Removal and Transfer of Relevant Documentation

This requirement implies that the creator needs to carry forward with the removed records all the information that is necessary to establish the identity and demonstrate the integrity of those records, as well as the information necessary to place the records in their relevant contexts.

Appendix 21b

Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records

Preamble

The baseline requirements outline the minimum conditions necessary to enable the preserver to attest to the authenticity of copies of inactive electronic records.

Baseline Requirements (Requirement Set B)

The preserver should be able to demonstrate that:

REQUIREMENT B.1: Controls over Records Transfer, Maintenance, and Reproduction	the procedures and system(s) used to transfer records to the archival institution or program; maintain them; and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that	
	B.1.a	Unbroken custody of the records is maintained;
	B.1.b	Security and control procedures are implemented and monitored; and
	B.1.c	The content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.

REQUIREMENT B.2: Documentation of Reproduction Process and its Effects	the activity of reproduction has been documented, and this documentation includes	
	B.2.a	The date of the records' reproduction and the name of the responsible person;
	B.2.b	The relationship between the records acquired from the creator and the copies produced by the preserver;
	B.2.c	The impact of the reproduction process on their form, content, accessibility and use; and
	B.2.d	In those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user;

REQUIREMENT B.3: Archival Description	the archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created.

Commentary on the Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records

The establishment and implementation of the baseline requirements take place as part of the function of managing preservation. The preservation function and the role of the baseline requirements within it are described in more detail in the "Preservation Task Force Report."

B.1 Controls over Records Transfer, Maintenance, and Reproduction

The controls over the transfer of electronic records to archival custody include establishing, implementing, and monitoring procedures for registering the records' transfer; verifying the authority for transfer; examining the records to determine whether they correspond to the records that are designated in the terms and conditions governing their transfer; and accessioning the records.

As part of the transfer process, the assessment of the authenticity of the creator's records, which has taken place as part of the appraisal process, should be verified. This includes verifying that the attributes relating to the records' identity and integrity have been carried forward with them (Requirement A.1), along with any relevant documentation (Requirement A.8).

The controls over the maintenance of electronic records once they have been transferred to archival custody are similar to several of the ones enumerated in the benchmark requirements. For example, the preserver should establish access privileges concerning the access, use, and reproduction of records (Requirement A.2); establish procedures to prevent, discover, and correct loss or corruption of records (Requirement A.3), as well as procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change (Requirement A.4). Once established, the privileges and procedures should be effectively implemented and regularly monitored. If authentication of the records is required, the preserver should establish specific rules regarding who is authorized to authenticate them and the means of authentication that will be used (Requirement A.6).

The controls over the reproduction of records include establishing, implementing, and monitoring reproduction procedures that are capable of ensuring that the content of the record is not changed in the course of reproduction.

B.2 Documentation of Reproduction Process and its Effects

Documenting the reproduction process and its effects is an essential means of demonstrating that the reproduction process is transparent (i.e., free from pretence or deceit). Such transparency is necessary to the effective fulfilment of the preserver's role as a trusted custodian of the records. Documenting the reproduction process and its effects is also important for the users of records since the history of reproduction is an essential part of the history of the record itself. Documentation of the process and its effects provides users of the records with a critical tool for assessing and interpreting the records.

B.3 Archival Description

Traditionally it has been a function of archival description to authenticate the records and perpetuate their administrative and documentary relationships. With electronic records, this function becomes critical. Once the records no longer exist except as authentic copies, the archival description is the primary source of information about the history of the record, that is, its various reproductions and the changes to the record that have resulted from them. Although it is true that the documentation of each reproduction of the record copies[1] may be preserved, the archival description summarizes the history of all the reproductions, thereby obviating the need to preserve all the documentation for each and every reproduction. In this respect, the description constitutes a collective attestation of the authenticity of the records and their relationships in the context of the fonds to which the records belong. This is different from a certificate of authenticity, which attests to the authenticity of individual records. The importance of this collective attestation is that it authenticates and perpetuates the relationships between and among records within the same fonds.

[1] Although, technically, every reproduction of a record that follows its acquisition by the preserver is an authentic copy, it is the only record that exists and, therefore, should normally be referred to as "the record" rather than as "the copy."

Appendix 21c

Digital Records Maintenance and Preservation Strategies[1]

This appendix includes a list of preservation strategies largely drawn from the UNESCO *Guidelines for the Preservation of Digital Heritage*,[2] which offers a framework for describing digital records preservation strategies that protect and maintain the accessibility of authentic copies of digital records throughout the chain of preservation.

The complete list of possible strategies adopted by InterPARES 2 is conceptually divided into two broad categories: a) maintenance strategies and b) preservation strategies.

A. Maintenance Strategies

Maintenance strategies[3] are the minimum necessary requirement to protect and maintain accessibility of authentic copies of digital records. There are eight primary maintenance strategies. All are necessary to ensure the records components will exist long enough for preservation strategies to be applied.

A1. Clear allocation of responsibilities

A person or office must be given unambiguous responsibility for managing records storage and protection. This is a technical responsibility that requires a specific skill set, dedicated resources, and an appropriate plan. This strategy can be undertaken by hiring a competent staff member devoted exclusively to this task or by assigning existing staff or an existing office a specific portion of time to carry out the responsibilities.

A2. Provision of the appropriate technical infrastructure

This includes all of the physical and administrative resources that enable the recordkeeping and/or maintenance processes (buildings, computer hardware, computer networks and the auxiliary staff necessary to maintain the same).

A3. System maintenance, support and replacement

The implementation of a plan for maintaining, updating and/or replacing hardware and software.

A4. Transfer of data to new storage media on a regular basis

The implementation of a plan for copying of data from one storage medium to another to avoid the impact of media decay. Such transfers should be undertaken in a systematic manner.

[1] Adapted from: Kevin Glick, "Electronic Records Preservation Strategies," (unpublished report, 2006).

[2] Colin Webb (2003), *Guidelines for the Preservation of Digital Heritage*. Prepared by the National Library of Australia for the Information Society Division, United Nations Educational, Scientific and Cultural Organization, report no. CI-2003/WS/3. Available at http://unesdoc.unesco.org/images/0013/001300/130071e.pdf.

[3] A maintenance strategy is defined as "A coherent set of objectives and methods for protecting (i.e., safeguarding authenticity and ensuring accessibility of) digital components and related information over time while still in active or semi-active use by the creator, and for reproducing the related authentic records and/or record aggregations" (InterPARES 2 Terminology Database, op. cit.).

A5. Adherence to appropriate conditions for storage media

The rate of media decay may be dramatically reduced by adhering to appropriate environmental conditions. For instance, excessive heat, humidity and dust endanger storage media.

A6. Redundancy and geographic location

The duplication of digital objects and the storage of the resulting multiple copies on different physical media protects them against media failure. Storage in different physical locations protects against poor environmental storage conditions, fire, flood, etc., at a particular storage site.

A7. System security

Controls should be implemented to ensure that digital components of records are exposed only to authorized users and/or processes. Such controls should include restricting physical access to places where computers are kept as well as restricting access to the digital records on the computers themselves. The latter can be accomplished through various means, including the use of passwords and/or biometric authentication to log on to the system.

A8. Disaster planning

The strategies listed above are designed to minimize accidental loss of data and maximize media longevity, but even with perfect storage conditions and excellent handling protocols, disasters may still happen. A disaster recovery plan should contain detailed procedures for restoring a damaged system and for guiding the effective recovery of recordkeeping and/ or preservation systems following a disaster.

B. Preservation Strategies

In addition to the maintenance strategies, every records preserver is responsible for establishing a trusted preservation system[4] for expressing one or more preservation strategies.[5] Twelve preservation strategies are listed below, in Section B, divided into four broadly defined groups. It is most likely that, in practice, a preserver will support two or more preservation strategies in addition to the eight maintenance strategies listed above in Section A.

B1. Use of standards

The use of widely available and supported standards increases the likelihood of stability and longer term support. Such standards may either be *de jure*,[6] if they have been formally agreed

[4] A trusted preservation system is defined as "The whole of the rules that control the preservation and use of the records of the creator and provide a circumstantial probability of the authenticity of the records, and the tools and mechanisms used to implement those rules" (Ibid.).

[5] A preservation strategy is defined as "A coherent set of objectives and methods for protecting (i.e., safeguarding authenticity and ensuring accessibility of) digital components and related information of inactive records over time, and for reproducing the related authentic records and/or archival aggregations" (Ibid.).

[6] A *de jure* standard is defined as a "Standard issued by an official standards-setting body, whether national (e.g., ANSI), multi-national (e.g., CEN) or international (e.g., ISO)" (Ibid.). For computer file formats, two recent *de jure* standards are PDF/A (PDF standard for archiving) and ODF (OASIS OpenDocument Format).

upon, or *de facto*,[7] if they have been widely adopted by industry. Standards can apply to many facets of a preservation system, including encoding methods, file formats, physical storage media, etc. Compliance with standards might also simplify the application and/or maximize the effectiveness of later preservation strategies. Standardization may be applied *prospectively*, by limiting the formats in which digital records may be transferred to the preserver; or *retrospectively*, by converting files received in other formats to standard ones.

B1.1. Self-describing formats (persistent object preservation, tagging)

Analysis and tagging of records so that the functions, relationships and structure of specific elements can be described. The re-presentation of content can be liberated from specific software applications and can be achieved using different applications as technology changes.

B1.2. Encapsulation

Binding together a record and the means of providing access to it, normally in a *wrapper* that describes what it is in a way that can be understood by a wide range of technologies (such as an XML document). The wrapper often includes metadata that describe or link to the correct tools.

B1.3. Restricting the range of formats to be managed (normalization)

Storing records in a limited number of formats only.[8] The selection of acceptable formats may continue to include new proprietary formats or new generations of existing proprietary formats, or it may be restricted to non-proprietary formats, to carry standardization one step further. One example of this approach is referred to as *durable encoding*, which recommends encoding records to conform to well-known data processing standards down to the level of encoding bits as ASCII or Unicode UTF-8, and objects as XML.

B1.4. Conversion

Transferring digital records from one hardware or software generation to another. As distinct from *refreshing*, which copies the data stream from one carrier to another, conversion entails transforming the logical form of a digital object so that the conceptual object can continue to be correctly rendered or presented by the new hardware or software. The most commonly proposed conversion method involves permanently transforming one logical format into another in line with technological change, so that all converted objects can be presented with prevailing technology. It is also possible to propose a "conversion on demand" or "conversion at the point of access" model. This approach is discussed below in Section B2.4.

B2. Technology dependence

These strategies continue to rely on the original hardware and/or software without changing the records.

[7] A *de facto* standard is defined as a "Standard not issued by any official standards-setting body, but nevertheless widely used and recognized by its users as a standard" (Ibid.). Well known and widely used computer file formats that are considered *de facto* standards include PDF, TIFF, DOC and ZIP.

[8] For a detailed analysis of current issues and trends in the selection of file, wrapper, tagging and encoding formats, together with recommendations for developing and implementing policies on selecting digital file formats for long-term preservation, see: Evelyn Peters McLellan (2006), "InterPARES 2 Project - General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation." Available in English at http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_english.pdf, and in French at http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_french.pdf.

B2.1. Technology preservation

Maintaining the original software and hardware with which digital records were presented.

B2.2. Reliance on backward compatibility

Trusting the ability of some software to correctly interpret and present digital components of records created with previous versions of the same software. In this strategy, the presentation is limited to a temporary conversion for viewing or for non-archival copying purposes, whereas conversion permanently changes records into the format supported by the current version of the software.

B2.3. Software re-engineering

Transforming software as technologies change. As such, it is similar to the transformation of record formats, discussed in sections B1.4.and B2.2. This may include anything from re-compiling source code for a new platform to re-coding the software from scratch in another programming language.

B2.4. Viewers and conversion at the point of access

The use of software tools or transformation methods that provide temporary accessibility when needed, using the original data stream.

B2.5. Emulation

Using software that makes one technology behave like another. In other words, making future technologies behave like the original environment of a preserved digital record, so that the original record could be presented in its original manifestation from the original, or converted, data streams.

B3. Non-digital approaches

Copying the digital records onto relatively stable analogue media, such as paper or microfilm; shifting the preservation burden to an analogue copy in place of the digital object. This approach destroys any functionality provided by the software, such as manipulability.

B4. Data restoration (digital archaeology)

Recovering records as bits from physical media followed by steps to restore the intelligibility of the recovered records. It is most often employed in the recovery of data from failed, damaged or degraded media, but methods to restore intelligibility have been used to rescue documents in obsolete formats.