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 THREE

AUTHENTICITY, RELIABILITY AND ACCURACY OF DIGITAL RECORDS IN THE ARTISTIC, SCIENTIFIC AND GOVERNMENTAL SECTORS

Domain 2 Task Force Report

[including Appendices 12 and 20]

by

John Roeder, The University of British Columbia
Philip Eppard, University of Albany, State University of New York
William Underwood, Georgia Tech Research Institute
Tracey P. Lauriault, Carleton University

Status: Final (public)

Version: Electronic

Submission Date: February 2007

Publication Date: 2008

Project Unit: Domain 2 Task Force

URL: http://www.interpares.org/display_file.cfm?doc=ip2_book_part_3_domain2_task_force.pdf

How to Cite: John Roeder, Philip Eppard, William Underwood and Tracey P. Lauriault, "Part Three—Authenticity, Reliability and Accuracy of Digital Records in the Artistic, Scientific and Governmental Sectors: Domain 2 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_3_domain2_task_force.pdf>

Table of Contents

Introduction

Background and mandate

The first InterPARES Project (1999-2001) addressed the problems of preserving administrative and legal records generated within databases and document management systems.[1] Such records, although fixed digitally on relatively unstable media, are intended to approximate the physical documents generated in the course of established business procedures in well-understood juridical contexts. Thus, the Project naturally focused on how to preserve their authenticity and reliability—those qualities that make them trustworthy as the representations of actions—during their inevitable rewriting from system to system, from medium to medium and from format to format, when they are susceptible to alteration.

For this investigation, InterPARES 1 drew concepts from contemporary archival diplomatics, a theory of record and record analysis rooted in a European practice that is a source of modern Western business and legal systems. Diplomatics identifies those features of documents that make them records—fixed, reliable, complete representations of transactions. It helps guide preservation, because preserving a record requires preserving all those features that make it a record.

Referring to this theory, the Project's Authenticity Task Force developed two sets of practical guidelines for ensuring the authenticity of digital records over time.[2] Each addresses a different phase in the lifecycle of a record, assuming a common distinction between active records that are maintained by the creator for current and future reference, and inactive records that have been transferred to the custody of an archive for long-term preservation. The *benchmark requirements* set forth a basis for presuming or verifying the authenticity of the creator's digital records, while the *baseline requirements* support the production of authentic copies of digital records after they have been transferred to the preserver's custody. Both sets of requirements define and give a basis for assessing the records' identity and integrity, which must be preserved for the copies to be authentic.[3]

---

[1] See http://www.interpares.org/ip1/ip1_index.cfm.

[2] See Authenticity Task Force, "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. Abridged versions of the benchmark and baseline requirements are provided in Appendices 21a and 21b, respectively. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_21.pdf.

[3] As the InterPARES 1 Preservation Task Force determined, "[e]mpirically, it is not possible to preserve an electronic record: it is only possible to preserve the ability to reproduce the record. That is because it is not possible to store an electronic record in the documentary form in which it is capable of serving as a record. There is inevitably a substantial difference between the digital representation of the record in storage and the form in which it is presented for use." (Kenneth Thibodeau et al., "Part Three – Trusting to Time: Preserving Authentic Records in the Long Term: Preservation Task Force Report," in Duranti, *Long-term Preservation*, ibid., 106. Online reprint available at http://www.interpares.org/book/interpares_book_f_part3.pdf). In other words, only the first instantiation of a digital record, before it is stored, is an original. Once the first instantiation is saved, and thus stored in the system in the form of one or more digital components, the original record ceases to exist. Consequently, all subsequent manifestations of stored records are, *ipso facto*, copies. InterPARES 2 research expanded on this concept by distinguishing between a *stored* digital record—effectively defined as 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—and a *manifested* digital record—effectively defined as a digital record that is visualized or rendered from a stored digital record 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) (see Luciana Duranti and Kenneth Thibodeau (2006), "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science* 6(1): 13–68. Note: A reprint of this article is included in Appendix 2.). In fact, "the primary

By focusing on the theoretical requirements for authenticity, the work of InterPARES 1 clarified obstacles to preservation in a thorough and coherent way, providing a complete framework in which to understand and resolve problems that many organizations have lately experienced. These problems were not simply the familiar results of degraded media and changing software. "For example, it highlighted the extent to which electronic systems are still being designed to manage data rather than records."[4] Case studies found that few systems contained entities satisfying the diplomatic definition of a record. Even systems that did contain records did not retain enough information about identity and integrity; so, by definition, the records could not be preserved authentically. The studies also encountered types of information displays that did not seem to have the fixity that one expects of records; for example, computer-monitor displays that assemble information from various, continuously updated sources. Like records, such displays inform the decisions and actions of organizations, but they are not stored or fixed, which raises the question of whether they could be preserved in any sense.

These findings call to mind problems of information management in activities far removed from business and law. One fifth of the data generated by the 1976 Viking exploration of Mars[5] and the works of nearly half of composers[6] and one-quarter of digital photographers[7] have been lost or threatened by technological obsolescence or inadequate preservation strategies. Challenges have been mounted to the trustworthiness of the records of electronic voting machines.[8] Every user of the Internet is familiar with broken hyperlinks, unplayable media and the challenge of determining whether information is authoritative and true.[9]

The nature of these activities gives hope that archival science can assist in finding solutions to these problems. Although artistic objects and experiences are not records in diplomatic theory (being final products, not by-products, of an activity), our appreciation of them generally requires knowledge of the actions and contexts in which they were produced.[10] Hypothesis-testing activities of science depend on the reproducibility of experiments, which in turn requires understanding exactly how recorded data were gathered and interpreted. And as governments mandate that their services be offered online, citizens will want their transactions mediated by interactive applications to be completely and accurately recorded in a way that allows them to trust the record.

But if these problems can indeed benefit from archival expertise, there are significant impediments to understanding them and to finding solutions through collaborations among

purpose of keeping the stored record is to be able to reproduce the manifest record, while the manifest record is preserved to communicate information to persons or other systems" (ibid., 51).

[4] Heather MacNeil et al., "Part One – Establishing and Maintaining Trust in Electronic Records: Authenticity Task Force Report," in Duranti, *Long-term Preservation*, op. cit., 24. Available at http://www.interpares.org/book/interpares_book_d_part1.pdf.

[5] See Terry Cook (1995), "It's Ten O'Clock, Do You Know Where Your Data Are?" *Technology Review* 98: 48–53; and Ross Harvey (2000), "An Amnesiac Society? Keeping Digital Data for Use in the Future." Paper presented at the LIANZA 2000 Conference, New Zealand, 15-18 October 2000.

[6] Michael Longton (2004), "InterPARES 2 Project - General Study 04 Final Report: Recordkeeping Practices of Composers," 1. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs04_final_report.pdf.

[7] Jessica Bushey and Marta Braun (2006), "InterPARES 2 Project - General Study 07 Final Report: Survey of Recordkeeping Practices of Photographers using Digital Technology," 22. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs07_final_report.pdf.

[8] See, for example, http://www.votetrustusa.org/.

[9] Chip Martel et al. (2001), "A General Model for Authentic Data Publication," 1. Available at http://www.cs.ucdavis.edu/~devanbu/files/model-paper.pdf; and Michael T. Goodrich et al. (2001), "Authenticated Data Structures for Graph and Geometric Searching," Technical report, Center for Geometric Computing, Brown University, 1. Available at http://www.cs.brown.edu/cgc/stms/papers/authDataStr.pdf.

[10] David Davies, *Art as Performance* (Oxford: Blackwell, 2004).

archivists, creators and computer scientists. In the artistic, scientific and governmental sectors, the concepts of authenticity and reliability have diverse meanings, whose relation to those in archival science is not always clear. Also, the structure and function of the digital entities created in art and science do not always resemble those in administrative and legal contexts, so it is not clear how well the requirements established by InterPARES 1—or even the archival concept of authenticity itself—apply. Indeed, one of the most interesting research questions for InterPARES 2 was whether it is possible to satisfy these requirements for records (or other digital entities) in activities further removed from traditional recordkeeping practices. One way to answer this question was suggested by the Authenticity Task Force report,[11] which emphasized how important it is to study documents within the context of the systems in which they are created.

Accordingly, the second phase of InterPARES began in 2002 to develop a theoretical understanding of the records generated in interactive, experiential and dynamic systems, of their process of creation and of the present and potential use of records in the artistic, scientific and governmental sectors. The Project was organized into three domains of research, each tasked to investigate different aspects of the problem. The Domain 2 Task Force investigated the concepts of authenticity, reliability and accuracy as they are understood theoretically and practically in the artistic, scientific and governmental sectors, and it considered how those understandings relate to the InterPARES 1 definitions. To this end, the following three mandates were established for Domain 2 during the Project's February 2003 plenary workshop in Vancouver:

1. to find out how the concepts of reliability, accuracy and authenticity are used by records creators in each of the Project's three focus sectors;
2. to find out which words are used in each focus to signify these concepts; and
3. to find out what, if any, significance the creators in each focus place on these concepts.

Research team

The following is a list of researchers and research assistants who contributed to the work of the Domain 2 Task Force throughout the Project:[12]

Co-chairs:

Philip Eppard	Jan 2002 - Dec 2006
Brent Lee	Jan 2002 - Dec 2005
John Roeder	Jan 2002 - Dec 2006
Bill Underwood	Jan 2002 - Dec 2006

Researchers:

Marta Braun	Ryerson University, Canada—Working Group 2.1
Ann Butler	New York University, USA—Working Group 2.1
Hannelore Dekeyser	Katholieke Universiteit Leuven, Belgium—Working Group 2.3
Philip Eppard	University of Albany, State University of New York, USA—Working Group 2.3

[11] MacNeil et al., "Authenticity Task Force Report," op. cit., 24.

[12] Researcher membership in Domain 2 changed somewhat over the five years of the Project. Among those who were interested in Domain 2 issues but were unable to participate for the full length of the Project are: Margaret Campbell, Nova Scotia Provincial Archives, Canada; Ben Howell-Davis, Davis International Associates, USA; Reagan Moore, San Diego Supercomputer Center, USA; and Xiaowei Qiu, State Archives Administration of China.

Ken Hawkins	National Archives and Records Administrations, USA—Working Group 2.2
Ian Lancashire	University of Toronto, Canada—Working Group 2.1
Brent Lee	University of Windsor, Canada—Working Group 2.1
Michael Murphy	Ryerson University, Canada—Working Group 2.1
Eun G. Park	McGill University, Canada—Working Group 2.2
Richard Pearce-Moses	Arizona State Library—Working Group 2.3
John Roeder	The University of British Columbia—Working Group 2.1
Andrew Rodger	Library and Archives Canada—Working Group 2.1
Bill Underwood	Georgia Tech Research Institute, USA—Working Group 2.3

Research Assistants:

Scott Amort	The University of British Columbia, Canada
Gary Barclay	The University of British Columbia, Canada
Lindsey Bergen	The University of British Columbia, Canada
Natalie Catto	The University of British Columbia, Canada
Heather Dean	The University of British Columbia, Canada
Shanna Fraser	The University of British Columbia, Canada
Jessica Glidewell	The University of British Columbia, Canada
Joshua Hauck-Whealton	University at Albany, State University of New York, USA
Ted Hoppenstedt	University at Albany, State University of New York, USA
Tracey P. Lauriault	Carleton University, Canada
Rachel McMullin	University at Albany, State University of New York, USA
Peter Runge	University at Albany, State University of New York, USA
Mary Beth Sullivan	University at Albany, State University of New York, USA
Carol Ward	University at Albany, State University of New York, USA
Reginald White	University at Albany, State University of New York, USA
Mark Wolfe	University at Albany, State University of New York, USA
Catherine Yasui	The University of British Columbia, Canada
Sherry Xie	The University of British Columbia, Canada
Jessica Zacher	University at Albany, State University of New York, USA

Research Questions and Methodology

The goals of Domain 2 were articulated in the original Project proposal as a series of research questions.[13] In brief, these questions ask: What do the concepts of reliability, accuracy and authenticity mean in the context of artistic, scientific and governmental activities? To what extent, and how, do records creators in these areas presume and verify their records to have these qualities? How do these presumptions, if they exist, relate to the conceptual requirements for authenticity that the UBC-MAS[14] and InterPARES 1 projects generated for database systems? What intellectual tools, such as guidelines, and what technologies would assist creators to create authentic, reliable and accurate records while still respecting legal obligations, cultural differences, freedom of expression and inquiry and right to privacy?

[13] See Appendix 12.
[14] See http://www.interpares.org/UBCProject/index.htm.

One new direction, implicit in these questions, was that InterPARES 2 did not look only for instances of ideal digital records as they were modeled by InterPARES 1. The arts and science focuses seemed too different to warrant any such presupposition and, in any event, InterPARES 1 itself had discovered that even the documents in administrative systems were in many cases far removed from the ideal. Rather, as suggested by the Authenticity Task Force, the research considered all entities existing in actual systems as well as creators' conceptions of the nature, by-products and products of their activities, and what they understood to be required for presuming authenticity, reliability and accuracy.

Domain 2 accomplished its work through several avenues of investigation. First, researchers combed the literature specific to each focus (i.e., arts, science and government) to find discussions of authenticity, reliability and accuracy, and of related but differently named concepts. They then constructed and published on the InterPARES Web site annotated bibliographies. These bibliographies served as research tools for the other activities of the Project, which were manifested in papers and presentations about the conceptual analysis.

To ground these mostly theoretical discussions, researchers also analyzed reports generated by InterPARES studies of current practices in these fields—both studies of specific cases and more general surveys and interviews of creators. The studies were chosen as exemplars of current practice that showed the potential to cast light on conceptions of the authenticity and reliability of digital records. Domain 2 helped design the research instruments for these studies, in part, to elicit creators' views on its research questions.

The bibliographic research, conceptual analysis and case study reviews provided the main inputs to the products of the InterPARES 2 Terminology Cross-domain, an interdisciplinary research unit directed by lexicographers and experts in knowledge organization. Through a rigorous procedure, the Terminology Cross-domain developed a glossary, dictionary and ontology (a formal description of the concepts existing in the community of creators and preservers studied by the Project).[15] These products rationalized and controlled the language used by InterPARES 2 researchers, who came from quite various disciplinary and national traditions. Domain 2 research provided support, nuance and context to the definitions listed in the glossary as the official working concepts of the Project. For example, the glossary definition of "authenticity" as "the trustworthiness of a record as a record; i.e., the quality of a record that is what it purports to be and that is free from tampering or corruption" is enhanced and modified by the thirteen alternative definitions (listed in the dictionary) that Domain 2 located just in the arts.

At the outset of InterPARES 2, it was hoped that Domain 2 would describe a theory of how to make and keep the records of dynamic and interactive systems in a way that would consider their diverse cultural and disciplinary environments. Progress was made towards such a theory: researchers proposed expansions to the traditional conceptions of a record and of metadata that were appropriate to the interactive and dynamic environments that the Project studied; and sets of guidelines for records creation and maintenance were developed to address the various problems discussed in the literature and observed in the case studies.

Lastly, Domain 2 initiated a test project to transform the documents in one of the case studies into preservably reliable, accurate and authentic records. This was informed not only by the theoretical investigations outlined above, but also by the Domain's participation in modeling sessions that identified the procedures, inputs, outputs, resources and controls on creation.

This report summarizes the results of Domain 2's work. Each of the following three sections addresses one of the InterPARES 2 Focus Task Forces. Within each section, the conceptual

[15] See http://www.interpares.org/ip2/ip2_terminology_db.cfm.

analysis is reviewed and compared with the analyses of the case studies and general studies relevant to that focus. The concluding three sections of this report consider the extent to which the tools developed by InterPARES 1 are adequate to cover the preservation concerns of the focuses, and set forth the basis for guidelines to assist creators in producing preservably authentic records, should they so desire.

At the most general level, Domain 2's findings are not surprising. For the most part, creators' conceptions conform to the various senses of the terms exposed in the theoretical literature of their discipline (for example, by those authors surveyed in the review of these concepts in the arts, summarized below); however, they do so in informal and overlapping ways. Artists, scientists and bureaucrats have very different ideas about the documents they create and reference, what needs to be kept and the features that are essential; terms that have a fairly precise meaning to the archival profession have very different, even contradictory, meanings to these creators. The diversity the Domain 2 researchers found has been reflected in some current thinking about the constructed nature of the concept of authenticity.[16] It is hoped that the specifics of the conceptual analysis presented here, which attempts to carve out semantic boundaries and make clear distinctions among similarly-named concepts, will promote better communication among all interested parties.

The Domain 2 researchers found that although the case studies were quite diverse, they shared many common problems: technological obsolescence, lack of control over creation procedures, insufficient documentation and uncertainty about what digital objects needed to be saved. The ubiquity of the problems identified helped focus the drafting of guidelines for making and maintaining digital materials.[17]

The Domain 2 researchers also found that the benchmark requirements of InterPARES 1 were indeed useful for measuring a presumption of the authenticity of creators' records, for many instances were observed in which documents could not be preserved because they lacked some essential attribute that the requirements identified. However, the researchers also found that it is difficult to apply, or even to adapt, the requirements to the variety of systems that the Project studied, and many cases were observed where they were not sufficient to preserve the kinds of authenticity that the creators valued. The conceptual analyses and experiences of the case studies provide valuable insight into these additional aspects of preservation.

Focus 3 – Government

Scope of the research

The governmental sector presented a terrain that was in many ways very familiar to InterPARES researchers. Indeed, the bulk of the work done studying issues around the creation, maintenance and preservation of digital records has been on the records of governments or similar bureaucratic entities. Yet, InterPARES 2 concentrated its focus on records in interactive, dynamic and experiential systems, particularly those records created in what has come to be called e-government—the use of digital technology to improve the delivery of governmental information and services to the citizenry. Such services typically delivered in an interactive mode through the World Wide Web pose new challenges to both creators and preservers. But the

[16] Heather MacNeil and Bonnie Mak (2007), "Constructions of Authenticity," *Library Trends: Recent Trends in CulturalHeritage Preservation* 56(1): 26–52.
[17] See Appendix 20.

constraints on the creators of government records result in a different environment than the ones studied by the other two focus areas of InterPARES 2. The freedom of expression enjoyed by records creators in the arts, for example, is not at all characteristic of the bureaucratic laws and regulations controlling the e-government environment. Accountability is the watchword with government records as public officials strive to ensure the rights of citizens while maintaining the ability to demonstrate their own faithful execution of duties in the public trust.

Conceptual analysis: authenticity, accuracy and reliability in the literature of e-government

The literature review studying the concepts of authenticity, accuracy, and reliability in e-government rather quickly shows that these concepts are seldom addressed directly. Most e-government literature is focused on delivery options and how to improve them—in short, obtaining the maximum efficiency in the use of information technology to meet the needs of citizens. Any concerns about authenticity in the electronic environment are generic, without singling out the records produced by these newer kinds of electronic systems. This is understandable, for consistency across records in different formats is appropriate. At the same time, it seems clear that older concerns about such issues have carried over from the paper environment. Of these three concepts addressed in Domain 2, it is authenticity that has received the most attention within the governmental sector.

A good example of how these issues have been discussed in the governmental sector can be seen in the electronic records guidelines issued by the New York State Office for Technology. The guidelines were developed "to provide general direction on how state and local government agencies can ensure the authenticity, integrity, security, and accessibility of electronic records (e-records)."[18]

Archivists are keenly aware of how terminology in their field has been used imprecisely, and this is a particular problem in the area of digital records.[19] It should not be surprising, therefore, that the terminology used in the governmental sector in discussing digital records is at times vague or inconsistent in its usage. This is particularly true for words like "authenticity," "accuracy" and "reliability," which are not technical terms at all, but words with common sense, everyday meanings.

The New York State guidelines include a glossary, thereby clearly acknowledging the need to define some of the technical terms used in the document as well as those more common words in need of a precise definition. "Authenticity" is included in the glossary but accuracy and reliability are not. The definition of authenticity, however, is very constricted: "[It] refers to the methods used to verify the source or origin of an e-record. Authenticity is closely related to the concept of *integrity*."[20] The InterPARES definition of authenticity is more in line with the concept of integrity defined in the New York State glossary, which begins by asserting that integrity "is the attribute that the record's contents have not been changed, deleted or otherwise altered."[21] This definition goes on to draw in accuracy as part of integrity, stating that "[i]n addition, integrity addresses the accuracy and timeliness of the contents of a record." Finally, this

[18] New York State Office for Technology (2002), "E-Records Guidelines: Ensuring the Security, Authenticity, Integrity, and Accessibility of Electronic Records." Part 4 of *Electronic Signature and Records Act (ESRA) Guidelines*. Available at http://www.oft.state.ny.us/arcPolicy/policy/ESRAGuidelines4.htm.
[19] This is one of the reasons why InterPARES 2 established its Terminology Cross-domain Task Force.
[20] New York State Office for Technology (2002), "Glossary." Part 5 of *Electronic Signature and Records Act (ESRA) Guidelines*. Available at http://www.oft.state.ny.us/arcPolicy/policy/ESRAGuidelines5.htm. Italics as in original.
[21] Ibid.

definition emphasizes the legal importance of maintaining authenticity and integrity, noting that "[b]oth authenticity and integrity are derived from the legal arena and have a strong bearing on the legal admissibility of records."

The fact that the accuracy of the content of a record should be subsumed under integrity is a bit surprising, since accuracy of content is not addressed in the ESRA guidelines themselves. Instead, accuracy is used as an attribute of systems: "The reliability and accuracy of the systems, processes and procedures used to create, capture, and maintain e-records are critical to demonstrating their authenticity and integrity."[22] Similarly, later in the guidelines, government workers are cautioned: "Make sure the system performs in an accurate, reliable, and consistent manner in the normal course of business" to ensure that records are acceptable "for legal, audit, and other purposes."[23] Thus, accuracy is addressed only peripherally, and generally as an attribute of the way a system should function. Otherwise, it might be assumed that maintaining the integrity of a record will also ensure that the accuracy of the contents of the record are also maintained intact.

Similarly, "reliability" is a concept that seems more to be a characteristic of systems than of records, as noted above in the quote about reliability and accuracy of systems being central to demonstrating the authenticity and integrity of digital records. This is expressed elsewhere in the ESRA guidelines more specifically as "the reliability of hardware and software" affecting "the authenticity and integrity of e-records."[24] In sum, then, accuracy and reliability are viewed primarily as attributes of systems or the functioning of systems. It seems to be implicit that reliable systems will keep reliable records. Perhaps the accuracy of the content of records is not addressed more specifically because its importance is taken as a given that is not unique to the world of digital records. Such a position, however, overlooks the greater potential for change in the content of records in a digital environment.

A more fruitful discussion can be had of the term "authenticity" and its closely allied term "integrity." The ESRA glossary definition clearly establishes a link between authenticity and authenticate when it says that authenticity "refers to the methods used to verify the source or origin of an e-record." The guidelines themselves, however, almost always link authenticity to integrity; indeed, they seem to be virtually synonymous.

Preference for use of the word "integrity" instead of "authenticity" is evident in the United States *E-Government Act of 2002*.[25] Here integrity encompasses authenticity, for it is defined as "guarding against improper information modification or destruction, and includes ensuring information nonrepudiation and authenticity,"[26] where nonrepudiation is defined as "the ability to ensure that a party to a contract or a communication cannot deny the authenticity of their signature on a document or the sending of a message that they originated."[27] Integrity is linked with confidentiality and access as part of information security, which is applicable to "protecting information and information systems."[28] "Accuracy" is referenced in the Act in the context of how the integration of federal information systems will help assure and validate the accuracy of information. Obviously, accuracy is regarded as important, but it is assumed as a given requirement for the content of records and thus is not given much direct attention here or

[22] Ibid., Part 4, op. cit.
[23] Ibid.
[24] Ibid.
[25] United States Congress (2002), *E-Government Act of 2002*. 107th Cong., 2d sess., 2002. H.R.2458.ENR.
[26] Ibid., Sec. 3542(b)(1)(A).
[27] SearchSecurity.com. Available at http://searchsecurity.techtarget.com/sDefinition/0,,sid14_gci761640,00.html.
[28] United States Congress, *E-Government Act of 2002*, op. cit., Sec. 3542(b)(1).

elsewhere in the literature. "Reliability" is a term not used in the Act, although one might infer that it would be considered as part of the discussion of information security.

Terminology and definitions found in documents on digital records and data on various U.S. state Web sites show a similar range of imprecision in application of these terms, though all are within a commonly accepted sense. The word "trustworthy" has been adopted by the state of Minnesota in its *Trustworthy Information Systems Handbook*, and trustworthy is defined as an attribute of records that "contain information that is reliable and authentic."[29] Clearly, the title indicates that trustworthy applies to systems as well as to records themselves. The handbook's glossary formally says, "Authenticity is a function of a record's preservation and is a measure of a record's reliability over time," while reliability is defined as "the measure of a record's authority and is determined solely by the circumstances of the record's creation." The state of Texas offers a definition of authenticity that is similar to that proposed by InterPARES, but reliability seems to refer to the ability to sustain and reproduce records accurately into the future.[30] The state of Wisconsin offers a similar definition of "reliable," while its definition of "authentic" is that "the retained electronic record correctly reflects the creator's input and can be substantiated."[31] Unlike most such documents, the Wisconsin standards also define "accurate" as meaning that "all information produced exhibits a high degree of legibility and readability and correctly reflects the original record when displayed on a retrieval device or reproduced on paper."

Authenticity, accuracy and reliability in the governmental sector case studies

InterPARES 2 researchers carried out eight case studies within the government focus area. The case study reports provide additional insight into how those working in the governmental sector regard the concepts of authenticity, reliability and accuracy in the context of their own electronic systems. The reports tend to confirm much of what the conceptual analysis shows about authenticity, reliability and accuracy in the e-government area. Often, authenticity is either presumed, providing that requisite procedures have been followed, or is tied to authentication methods, such as PKI (Public Key Infrastructure). There is concern with accuracy of information, but it is likely to be limited to the accuracy of data at the point of creation, after which accuracy (like authenticity) is presumed to be protected by procedural controls.

One of the more useful discussions of these issues in the case study reports is in case study 20, Ireland's Revenue On-Line Service (ROS). Here there is a presumption of the "authenticity of received, signed and submitted tax forms from authorised users."[32] In addition, a "chain of authenticity" is presumed based on "user log-ins, digital certificates and PKI."[33] This "security wrapper" is retained "to confer authenticity and non-repudiation over time."[34] The study acknowledges that it is unclear whether this approach will work over time, however. Nevertheless, it is clear that the ROS relies on these external controls to convey authenticity and

[29] Minnesota Historical Society, State Archives Department (2002), *Trustworthy Information Systems Handbook*. Version 4. Available at http://www.mnhs.org/preserve/records/tis/tableofcontents.html.
[30] Texas Department of Information Resources (2004), *Architecture Components for the Enterprise. Data and E-Records Management Domain*. Available at http://www.dir.state.tx.us/ace/documents/phase1toc.htm.
[31] Wisconsin Department of Administration (2001), *Administrative Rules: Adm 12, Electronic Records Management - Standards and Requirements*. Available at http://www.legis.state.wi.us/rsb/code/adm/adm.html.
[32] John McDonough, Ken Hannigan and Tom Quinlan (2005), "InterPARES 2 Project - Case Study 20 Final Report: Revenue On-Line Service (ROS)," 77. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs20_final_report.pdf.
[33] Ibid.
[34] Ibid.

sees this authenticity as carrying through time. The discussion of reliability in this case study suggests that it is a concept not well differentiated from authenticity. Again, it is the controlled environment, through the use of passwords and PKI, which confers reliability on records. In this case study, accuracy relates to the factual accuracy of the data in records, but since individuals and revenue agents may enter inaccurate data, full accuracy of the data cannot be guaranteed. "A certain number of business rules and logic to check calculations" are built into the system to help guard against inaccurate data being entered.[35]

In the City of Vancouver Geographic Information System (VanMap) case study (case study 24), accuracy is sometimes referred to as "data quality." There is a concern that information added to VanMap be as accurate as possible before it is entered. Since data comes from external sources, however, its accuracy cannot be guaranteed. However, the VanMap team does guarantee that information is as accurate as it was when it was entered into the system, that "the data are not altered in such a manner as to affect their accuracy and reliability."[36] Note that in this answer reliability is treated as a characteristic of data. When decisions based on information in VanMap are being made, staff may verify the data or seek other kinds of independent verification. Clearly there is a concern for the factual accuracy of data in the system, since there is a recognition that it can affect decision-making. In fact, this case study treats authenticity as residing largely in the accuracy of the data.

Case study 21 looks at the electronic filing system in place at the Supreme Court of Singapore. In this case study, "reliability" at first seems to refer to the overall system, for it relates to the ability of the court to process submissions from attorneys and litigants in an efficient manner and to keep track of the large number of files in the system. But the records are considered reliable as well, because "they are created and modified in a controlled environment with access privileges assigned to respective action officers based on their job responsibilities."[37]

Accuracy in the Singapore case relates to "the provision of accurate information from the case records."[38] Various methods are in place to ensure that the records are accurate, although the concern with accuracy seems to be centred more on creation of the records than on their maintenance. Perhaps this is seen more as a matter of protecting authenticity, for, once the records are created, preserving their authenticity would seem to include preserving accuracy. Here one sees also the confusion of authenticity and authentication. The use of authentication technologies is the source of a presumption of authenticity. By use of PKI and other security controls, alteration or tampering of information in the files is prevented and authenticity is ensured, although not necessarily for the long term.

Of all the case studies in the government focus, the computerization of the Alsace-Moselle land registry (case study 18) offers perhaps the most sophisticated technological awareness of issues of accuracy, authenticity and reliability. For more than a century, a rigorous system was in place to provide "accurate, reliable, and authentic information" through a paper recordkeeping system that has now been computerized.[39] The case study report details all the steps taken in development of the new system. Here it is only important to pay attention to how the concepts of

[35] Ibid., 75.
[36] Evelyn McLellan (2005), "InterPARES 2 Project - Case Study 24 Final Report: City of Vancouver Geographic Information System (VanMap)," 26. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs24_final_report.pdf.
[37] Elaine Goh (2005), "InterPARES 2 Project - Case Study 21 Final Report: The Electronic Filing System (EFS) of the Supreme Court of Singapore," 39. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs21_final_report.pdf.
[38] Ibid.
[39] 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," 20. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs18_final_report.pdf.

accuracy, authenticity and reliability are viewed in the process. First, one can note that quality of the data is the way in which the concept of accuracy is expressed. Authenticity, integrity and reliability are all regarded as attributes of data or information, and the terms are often used interchangeably. Authentication techniques are key features of the system and are seen as central to the guarantee of authenticity, integrity or reliability.

A somewhat different take on the concept of accuracy emerges in the case study of the Archives of Ontario Web Exhibits (case study 05).[40] Here accuracy is seen not as relating to the digital components of the records so much as to the concept of historical or narrative accuracy of an exhibit as a whole. In other words, the creators of the Web exhibits seek to present factual information with proper documentation but leave interpretation to those viewing the exhibit. Although this seems to be a significantly different conception of accuracy in digital records, one could argue that this is analogous to the concern for correct information being entered into tax forms in the ROS case study noted above. In case study 05, authenticity is presumed, particularly within the environment of the creator, but whether these Web exhibits can be considered to be "authentic" when viewed by users will depend in part on how they are displayed in any given technological environment.

This sampling of findings on authenticity, reliability and accuracy from the InterPARES 2 case studies in the government focus indicates that there is an awareness of the importance of these three concepts. Reliability is less often addressed, and when it is it seems to be used as a synonym for authenticity or at least is seen as inextricably connected with authenticity. Both authenticity and reliability are deemed to be protected by procedural controls and authentication techniques. Accuracy is sometimes equated with data quality, a somewhat fuzzier concept, and one which was not detected in the literature review in the government focus. The controlled bureaucratic process in government lends a sense of confidence that structure and procedure will help maintain these essential characteristics of government records.

Conclusions

Because of the long-recognized necessity for government to be accountable for its actions, the need to maintain records that can be demonstrated to be authentic and reliable is not a new concept for those responsible for creating and maintaining government records. As Minnesota's *Trustworthy Information Systems Handbook* states, "We need trustworthy information systems to ensure our accountability as government agencies."[41] Reliability and authenticity are characteristics of information and records that are essential to trustworthiness. The fact that so much attention has been paid to the challenges posed by digital records in the governmental sector is evidence that concern for these issues has taken strong root. Government archives and records managers have taken the lead in impressing these issues on executives, agency heads and information technology personnel. The InterPARES research in Domain 2, however, suggests that terminology is used loosely and that adjectives such as "authentic," "trustworthy" and "reliable" are applied at different times to information, data, records and systems. Accuracy is either considered a characteristic of authenticity or something that is generally beyond the control of government workers, insofar as it relates to the factual accuracy of data entered into forms by others. As with paper records, authenticity is often presumed, particularly in instances

[40] Jim Suderman et al. (2004), "InterPARES 2 Project - Case Study 05 Final Report: Archives of Ontario Web Exhibits." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs05_final_report.pdf.
[41] Minnesota Historical Society, *Trustworthy Information Systems Handbook*, op. cit.

where authentication techniques are employed. Although there is a commendable respect for authenticity, reliability and accuracy in the government focus, there needs to be a greater sense of the complexities involved in maintaining these characteristics for records requiring long-term preservation.

Focus 2 – the Sciences[42]

Scope of the research

The crucial importance to society of scientific research and its ready use of the latest technologies were key factors in leading InterPARES 2 researchers to make scientific activities one of the Project's three focuses for investigation. Archivists have done significant work in the area of documentation and appraisal in the sciences—work that has been supplemented by the scientific community's efforts in recent years[43]—to address the impact of information technologies on scientific research and recordkeeping. With traditional archival definitions of the concept of record under review in the digital era, scientific activities also seemed to be a fruitful area of study because of the different perspectives that scientific researchers have about records and recordkeeping practices.[44]

The sciences are a broad and heterogeneous area for study, and it was not possible for InterPARES researchers to investigate all branches of the world of science. The study of the scientific concepts of authenticity, accuracy and reliability, however, began with a broad literature review across the disciplines, seeking specific discussions of, or references to, those three key concepts. The findings of the literature review were supplemented with data from various InterPARES 2 case studies that focused on scientific data and records creation, management, appraisal and retention. Finally, broader, more disciple-wide perspectives on these issues were provided by a number of general studies carried out by the Focus 2 researchers.

---

[42] InterPARES wishes to acknowledge SUNY Graduate Research Assistant, Joshua Hauck-Whealton, for his help in compiling the case study and bibliographic data used in this section.

[43] See, for example, National Research Council, Commission on Physical Sciences, Mathematics and Applications, *Preserving Scientific Data on Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources* (Washington, D.C.: National Academy Press, 1995). Available at http://www.nap.edu/catalog.php?record_id=4871#toc; National Science Foundation, *Report of the National Science Board: Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*, NSB-05-40, September 2005. Available at http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf; David F. Strong and Peter B. Leach, *National Consultation on Access to Scientific Data: Final Report* (Ottawa: Canadian Institute for Scientific and Technical Information, National Research Council Canada, 2005). Available at http://ncasrd-cnadrs.scitech.gc.ca/NCASRDReport_e.pdf; Kenneth Thibodeau (1995), "Preserving Scientific Data on Our Physical Universe," *IASSIST Quaterly* 19(4): 26–29. Available at http://iassistdata.org/publications/iq/iq19/iqvol194thibodeau.pdf; Joan Warnow-Blewett, Joel Genuth and Spencer R. Weart, *AIP Study of Multi-Institutional Collaborations: Final Report. Highlights and Project Documentations* (College Park, MD: Center for History of Physics, American Institute of Physics, 2001). Available at http://www.aip.org/history/pubs/collabs/highlights.pdf; Library of Congress, *National Digital Information Infrastructure and Preservation Program (NDIIPP), Digital Preservation*. Available at http://www.digitalpreservation.gov/index.html; David L. Brown, Grace Welch and Christine Cullingworth (2005), "Archiving, Management and Preservation of Geospatial Data: Summary Report and Recommendations," GeoConnections Policy Advisory Node: Working Group on Archiving and Preserving Geospatial Data. Available at http://www.geoconnections.org/publications/policyDocs/keyDocs/geospatial_data_mgt_summary_report_20050208_E.pdf; and CODATA Working Group on Archiving Scientific Data, The Committee on Data for Science and Technology (CODATA) of the International Council for Science. Available at http://www.nrf.ac.za/codata/.

[44] See Tracey P. Lauriault, Barbara L. Craig, D. R. Fraser Taylor and Peter L. Pulsifer (2007), "Today's Data are Part of Tomorrow's Research: Archival Issues in the Sciences," *Archivaria* 64 (Fall): 123–179.

Conceptual analysis: authenticity, accuracy and reliability in the literature of the sciences

A recent European Task Force on permanent access to scientific records neatly lays out the scope of the definitional problem:

> The definition of "the records of science" must be broad to ensure an accurate record of the research process and its results is created. Within this definition it is essential to include both the formal, structured 'minutes of science' (published records in the formal refereed scientific literature) and the less structured, informal communication mechanisms which are now commonly used to share topical ideas and information (such as Web sites, moderated bulletin boards and email). In addition, through improvements in computing and networks linked with powerful data and text mining techniques, new research practices have developed which are data-intensive and highly collaborative. Scientific records may take the form of raw data, aggregated into datasets relevant to a particular topic or field. These can be numeric, graphic or textual and may contain embedded logic (chemical compound structures, crystallographic data, genome sequences).[45]

This wide-ranging definition includes the published results of scientific research, which would not be considered "records" in the traditional archival sense, as well as forms of documentary communication, such as e-mail, readily recognizable as records. But the real emphasis in this definition seems to be on scientific data and the various aggregations in which they are gathered and utilized for research purposes. InterPARES 2 research into scientific activities confirms that most scientists do not think of records in the archival sense but rather tend to equate "scientific records" with "scientific data," sometimes using the term "scientific data records." Thus, it is important for archivists to understand that for scientists, scientific data can also be defined more precisely to mean "numerical quantities or other factual attributes generated by scientists and derived during the research process (through observations, experiments, calculations and analysis),"[46] and "numbers, images, video or audio streams, software and software versioning information, algorithms, equations, animations, or models/simulations."[47] Although the former definition is more in keeping with the definition of *data* provided in some archival contexts,[48] it is interesting to note that the latter definition circumscribes a broader concept that includes digital entities that, depending on their context of creation and use, archivists would more readily identify as records.

Nevertheless, the Focus 2 research suggests that, in the eyes of many scientists, maintaining archives in the sciences is less an act of recordkeeping and more an act of data management and processing. "A key component of creating the public archive of information is the efficient capture and curation of the data—data processing."[49] Parallel to this difference in focus is the perception that the scientific community evinces less overt interest in authenticity than it does in accuracy and reliability.

---

[45] European Task Force Permanent Access (2005), "Permanent Access to the Records of Science: Proposal for a Research & Development Programme," 3–4. Available at http://www.alliancepermanentaccess.eu/Proposal%20Research%20and%20Development.doc.

[46] CODATA Working Group on Archiving Scientific Data, op. cit.

[47] National Science Foundation, *Report of the National Science Board*, op. cit., 18.

[48] For example, the word *data* is defined as "Facts, ideas, or discrete pieces of information, especially when in the form originally collected and unanalyzed," by Richard Pearce-Moses in his *A Glossary of Archival and Records Terminology* (Chicago: Society of American Archivists, 2005). Available at http://www.archivists.org/glossary/.

[49] Helen M. Berman et al. (2000), "The Protein Data Bank," *Nucleic Acids Research* 28(1): 235–242. Available at http://nar.oxfordjournals.org/cgi/content/full/28/1/235.

Of the three concepts of authenticity, reliability and accuracy, it is the last concept, accuracy, that receives the greatest attention in scientific literature; more specifically, concepts related to data quality. This is no doubt attributable to the fact that scientists focus on data, not records, and the accuracy of data is obviously crucial to the validity of scientific research. Conversely, the concept of accuracy is not as prominent in archival theory, but the InterPARES definition of accuracy is broad enough to include data and datasets as well as documents and records. Data are different than records, of course. Data are often still in the process of being used and modified, and have thus not been properly "set aside" in the fashion of records.

In an archival context, an accurate record is one that contains correct, precise and exact data, which is often adjudged in relation to the absoluteness of the data it reports or its perfect or exclusive pertinence to the matter in question. Furthermore, in an archival context, the accuracy of a record is *assumed* when the record is created and used in the course of business processes to carry out business functions, based on the assumption that inaccurate records harm business interests.[50] This assumptive approach toward the assessment of accuracy stands in marked contrast to the approach used in scientific research, where, because errors and uncertainty are a given, a general analytical tenant is that "no number has meaning unless it is accompanied by an estimate of uncertainty."[51] It is for this very reason, in fact, that accuracy is seen by most scientists as the most common and critical metadata element.

To an archivist, an authentic record does not have to be an accurate record, however. Although it is true that an authentic record is as reliable and accurate as it was when first generated, this is not the same thing as saying that authenticity ensures that the content of a record at the point of its creation is accurate. Thus, authenticity alone does not "automatically imply that the content of a record is reliable"[52] or accurate. Scientists, on the other hand, give primacy to data quality, which includes the concept of authenticity, normally articulated as data provenance or lineage. In this context, data accuracy is critical and the data need to be reliable. Data quality is normally articulated in a dataset's metadata; without metadata or data quality parameters, a scientist will not use, trust or rely on those data. Although each scientific discipline differs in how it defines scientific data quality, most include some or most of the following data quality elements: positional accuracy; attribute and thematic accuracy; completeness; semantic accuracy; and temporal information, reliability, lineage, logical consistency and objectivity.[53] In the fields of Geography and Geomatics, a number of standards have been published regarding the quality of geographic data.[54] The procedures described in the International Standard, ISO 19113:2002, provide a consistent and standard manner to determine and report a dataset's quality. The International Cartographic Association has also written a book on the subject, entitled *Elements of Data Quality*.[55]

The concept of precision is related but distinct from accuracy. "*Precision* refers to how *exact* and *reproducible* a measurement or estimate is, irrespective of its accuracy, while *accuracy* refers to how *close* a measurement or estimate is to the correct value."[56] A device used for

[50] See *Creator Guideline*s, guideline 4, in Appendix 20.
[51] National Research Council, *Preserving Scientific Data*, op. cit., 37.
[52] Pearce-Moses, *A Glossary of Archival and Records Terminology*, op. cit.
[53] Stephen C. Guptill and Joel L. Morrison (eds.), *Elements of Data Quality* (Oxford: Elsevier Science, 1995).
[54] International Organization for Standardization, ISO 19113:2002 - Geographic Information—Quality Principles; International Organization for Standardization , ISO 19114:2003 - Geographic Information—Quality evaluation procedures.
[55] Guptill and Morrison, *Elements of Data Quality*, op. cit.
[56] Randy Preston (2006), "InterPARES 2 Project - General Study 09 Final Report: Digital Recordkeeping Practices of GIS Archaeologists Worldwide: Results of a Web-based Survey," 74. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs09_final_report.pdf. Italics as in original.

measurement may be precise, by returning measurements that are always the same or close to the same, but not be accurate. A dataset may be reasonably accurate when all data points are close to the actual reality, but fail to be precise because of a wide spread of data. Although the InterPARES definition of accuracy encompasses the concepts of correctness and precision, the two concepts have different and important meanings within the sciences.

An additional feature of the concept of accuracy in the sciences is how it is affected by the idea of timeliness. It is generally assumed that there will be advances in the methods of data collection, and thus "evolving, improving accuracy of the determination" of data.[57] As a result, a more recent measurement will be considered more accurate; that is, more correct or closer to the actual reality. This suggests that data have a kind of shelf life, since the creation of a new (and more accurate) dataset will render an older dataset obsolete. It is recommended that obsolete data sets be discarded: "If the data have been completely superseded by better data . . . destruction of the records may be in order."[58]

It is useful here to understand the difference between this idea of data shelf life and another concept referred to in the literature as "currency" or "temporal accuracy." Both concepts suggest that older data should be replaced by more current data. However, the idea of currency reflects the fact that the reality itself has changed and that the presence of more accurate data is not just the result of improved methods or instruments. Currency, therefore, applies in situations where the facts being measured shift and change, such as in the compilation of maps or in population surveys. By the strictest definitions, data that no longer have currency are still accurate, as they truthfully depict reality the way it was at the time the measurement was taken. However, they may no longer be useful to the scientist for whom temporal accuracy is crucial.

As indicated earlier, the concept of authenticity does not loom large in the scientific literature. Any concern for authenticity that occurs is often part of the greater concern for accuracy. Several sources stress the need to properly assign responsibility for a given source of data so that errors can be fixed through correspondence with the creators. An article on digital signatures in electronic health records asserts, "The attribution of responsibility is a means to ensure the accurateness of the information."[59] In a paper on the protein data bank, the author writes, "In almost all cases, serious errors detected by these checks are corrected through annotation and correspondence with the authors."[60]

Ironically, however, although most organizations aim to ensure that their data are accurate, reliable and authentic, the Focus 2 case and general studies observed that many of these same organizations add disclaimers to absolve themselves of any responsibility for damages that may result from the use of their data. These types of disclaimers were used by a number of the scientific data portals surveyed in general study 10. Examples include the Antarctic Digital Database (ADD) (data portal IP2SF27), which cautions that its maps, when combined, may reflect some inconsistencies, particularly when older datasets are included; the National Geophysical Data Center (data portal IP2SF26) and the World Data Center for Solar Terrestrial Physics (data portal IP2SF10), both of which indicate that the Government of the United States

[57] National Research Council, Committee on Issues in the Transborder Flow of Scientific Data. *Bits of Power: Issues in Global Access to Scientific Data* (Washington, D.C.: National Academy Press, 1997), 48. Available at http://www.nap.edu/catalog.php?record_id=5504#toc.
[58] John L. Faundeen (2003), "The Challenge of Archiving and Preserving Remotely Sensed Data," *Data Science Journal* 2: 162. Available at http://www.jstage.jst.go.jp/article/dsj/2/0/159/_pdf.
[59] J. J. Bos (1996), "Digital Signatures and the Electronic Health Records: Providing Legal and Security Guarantees," *International Journal of Bio-Medical Computing* 42(1-2): 159.
[60] Berman et al., "The Protein Data Bank," op. cit., 237.

and its employees cannot be held accountable for any data quality warranties; the FMRI Data Center (data portal IP2SF7), which absolves itself from liability in relation to data quality; the Indiana University Bio Archive (data portal IP2SF5), which reminds users that data contain errors; and the British Atmospheric Data Centre (BADC) (data portal IP2SF1), which absolves itself from responsibility of data on its Web site and once downloaded onto the user's computer.[61] A slightly different type of disclaimer is provided by the Cybercartographic Atlas (case study 06), which states that its content is intended to be used primarily for educational purposes and that the "[d]isclaimers and caveats on the Web site are intended to limit the [legal] responsibility of the Creator."[62]

These examples suggest that although scientists are unlikely to use the word authenticity, they are nevertheless very concerned about the identity and integrity of the data they use. Identity and integrity are the key components of authenticity in the understanding of InterPARES.[63] Within a science context, identity is established by metadata that record the phenomena observed and the kinds of measurements obtained and by which instruments and at what time and place and under whose responsibility. Integrity, on the other hand, means that the data have not been altered, either by unauthorized tampering or by corruption due to technical failure since they were first created. In the sciences, data integrity is often guarded first by *authentication* and other security measures to prevent unauthorized tampering, then by checksums or other techniques to spot altered bits.

As used in the field of Communications Science and Engineering, *authentication* refers to "security measures designed to protect a communication system against fraudulent transmissions and establish the authenticity of a message," while an *authenticator* is a "letter, numeral or groups of letters attesting to the authenticity of a message or transmission."[64]

As in other areas, however, the process of authentication and the use of authentication technologies are not sufficient in themselves to be a guarantor of authenticity of preserved scientific records over time, since authentication can, at best, only establish the authenticity of a record at *a specific point in time*. Moreover, authentication may, in many instances, simply mean establishing the identity (and therefore the trustworthiness) of an agent associated with a record, not of the data or record itself: "Authentication is a security service that consists of verifying that someone's identity is as claimed."[65] On the other hand, authentication procedures and technologies can be an important and perhaps, in some cases, essential element in an overall preservation strategy that is designed to guarantee the authenticity of preserved scientific records over time.[66]

---

[61] See Tracey P. Lauriault and Barbara L. Craig (2007), "InterPARES 2 Project - General Study 10 Final Report: Preservation Practices of Scientific Data Portals." Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs10_final_report.pdf.

[62] Tracey P. Lauriault and Yvette Hackett (2005), "InterPARES 2 Project - Case Study 06 Final Report: Cybercartographic Atlas of Antarctica," 5, 27. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs06_final_report.zip.

[63] See MacNeil et al., "Authenticity Task Force Report," op. cit., 47, specifically, the section titled, "Conceptual findings: the requirements for authenticity."

[64] McGraw-Hill, *Dictionary of Scientific and Technical Terms*, sixth edition (New York: McGraw-Hill, 2003).

[65] Audun Jøsang and Mary Anne Patton, "User Interface Requirements for Authentication of Communication," in *Proceedings of the Fourth Australian User interface Conference on User interfaces 2003 - Volume 18*, R. Biddle and B. Thomas, eds. ACM International Conference Proceeding Series, vol. 36 (Darlinghurst, Australia: Australian Computer Society, 2003), 75. Available at http://portal.acm.org/ft_gateway.cfm?id=820105&type=pdf&coll=&dl=&CFID=15151515&CFTOKEN=6184618.

[66] See, for example, William E. Underwood (2002), "A Formal Method for Analyzing the Authenticity Properties of Procedures for Preserving Digital Records," in *Proceedings of the 2002 International Conference on Digital Archive Technologies (ICDAT2002),* December 19-20, 2002, Academia Sinica, Taipei, Taiwan, 53–64. Available at http://perpos.gtri.gatech.edu/publications/ICDAT2002.pdf#page=1. In this study, the author proposes a formal method for analyzing records management and archival procedures and systems to determine whether they maintain and preserve authentic records over time. The analysis procedure is based on a formalization of archival and diplomatic concepts and principles as

Within the sciences, the trustworthiness of data is also conceived in terms of "data quality," which, although it includes the concept of authenticity, is normally articulated as data provenance or lineage. Data quality is normally articulated in a dataset's metadata; without metadata or data quality parameters, a scientist will not use, trust or rely on those data. Each scientific discipline differs in how it defines scientific data quality, as is demonstrated from the results of the Focus 2 case and general studies. However, most disciplines include some or most of the following data quality elements: positional accuracy; attribute and thematic accuracy; completeness; semantic accuracy; and temporal information, reliability, lineage, logical consistency and objectivity.[67]

Data "lineage" is information about the chain of transmission, from the moment the data were originally recorded, that brought the data to the user. Lineage speaks to the history of a dataset, its lifecycle from data collection to its many stages of compilations, corrections, conversions and transformations and the generation of new interpreted products. This concept might also be characterized as "data provenance" and is clearly related to data integrity. As with traditional paper records, the provenance of a particular scientific dataset is essential in establishing its accuracy and currency.[68]

Not surprisingly, therefore, in a science context, the lifecycle of a dataset, from acquisition to compilation and derivation, comprises important areas of concern to accuracy.[69] With respect to accuracy, *acquisition* is the most important stage in the lifecycle of a dataset, since it is the point where the original observations are collected and where "fundamental assumptions, calibrations and corrections are made."[70] *Compilation* is the stage where a database is created; it occurs when the data are assembled into some sort of comprehensive arrangement or into a scientific dataset, and it is a phase during which many errors can be introduced. *Derivation* is the stage where data are being manipulated; the output of this process is a representation, interpolations, averaging and any number of manipulative techniques that may change the form, format or structure of the data. This may or may not be a reversible phase and is a diversion point from the original observations. For this reason, keeping the raw data as well as derived data is important.

This is why, in a scientific context, data accuracy is critical and why the data need to be reliable. Data quality is normally articulated in a dataset's metadata; without metadata or data quality parameters, a scientist will not use, trust or rely on those data. Metadata are essential for the dissemination of scientific data whereby "a data set without metadata, or with metadata that do not support effective access and assessment of data lineage and quality, has little long-term use."[71] In fact, as the general study 10 findings demonstrate, data portal discovery services rely on metadata descriptions, which are seen as a form of "truth in labelling." For users of these portals, and indeed among scientists in general, it is considered "axiomatic that a database has

definitions and axioms. Concepts such as digital record, record series and archival integrity are defined and axioms characterizing authentic documents and authentic records are formulated. A procedure is described for storing and retrieving the digital records of a record creator that incorporates elements to ensure the integrity and authenticity of the records. The theories of record integrity and authenticity are used with theories of communications security and belief to prove that the procedure achieves its goal of preserving the integrity and authenticity of the digital records.

[67] Guptill and Morrison, *Elements of Data Quality*, op. cit.

[68] Significant research describing the provenance of data in molecular genetics databases is one example of the importance of this concept for validating research (see Mark Greenwood et al. (2003), "Provenance of e-Science Experiments: Experience from Bioinformatics," in *Proceedings UK e-Science All Hands Meeting 2003*, Simon J. Cox, ed. Available at http://www.nesc.ac.uk/events/ahm2003/AHMCD/pdf/047.pdf).

[69] See Derek G. Clarke and David M. Clark, "Chapter 2: Lineage," in Guptill and Morrison, *Elements of Data Quality*, op. cit., 13–30.

[70] Ibid., 18.

[71] National Research Council, *Preserving Scientific Data*, op. cit., 36.

limited utility unless the auxiliary information required to understand and use it correctly—the metadata—is included in the record."[72] In the sciences, metadata are also a means of attesting to and assessing a dataset's authenticity. In other words, authenticity in the sciences is linked to a clear lineage recorded in the accumulating metadata surrounding data, which closely parallels the situation with respect to digital records in general. Both data and their cumulative and related metadata must be present, clear, unambiguous and un-compromised. In the absence of metadata, it is possible to gain some understanding of a scientific dataset if there are associated peer review papers and reports that describe them; however, this would be a more laborious process.

As expected, one can find ample evidence of a concern for data accuracy and authentication in the legal arena. For example, satellite images have been admitted into evidence in a few legal cases. "The admissibility of remote sensing information must be examined within the context of the general requirements for admission of scientific evidence and expert opinion."[73] A litigator seeking the admission of remote sensing data as evidence must (1) qualify an expert, (2) authenticate and prove the contents of the data and (3) establish that proper and accepted processing techniques were employed. The use of an archive history file accompanying the final satellite imagery exhibit provides the potential for objective, external authentication and establishes that appropriate techniques and methodologies were employed in the creation of the exhibit. An archive history file is a document listing (1) all the data used in the creation of the final exhibit, (2) all the tools used in the creation of the final exhibit and (3) all the processes and methods used to create an exhibit.[74]

When one moves away from pure scientific data and into the realm of records as defined by archival theory, a more explicit concern for authenticity does surface. A good example here is the laboratory notebook. The Chemical Sciences Roundtable, among many other groups, has discussed the problems of moving from bound paper notebooks to electronic notebooks. Since these lab notebooks can be used to determine precedence in such things as patent cases, they are legal documents and their authenticity must be established to give them evidential value. As Roundtable panellists explained, "One of the purposes of an electronic notebook is to have a historical record that is used, among other things, for establishing priority and for integrity concerns in science."[75]

The concept of reliability in the sciences is also influenced by the focus on data rather than on records. Because of the focus on the *accuracy* of data, the concept of reliability is more likely to be used in reference to *collections* of data. Scientists presume scientific data to be reliable because they were collected by a federal or state agency or because of the professional reputation of the scientist who collected the data. In the sciences, the concept of reliability is closely associated with the concepts of reproducibility and accuracy. More generally, reliability is a quality that can be attributed to a person, as in a reliable person; to a device, such as a reliable machine; or to a system that is organized to accomplish certain ends, as in a reliable computer or records system. It is the individual assessor who determines what attributes are required before reliability can be reasonably inferred. Thus, to scientists, reliable data are data collected by a

[72] Ibid., 31, as cited in Lauriault and Craig, "General Study 10 Final Report," op. cit., 76.
[73] Sharon. H. Hodge (1997), "Satellite Data and Environmental Law: Technology Ripe for Litigation Application," *Pace Environmental Law Review* 14: 714. Hodge's article references the following cases: *United States v. Reserve Mining Co.*, 380 F.Supp. 11 (D.Minn. 1974) and *Gasser v. United States*, 14 Cl.Ct.476 (1988).
[74] Ibid.
[75] National Research Council, Chemical Sciences Roundtable, *Impact of Advances in Computing and Communications Technologies on Chemical Science and Technology: Report of a Workshop* (Washington, D.C.: National Academy Press, 1999), 173. Available at http://www.nap.edu/catalog.php?record_id=9591#toc.

competent scientist using procedures and instruments that are reliable. Reliability is a matter of degree, however. The reliability of data is determined by examining information about their provenance, asking: Were the data created by a competent person; that is, a person who has professional credentials or is certified by a standards organization?

Thus, a set of data that was generated by a reliable person using trustworthy methodology and that remains complete and uncorrupted might be said to be "reliable." That would mean that the majority of its data points are correct and precise (accurate) and that their integrity has not been compromised. If the methodology was not sound, then the accuracy of the data may be low and the dataset would be unreliable. If the data are internally inconsistent—for example, by giving a wide range of data points or conflicting answers—then the data would be regarded as imprecise and thus unreliable. If the dataset is not complete or is otherwise corrupt, because of a failure at the point of collection to capture the full range of data or because of data loss at a later time, then the set is unreliable.

This difference in usage between the individual data point and the collective dataset is presumably a result of the scientific need for large collections of data. A single data point, although it may in fact be accurate, cannot be trusted by scientists to stand for a fact. A large set of data is needed for data to be reliable. The reason for this is that scientists do not expect absolute accuracy. Regardless of the precision of instruments and the soundness of the methodology, instruments can deliver noisy data leading to mistaken conclusions by the scientist. Therefore, single data points or small datasets cannot be trusted. Datasets need to be robust to be reliable so that the inevitable errors are diluted. Robustness implies large collections of data where the individual entries of data are complete.

In some situations, robustness can be a substitute for accuracy in producing a reliable dataset. Some measurements, such as certain measurements generated in the medical field, are difficult to make accurately. As one article notes, "The very nature of biomedical objects is volatile and irregular . . ."[76] In other areas of inquiry, such as meteorology, obtaining data quickly is more important than achieving high levels of precision and accuracy. When data are inaccurate, having numerous versions of the same observations can help smooth the inaccuracies. With a large dataset, it is also possible to see the overall precision of the data. Outliers can be seen for what they are and examined, thereby increasing the trustworthiness of the dataset. "Especially in areas of high data density, inaccuracies can be detected by humans or by computers from comparison with other data points, making it possible to bypass the inaccuracy."[77]

Authenticity, accuracy and reliability in the scientific sector case and general studies

The case and general studies carried out in Focus 2 tend to confirm the findings on authenticity, accuracy and reliability evident in the literature review. For example, case study 14 (Archaeological Records in a Geographical Information System) found: "There is more concern

[76] A. Minitski, A. Mogilner, C. MacKnight and K. Rockwood (2003), "Data Integration and Knowledge Discovery in Biomedical Databases. Reliable Information from Unreliable Sources," *Data Science Journal* 2: 25. Available at http://journals.eecs.qub.ac.uk/codata/journal/contents/2_03/2_03pdfs/DS131.pdf.

[77] National Research Council, Steering Committee for the Study on the Long-term Retention of Selected Scientific and Technical Records of the Federal Government. *Study on the Long-term Retention of Selected Scientific and Technical Records of the Federal Government: Working Paper*s (Washington, D.C.: National Academy Press, 1995), 58. Available at http://books.nap.edu/catalog.php?record_id=9478#toc.

over the reliability and accuracy of the records than the authenticity."[78] It might be argued that authenticity is a by-product of this emphasis on accuracy, but, as a distinct concept, authenticity is not developed in the studies. The lack of a strong understanding of authenticity is a result of the scientific emphasis on data and information over the records that contain them.

In case study 14, the archaeologists involved in the operations of the GIS have only thought about the concepts of reliability, accuracy and authenticity in terms of the data rather than in terms of the record. From the creator's viewpoint, reliability and accuracy relate to the reliability of the data source, so it is assumed that if the source is reliable the data will be reliable and accurate. In fact, the archaeologists in this case study assumed authenticity on the grounds that the datasets were obtained from a state repository or from a researcher trusted as a professional who maintains information securely. Yet, there was a sense of discomfort with the concept of authenticity as applied to records because the archaeologists saw their work as an ongoing compilation that could not be broken up into discrete units and treated as records.

Case study 19 (Preservation and Authentication of Electronic Engineering and Manufacturing Records) reported on an engineering experiment to develop an open-source preservation format for digital computer-aided design (CAD) records of solid models used in high-tolerance manufacturing of complex assemblies. The business owners in this study use CAD records in the science-based manufacturing of high-assurance, high tolerance machined piece parts for the U.S. government. In their words, "there is a critical, unsolved business requirement to maintain authentic records over time to enable the production of the pieces as long as the business requires them, with the assurance that they meet the same strict standards (tolerances) as the original piece."[79] The intent of the experiment was to preserve not only the geometric specifications of the model but also its semantically encoded metadata, joined to make a "new logical preservation format" for archival purposes. By "logical preservation format," the experiment partners in this study meant a format encompassing not only the fixed form and content of information representing the model but also instructions encoded within its metadata so that reasoning engines of the future can conduct "proofs" against the object to authenticate it as fit to support the procedural action for which it was designed to be used. Because the digital objects are held by trusted parties in a secure environment and the overriding need is the ability to preserve the CAD records for use in manufacturing pieces accurately, authenticity in the archival sense of the concept was seen to be of less concern to the partners involved in this study.

As previously stated, this lack of a strong concept of authenticity does not mean that scientists are not concerned with the issue, merely that they express the concern in terms of accuracy, reliability or integrity. Case study 08, which looked at Mars Global Surveyor Data Records in the Planetary Data System said, "Project team members, PDS managers and engineers and other Planetary Scientists do not traditionally use the term authentic to characterize the data products that they create, maintain and use. They are concerned that the data records are complete, reliable, accurate, and that the integrity of the data record is assured."[80] To this end, there are data processing plans, manuals, specifications and workbooks

[78] Richard Pearce-Moses, Erin O'Meara and Randy Preston (2004), "InterPARES 2 Project - Case Study 14 Final Report: Archaeological Records in a Geographical Information System: Research in the American Southwest," 29. Available at http://www.interpares.org/display_file.cfm?doc=ip2_c14_final_report.pdf.

[79] Kenneth Hawkins (2006), "InterPARES 2 Project - Case Study 19 Final Report: Preservation and Authentication of Electronic Engineering and Manufacturing Records," 4. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs19_final_report.pdf.

[80] William Underwood (2005), "InterPARES 2 Project - Case Study 08 Final Report: Mars Global Surveyor Data Records in the Planetary Data System," 22. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs08_final_report.pdf.

to guide processing, transferring and data preparation. Further, the data are peer reviewed for accuracy and reliability and are validated through a system that also conducts checksums.

Integrity is particularly important in cases where data have passed through several processes after first being received. Metadata that support integrity are sometimes referred to here as lineage data; that is, a record of the stages through which the data have passed. Case study 06, Cybercartographic Atlas of Antarctica (CAA), is instructive on this point. To ensure reliability, authenticity and accuracy of the digital entities and documentation in the CAA, data are acquired from authoritative sources and are peer-reviewed. Each would have been assessed against the Elements of Spatial Data Quality, which include:

- lineage
- positional accuracy
- attribute/thematic accuracy
- completeness
- logical consistency
- semantic accuracy
- temporal information[81]

Authenticity in geography is captured in standard metadata as data lineage. Lineage, a mandatory metadata element, includes the history of a geographical dataset. Key elements in the metadata identify characteristics such as scale, accuracy, age and limitations on use. Within the geomatics profession, certain data management practices have also been adopted (e.g., inclusion of source data, documentation of source data rendered and how these data have been modified). The reputation of the institution or scientist is also a factor; thus, the CAA relies on the professional practices and authority of the institutions from which data are derived, and adheres to cartographic professional practices to choose the right level of data accuracy and to select cartographers for the right representation, a process that is very much reliant on metadata and professional practices. An editorial group reviews the content of the CAA to ensure thematic accuracy. In addition, the CAA production environment is protected by security measures such as physical security, password protection and careful control of access depending on type of user.

Case study 26 (MOST Satellite Mission) reiterates the concept that robustness in a dataset can replace reliability. Raw data received from the satellite will sometimes be corrupted by technical failures. However, by processing large collections of data, the scientists are able to deal with this problem. During the processing, the presence of the errors is diluted and the end results, called "reductions," are not affected. "Whether or not these false or corrupt data are included in the calculations for the reduction, does not affect the reliability of the outcome."[82]

The MOST scientists consider data that they receive "authentic" if there is no indication that the data received differ from the data recorded by the satellite instruments. The data are only accurate, however, to the extent that they truly represent the physical phenomena being observed, within the capability of the instruments. The researchers generate other data algorithmically from the original data and consider the derived data accurate to the degree that the algorithm transforms all the original data as expected.

General study 10 (Preservation Practices of Scientific Data Portals) was undertaken to collect information about the actual practices, standards and protocols currently used by broadly defined existing data services, archives, repositories, portals or catalogues in the sciences. Although the

[81] Lauriault and Hackett, "Case Study 06 Final Report," op. cit., 19.
[82] Bart Ballaux (2005), "InterPARES 2 Project - Case Study 26 Final Report: MOST Satellite Mission: Preservation of Space Telescope Data," 13. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs26_final_report.pdf.

sample size from each scientific discipline is small, thus limiting cross-disciplinary analysis, the study does provide a deeper understanding of practices in the natural and physical sciences as these pertain to portals, selected case studies and their associated data, as well as an exploratory review that considers the importance of issues such as accuracy, reliability and authenticity in the management of scientific data exchanged through portals. The findings of general study 10 provide further evidence of the appreciation among scientists of the concept of authenticity, despite their general lack of familiarity with the term as it is used in an archival context.[83] Among the thirty-two scientific data portals surveyed in this study, the term "authentication" is often used, and many of the qualities of authenticity (as related to the concept in an archival context) are discussed despite the fact that the term "authenticity" is never used.

Observations derived from these few case studies suggest that accuracy is associated with the risk of having inaccurate data: the more legal requirements there are, the more rigorous are the quality checks. Also, the more automated the process is, the more technical the checksums are and the more reliant the creators are on the technical systems in place and the less reliant on human checks: this is the case with the NASA Mars Surveyor Data, the Engineering Drawing study and the MOST satellite data. Professional practice, however, is very important in the Cybercartographic and the Archaeology case studies, as is a reliance on the trust associated with the integrity and authority of external data providers.

Conclusions

There is no question that the concepts of authenticity, accuracy and reliability are important in the preservation of scientific records of science. Since sound scientific research is dependent on the accuracy of data gleaned from scientific experiments, it is logical that the concept of accuracy looms larger than the concepts of authenticity and reliability. Questions about the accuracy of the data maintained over time are not dissimilar to questions relating to the authenticity of records maintained over time. This is clearly evident in the cases where satellite data have been used as evidence in legal proceedings, as well as in the case of laboratory notebooks and in many of the scientific data portals surveyed in general study 10. Concerns for data lineage can be seen as analogous to archival concerns over provenance and the chain of custody, and the recognition that reliable datasets are connected to authoritative data collectors has echoes of archival concerns for the authority of records creators. The differences in scientific use of these three concepts is more one of emphasis reflecting the particular concerns of scientists, but there is no evidence of real disregard for the concepts of authenticity, accuracy and reliability as viewed from the archival perspective.

Another important finding is the relative importance that scientists place on the content of a record in terms of its data quality (i.e., the accuracy of its content) when appraising its long-term value, something that archivists have hitherto generally considered irrelevant when conducting appraisals. In fact, many scientists, especially those in geomatics, argue that the data quality of a record should be an important factor in the decision of what scientific data to preserve and that archivists must, therefore, consider data quality in their appraisals if they are to acquire data from the sciences. To this end, the fact that dataset users are expected to recognize that the analysis and interpretation of a dataset requires discipline-specific background knowledge and expertise suggests that archivists will fare better at archiving specific types of scientific data if they

[83] Tracey P. Lauriault and Barbara L. Craig (2007), "InterPARES 2 Project - General Study 10 Final Report: Preservation Practices of Scientific Data Portals." Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs10_final_report.zip.

collaborate with scientists and specialists in the field. Alternatively, archivists can trust that either the scientists or the bodies managing the data will have already appraised the data in their custody; in which case, the archivists can instead work with scientists and their related institutions to add specific archiving practices into the data creation, management and preservation processes.

Compounding things further is the often disparate ways that the term "record" is understood and used by archivists and scientists. For many scientists, record is synonymous with data, databases and related information; entities that in an archival context are not generally considered records, except in very special and limited circumstances. As the Focus 2 research suggests, this is not, to a large degree, simply a matter of semantics; rather, it is a fundamental difference in perspective between scientists (creators) and archivists (preservers), exacerbated by the emergence in all disciplines of often highly ephemeral interactive and/or dynamic information that exists only in digital form. More importantly, it appears that the nature of the "record" within the digital environment may be changing dramatically. If so, traditional archival science will have to adapt to these changes in both theoretical and practical terms if it is to preserve this new information environment in the archives of the twenty first century.

In summary, although scientists indeed do recognize the importance of maintaining the "records" of scientific work in authentic, accurate and reliable form, it is important for both communities to be sensitive to differences in terminology usage as well as to fundamental conceptual differences regarding the very essence of what constitutes a record in both disciplines so they can work together to meet a common interest in long-term preservation.

Focus 1 – the Arts

Scope of the research

Although there is no lack of interesting questions about digital preservation in the sciences and government, the artistic creative activities contemplated by Focus 1 are so multifarious as to call into question some fundamental assumptions upon which the InterPARES Project was founded. Could such diverse digital objects be compared at all? Do artists working in what were historically different media share any common conceptions about preservation issues? Could the qualities of records identified by archival science have any consistent meaning to such diverse creators?

To come to grips with the theoretical and historical aspects of these issues, Domain 2 researchers combed the disciplinary literature for citations and explanations of concepts related to the identity, nature and preservation of artworks. Annotated bibliographies for music, dance, photography, moving images, sound recording, visual art, electronic literature, theatre and architecture were compiled, posted on the Project's Web site[84] and incorporated into the bibliographic database. Other bibliographic references were gathered from the case study reports.

Owing to certain weaknesses in the way the bibliographies were constructed, they served as research tools, not as final products of the Domain. InterPARES 2 did not have deep expertise in all disciplines, so some searches and annotations were more exhaustive and penetrating than others, and some bibliographies include items that are only tangentially relevant to the research questions of the Project. Also, some disciplines were quicker than others to adopt digital media and to realize the challenges that such a move posed for preservation, so many references deal

[84] See http://www.interpares.org/ip2/ip2_documents.cfm?cat=biblio.

with the concept of authenticity only in non-electronic media. Finally, since much of this work was conducted in the early stages of the InterPARES 2 Project (2002-2003), it does not reflect the most recent disciplinary thinking. To cite only a few examples, the bibliographies do not include major initiatives in electronic literature,[85] some significant developments in the philosophy of aesthetics[86] or descriptions of important preservation projects by the Variable Media Network and MUSTICA investigators in 2004. This report attempts to incorporate insights from such recent research.

Nevertheless, the bibliographies provided a good starting point for an historically informed analysis of the concepts. This was conducted by Domain 2 scholars of the arts and presented in several research papers. Their analysis is summarized below.

Grounding this conceptual survey was the work of the case studies researchers who investigated how selected creators thought about issues of identity, integrity and preservation in the context of specific works of art. Seven case studies treated the works of eleven creators, covering music, dance, theatre, moving images, interactive media installation and online publication. Although these disciplines have historically been fairly distinct, the Domain 2 researchers found that digital technology has fostered much interdisciplinary collaboration. The sharing of technology among disciplines has also helped reconcile different disciplinary conceptions about the nature of art and what needs to be preserved.

Conceptual analysis: authenticity, accuracy and reliability in the literature of the arts

Art practice and theory rarely concern themselves with by-products. To be sure, artists create (then often neglect) documents pertaining to their creative activities. Those documents that represent the transactional relations of artists and their patrons, such as commissions, contracts and correspondence, do not differ substantively from similarly functional documents of any other creator to the extent that their form and content are governed by the legal contexts in which they arise. But artists also generate other by-products that have no equivalent in the records of business, government or science. For example, most artists make and keep sketches—collections of ideas (sometimes fragmentary but sometimes even apparently complete works)—that are never published as final products. From an artistic sketch, unlike from a draft of a legal record, one might not determine the form of the work(s) that will result. This is partly because many sketches are merely components, not yet integrated into a whole. But the deeper reason is that artworks themselves are so different from records. They may have little to do with facts. They are complete and effective simply when the creator finishes them, not because they instantiate a fixed form or result from following a fixed procedure. Many are ephemeral, constituted essentially as experiences rather than as concrete documents or objects. And many are interactive, with their content and form determined partly by input from agents outside the artist's control.

Not surprisingly, then, it is difficult for artists to relate archival conceptions of a record to their documents, digital or otherwise. They apply the terms "authenticity" and "accuracy" to final products instead—to the objects or experiences that are the focus of aesthetic appreciation.

[85] See, for example, Nick Montfort and Noah Wardrip-Fruin (2004), "Acid-Free Bits: Recommendations for Long-Lasting Electronic Literature," Version 1.0, June 14, 2004. The Electronic Literature Organization. Available at http://www.eliterature.org/pad/afb.html; and Alan Liu et al. (2005), "Born-Again Bits: A Framework for Migrating Electronic Literature," Version 1.1, August 5, 2005. The Electronic Literature Organization. Available at http://www.eliterature.org/pad/bab.html.

[86] Davies, *Art as Performance*, op. cit.

Certainly, it is important to know that an artwork is what it purports to be, and has not been forged, tampered with or otherwise corrupted; the identity and integrity of artworks are important historically and culturally and may affect their financial value.

For most artists and audiences of art, the word "authentic" carries a primary sense of "original." In diplomatics, an original is a record that is primitive,[87] complete and effective. However what constitutes an original artwork varies considerably with the diverse conceptions of the nature of art.

The notion of authenticity as originality is most straightforward, and conforms most clearly to diplomatic conceptions, for a "singular" artwork[88] that is a (relatively durable) physical object. Artists may date and sign these objects, and some artists include elements (a special symbol, or a self-portrait in a crowd scene) that brand the work with their identity. Some of these elements can be understood as intended to provide the work with authenticity by establishing its origin with the artist. However, since most such elements are easy to forge, the originality of a singular artwork is best established "by a complete and dependable record"[89] of where the object has been since it left the artist's hands. This record may also include information about alterations or "restorations" made to the object, thus addressing its integrity as well as its identity. In contrast to administrative records, whose originality can be determined by observing whether they manifest all the necessary elements of the documentary form that defines them, artworks need not conform to a pre-established form. Thus, provenance is the principal testament of originality; that is, of authenticity. In this connection, some of the artistic literature also uses the term "authentic" interchangeably with "genuine." "Genuineness is based on and reflects a direct causal relation to the artist."[90]

If no such record of provenance exists—as is frequently the case—the originality of an object can be judged by "expert . . . comparison with works already accepted and works already rejected as . . . by the same artist."[91] In effect, this judgment is an "authentication" of a work, like the authentication of a record—the declaration of its authenticity at a specific point in time by a juridical person entrusted with the authority to make such declaration—that is necessary when its authenticity cannot otherwise be presumed. An expert's examination of the materials of the work is analogous to a diplomatic analysis of the extrinsic elements of a document's form. Expert authentication may also consider the structure of the work—the relations of the parts to each other and their function in the whole—which is analogous to analyzing the intrinsic documentary form of records. But artworks, even in the same genre by the same artist, may differ widely from each other in these respects, and the many examples of expert-deceiving forgery teach us that the criteria for authentication based on intellectual structure are, at best, provisional.

Conceptions of authenticity become more complicated for "multiple" artworks, which can occur in different places at the same time; for example, literature, prints, music, films, dances and installation art. The authoritativeness and effectiveness of any occurrence of such a work do not depend on its "primitiveness," but the words "original" and "authentic" are still used to refer

[87] That is, the first complete and effective instantiation of the record.
[88] "Singular artworks are unique, occurring at only one place at a time. Paintings, collages, carved sculptures, and Polaroids are typical examples of singular works" (Guy Rohrbaugh, "Ontology of Art," in *The Routledge Companion to Aesthetics.* 2nd ed. B. Gaut and D.M. Lopes, eds. (New York: Routledge, 2005), 242. Online reprint available at http://web.mac.com/rohrbaugh/iWeb/Site/Philosophy_files/encyclopedia3.pdf).
[89] Nelson Goodman (1996), "Authenticity," in *The Dictionary of Art*, Jane Turner, ed. (New York: Grove, 1996), 834.
[90] Jerrold Levinson (1990), "Autographic and Allographic Art Revisited," in *Music, Art, and Metaphysics: Essays in Philosophical Aesthetics* (Ithaca: Cornell University Press, 1990), 106.
[91] Goodman, "Authenticity," op. cit.

to the link between it and its creator. For example, to discredit unauthorized circulation of his novel, *Ulysses*, James Joyce provided the authorized publisher with a letter certifying that a particular edition—an unspecified number of physical objects—would be "the only authentic one."[92] For sound recordings that are cloned for publication, the term "master" is used instead of "original" to designate the authoritative source. Here "authenticity" is also used simply to indicate how exactly the copies reproduce the aural experience of the original recording, without regard to the original's status as a record.

Many of these multiple artworks result from executing instructions with specific instruments. The instructions can be executed with the instruments at different places or times, producing multiple objects or experiences that nevertheless all arise from the same procedure. For example, oily ink is rolled over a grease-pencil drawing on a moistened limestone plate, which is then pressed against paper; with each pressing, an instance of the work—a lithographic print—results. Moreover, some types of multiple works (music, dance, theatre) are created for performance: they are temporal experiences resulting from the actions of "interpreters" who execute the instructions with the specific instruments but who are also allowed by convention to add other information. In such contexts, the word "authentic" is used to indicate the causal link of the instructions and instruments to the creator. For example, a study of the emulation of an interactive video artwork, *Erl King*, observes that "the original [computer] code was written by the artists and their collaborators, and was therefore deemed critical to the authenticity of the work."[93]

For performed artworks, however, it is important to distinguish the authenticity of the documents conveying the instructions from what is called the authenticity of the performance. The latter entails notions of accuracy: for example, a musical performance is "accurate" to the degree that it realizes all the instructions in the score. Note, however, that scores do not make explicit all information necessary for accurate performance; performers must also adhere to implicit conventions of "performance practice," specific to the composer's time and place, that may modify the meanings of the symbols on the score. Some theorists call such accurate performance "authentic"—"a performance that reproduces all that is constitutive of the work's individuality."[94] Since the accurate and conventional execution of instructions produces an authentic instance of a multiple work, the notion of authenticity is detached from the property of primitiveness that characterizes the archival conception of originality.

Performance authenticity may be regarded as a matter of degree: to the extent that a performance is accurate as described above, it is "authentic." This contrasts with the meaning in diplomatics, in which only reliability is a question of degree, while authenticity is an absolute. Indeed, one might question whether performance authenticity has anything to do with preservation at all, since performances are ephemeral, not fixed. At least one can say, however, that performance authenticity is only possible if instructions and instruments (and knowledge of interpretative conventions) are authentically preserved.

Two further, opposing senses of "authenticity" stem from two complementary purposes of art, evident in the following statement: photographs "are authentic to the extent that they do

[92] Letter excerpt from James Joyce to Bennett Cerf (April 2, 1932), cited in Robert Spoo (1998), "Copyright Protectionism and Its Discontents: The Case of James Joyce's *Ulysses* in America," *Yale Law Journal* 108(3): 659.
[93] Caitlin Jones (2004), "Does Hardware Dictate Meaning? Three Variable Media Conservation Case Studies," *horizon*^{0} 18(2). Available at http://www.horizonzero.ca/textsite/ghost.php?is=18&file=6&tlang=0.
[94] Stephen Davies (1991), "The Ontology of Musical Works and the Authenticity of their Performance," *Nous* 25: 21–41. Reprinted in *Themes in the Philosophy of Music* (New York: Oxford University Press, 2003), 74.

justice to the facts of reality, and they are authentic in quite another sense by expressing the qualities of human experience by any means suitable to that purpose."[95]

The first meaning is verisimilitude. Viewers of a photograph may expect that the more fully it accommodates the detail, tonal range and perspective that they would perceive in the real object—the more the image is seen as the real thing is seen—the more truthful it is.[96] In this context, then, authenticity signifies accuracy—the quality of a work that facilitates the viewer seeing the photographed subject as if seeing the actual subject. The verisimilitude of a work may also be attributed to the way the work was made; for example, the strictures on props, sets and camera technique promulgated by the Danish filmmakers' collective Dogme95, whose "supreme goal is to force the truth out of . . . characters and settings."[97] Such emphasis on procedure recalls archival conceptions of reliability, which a record possesses if it is capable of standing "for the fact it is about,"[98] because it is authored by a competent person, created according to a controlled procedure, and complete in its form.

This idea of accuracy (conformance to perception) differs both from accuracy of performance—how exactly instructions are executed—and from scientific notions of accuracy, neither of which refer to how a subject is perceived. Also, it is undercut by many artists' realization that any record involves innumerable subjective decisions. In the words of the photographer Richard Avedon, "A portrait is not a likeness. The moment an emotion or fact is transformed into a photograph it is no longer a fact but an opinion. There is no such thing as inaccuracy in a photograph. All photographs are accurate. None of them is the truth."[99] Similarly, theorists demonstrate "the unreliability of the photograph as a record, and how much a construction it is . . . The photograph fails as a fact."[100]

The opposing meaning of authenticity reflects the belief that the primary purpose of art is to represent subjective experiences, not facts, an attitude that understandably prevails in such non-representational arts as music. In this context, authenticity denotes the degree to which an artwork manifests the individuality and essence of its creator or of the culture in which it was created. The term is used by critics "to bestow integrity, or its lack, on a performer, such that an 'authentic' performer exhibits realism, lack of pretence, or the like."[101] It might be said in such cases that the artist is the artwork, unmediated by any records. The 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. Such "personal" authenticity may work against verisimilitude,[102] as evidenced by Dogme95's prohibition on crediting a film's director.[103]

---

[95] Rudolph Arnheim (1993), "The Two Authenticities of the Photographic Media," *Journal of Aesthetics and Art Criticism* 51(4): 537.
[96] Jerry L. Thompson (2002), "Truth and Photography," *Yale Review* 90(1): 25–53.
[97] Lars von Trier and Thomas Vinterberg (1995), "The Vow of Chastity." Available at http://www.dogme95.dk/the_vow/vow.html.
[98] From the definition for "reliability" from the InterPARES 2 Terminology Database. Available at http://www.interpares.org/ip2/ip2_terminology_db.cfm.
[99] Richard Avedon, Foreword to *In the American West 1979-1984* (New York: Harry N. Abrams, 1985). Available at http://www.richardavedon.com/#mi=1&pt=0&pi=11019&p=-1&at=-1.
[100] Aphrodite Désirée Navab (2003), Review of *Transforming Images: How Photography Complicates the Picture*, by Barbara E. Savedoff. *Journal of Aesthetic Education* 37(2): 114–121.
[101] Allan Moore (2002), "Authenticity as Authentication," *Popular Music* 21(2): 210. Available at http://journals.cambridge.org/article_S0261143002002131.
[102] Peter Kivy. *Authenticities: Philosophical Reflections on Musical Performance* (Ithaca and London: Cornell University Press, 1995).
[103] von Trier and Vinterberg, "The Vow of Chastity," op. cit., vow no. 10.

However, the notion of authorship, and its role in establishing authenticity, is problematized by many works that incorporate mass-produced materials or that combine original work with excerpts from other artworks. The digital representation of sound and visuals greatly facilitates such appropriation. The aesthetic effect of such works is attributable to multiple authors, some of whom may not have intended it, and conflicts over copyright inevitably follow. "Questions about display and preservation require an interpretation of exactly what constitutes the work and who is authorized to make decisions that will shape how it is received."[104]

On the basis of this review of the literature, it is apparent that those who wish to preserve authentic artworks should be aware of the web of overlapping and sometimes contradictory meanings around the concepts of authenticity, accuracy and reliability. The terms are applied more frequently to final products than to the by-products of artistic creation. Authenticity refers to different properties depending upon what kind of artwork is being referenced. It is sometimes equated with accuracy. Reliability is almost never mentioned, although it is implicit in the relatively few instances of art-making procedures.

Little of the literature deals more than superficially with the problems of preserving digital artworks. For example, the Domain 2 researchers found numerous discussions of how the verisimilitude of photography is compromised by editing techniques, but the literature generally fails to address the practical aspects of how to create and manage digital images as reliable records and preserve their authenticity over the long term. However, the distinction between singular and multiple artworks makes an interesting and suggestive parallel to the distinction between fixed, physical records and the kind of ephemeral displays of information that are constituted by the execution of instructions in a computer system.[105] To understand this parallel more fully, and to investigate how problems of identity and integrity can affect the possibility of preserving digital objects, it is necessary to review InterPARES 2's analysis of actual digital art.

Authenticity, accuracy and reliability in the artistic sector case and general studies

Domain 2 had an especially rich source of information from the case studies in the creative and performing arts. Completed studies are listed below, tagged with the code assigned to them by the Project.

- CS01 Arbo Cyber, théâtre (?)
- CS02 Performance Artist Stelarc
- CS03 *Horizon Zero*/Zero Horizon Online Magazine and Media Database
- CS09(1) Digital Moving Images: Altair4 di Roma. A Multimedia Archaeological Project: *The House of Julius Polybius*
- CS09(2) Digital Moving Images: National Film Board of Canada
- CS09(3) Digital Moving Images: Commercial Film Studio
- CS09(4) Digital Moving Images: WGBH Boston
- CS10 *The Danube Exodus*: Interactive Multimedia Piece
- CS13 *Obsessed Again…*
- CS15 *Waking Dream*

[104] Martha Buskirk. *The Contingent Object of Contemporary Art* (Cambridge, MA: MIT Press, 2005), 23.
[105] A similar idea, without reference to the arts, is suggested in Helen Heslop, Simon Davis and Andrew Wilson (2002), "An Approach to the Preservation of Digital Records," National Archives of Australia. Available at http://www.naa.gov.au/Images/An-approach-Green-Paper_tcm2-888.pdf.

Various perspectives on each case are posted on the InterPARES 2 Web site, including: (1) the proposal; (2) a final report that answers the 23 questions of the research instrument; (3) a characterization of the creation of digital objects; (4) an activity model (for selected case studies only); and (5) a diplomatic analysis of the digital objects, attempting to identify the presence of records as those were defined by InterPARES 1. Although the case study characterizations, activity models and diplomatic analyses are especially pertinent to the research questions of the Project's other Domains, they also cast light on conceptions of authenticity, reliability and accuracy in interactive and dynamic systems of the creative and performing arts. So do the answers to some of the twenty-three questions, which the Domain 2 researchers helped design for this purpose.

In addition to the case studies, three general studies cast further light on the research questions. Two Web-based surveys solicited the comments of composers (general study 04) and of photographers (general study 07) on the nature of the digital objects they create, and the problems they have encountered with maintenance and preservation. Also, the MUSTICA project (general study 03), a collaboration of InterPARES researchers with researchers in French music-composition studios, yielded valuable insights from institutions that have made the maintenance of interactive digital music a high priority.[106]

Consistent with the results of the conceptual analysis, case studies researchers found that artists did not always distinguish between products and by-products (that is, between publications and records) when thinking about authenticity, reliability and accuracy. In case study 03 (*HorizonZero*) and case study 09(1) (Altair4 di Roma), the final products were posted to servers, or distributed on disks, whereby they lost their archival bond to the objects that were created in the course of producing them. Artists are more concerned about preserving the final products than the by-products, although they recognize the necessity of the latter to the former. Indeed, it seems sensible to consider how to preserve both, especially since they involve many of the same issues of technological context, identity and integrity.

In any case, most of the artists studied by InterPARES 2 understand authenticity, first and foremost, to denote the causal link between them and the products or by-products of their activities. For example, to Stelarc (case study 02), a work is authentic if its content is his; any original performance (by him) is authentic but re-creations of the same actions by others would not be.[107] Similarly, the creators of the documentary Web site studied in case study 01 (Arbo) believe that authenticity is guaranteed if the artists who worked on the video and sound recordings during the original performances are the same who adapt them for the site. "'Authenticity' is maintained by the artist's constant presence."[108]

Generally, this sort of authenticity is assumed to be ensured by the creators' control over the creation and organization of their digital objects, and by their marking the identity of those objects with metadata. For example, in the interactive multimedia installation of case study 10

[106] See Jennifer Douglas (2006), "InterPARES 2 Project - General Study 03 Final Report: Preserving Interactive Digital Music - The MUSTICA Initiative." Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs03_final_report.pdf; John Roeder (2006), "Authenticity of Digital Music: Key Insights from Interviews in the MUSTICA Project," version 2. Available at http://www.interpares.org/display_file.cfm?doc=ip2_gs03_summary_report_ROEDER_v2.pdf; and Bruno Bachimont et al. (2003), "Preserving Interactive Digital Music: A Report on the MUSTICA Research Initiative," in *Proceedings of the Third International Conference on WEB Delivering of Music (WEDELMUSIC 2003)*, 15-17 September 2003, Leeds, UK (Washington, D.C.: IEEE Computer Society Press, 2003). Available at http://polaris.gseis.ucla.edu/blanchette/papers/wedelmusic.pdf.

[107] Henry Daniel and Cara Payne (2004), "InterPARES 2 Project - Case Study 02 Final Report: Performance Artist Stelarc." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs02_final_report.pdf.

[108] Martine Cardin (2004), "InterPARES 2 Project - Case Study 01 Final Report: Arbo Cyber, théâter (?)," 28. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs01_final_report_english.pdf.

(*The Danube Exodus*), authenticity, quality and reliability are guaranteed by the authors being able to oversee and control the publication or finalization, and then "stamp" the work with credits and copyright statements. It lasts only as long as the artists exercise stewardship over the product.[109] In strongly market-driven projects with rapidly changing tools, such as the commercial animation studio of case study 09(3)[110] and the contract-multimedia production company of case study 09(1),[111] where the incentive to preserve is almost nil, identity and integrity are ensured by restricting access to the creation or alteration of digital objects and by marking the objects with version numbers.

In auteur-driven projects, whose creators are concerned with their personal legacy, authenticity means that any supposed instance of the work appears the same as the original, according to the judgment of the creator. This recalls the academic definitions of "authentic performance" as one that accurately presents all the work's essential features. The importance of accuracy to the creators in case study 09(2) (National Film Board of Canada) is indicated by the "nervous breakdowns" some animators are reported to have suffered when confronted with versions of their work that had been degraded by migration to lower-quality display systems.[112] Some creators understood the notion of "reliability" in the same sense. In case study 01, the artists understood it to mean how well a video recording represented their conception of the work, with no regard to how it was made. They presume their records to be "reliable" because they believe the records are impossible to fake or, at least, that no one would want to do so.[113]

Often, in the most controlled contexts, "authenticity" is conceived purely as the "usability" of digital objects; that is, whether the objects (by-products) will function as expected in the software that is used to generate the final product.[114] In other words, authenticity is conflated with a kind of reliability. In collaborative efforts, the term "reliability" was not used independently at all,[115] or it was understood as a (desirable) characteristic of the systems that are displaying the documents, not as a (desirable) characteristic of the documents themselves. A reliable system, in this sense, displays the same information the same way every time it is called up.[116]

In a somewhat different sense, reliability is a concern in works for performance, such as is noted in case studies 13 (*Obsessed Again…*) and 15 (*Waking Dream*), whose creators collaborate with performers, revising the work's instructions until they communicate the intention

[109] Sally Hubbard (2006), "InterPARES 2 Project - Case Study 10 Final Report: *The Danube Exodus*," 7–8. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs10_final_report.pdf.
[110] James Turner et al. (2004), "InterPARES 2 Project - Case Study 09(3) Final Report: Digital Moving Images - Commercial Film Studio." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs09-3_final_report.pdf.
[111] Isabella Orefice (2004), "InterPARES 2 Project - Case Study 09(1) Final Report: Digital Moving Images - Altair4 di Roma, A Multimedia Archaeological Project: The House of Julius Polybius." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs09-1_final_report.pdf.
[112] Andrew Rodger (2006), "InterPARES 2 Project - Case Study 09(2) Final Report: Digital Moving Images - National Film Board of Canada," 11. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs09-2_final_report.pdf.
[113] Cardin, "Case Study 01 Final Report," op. cit., 41.
[114] See Orefice, "Case Study 09(1) Final Report," op. cit., 5; and Turner et al., "Case Study 09(3) Final Report," op. cit., 9, 18.
[115] See Brent Lee (2004), "InterPARES 2 Project - Case Study 03 Final Report: HorizonZero/Zero Horizon Online Magazine and Media Database." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs03_final_report.pdf; Rodger, "Case Study 09(2) Final Report," op. cit.; Mary Ide (2005), "InterPARES 2 Project - Case Study 09(4) Final Report: Digital Moving Images - WGBH Boston." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs09-4_final_report.pdf; Hubbard, "Case Study 10 Final Report," op. cit.; J. Scott Amort (2004), "InterPARES 2 Project - Case Study 13 Final Report: *Obsessed Again....*" Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs13_final_report.pdf; Sydney Fels and Seth Dalby (2004), "InterPARES 2 Project - Case Study 15 Final Report: *Waking Dream*." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs15_final_report.pdf; and Nadine Hafner, Janine Johnston, Tracey Krause and Keum Hee Yu (2006), "InterPARES 2 Project - Case Study 22 Final Report: Electronic Café International (ECI)." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs22_final_report_DRAFT.pdf.
[116] Daniel and Payne, "Case Study 02 Final Report," op. cit., 4.

completely, and revising the digital instruments until they are adequate and capable of supporting repeated performances. However, both works considered in these two case studies were revised for each new performance, resulting in different versions of the work, so the digital objects were never fixed enough to apply the term.

It is interesting that diplomatic analysis of even the most interactive and dynamic artworks found the presence of records, or near-records, in the creators' systems, even when the creators did not conceive of their objects in that way. This can be explained by the fact that many artists working digitally save their files in some organized fashion, maintaining enough of a file-organization system to create an archival bond among the digital by-products of their activities Also, the interactive and dynamic features of the works are grounded in every case on fixed instructions and instruments. The work (the focus of appreciation) is interactive, but the records are not. Generally, however, these will not be reliable in the future, in the sense that they could be used to re-perform/re-generate the work, because they are tied to specific technical platforms and standards that change rapidly.[117]

The main preservation challenge, as many studies noted, is to preserve the technological context of these documents, or to find new technological contexts in which equivalent experiences can be generated. The question that hovered over all the case studies, therefore, was how equivalency; that is, accuracy, could be judged in the creator's absence. Here, the MUSTICA researchers' experiences are especially relevant. They all identify the necessity of preserving the instructions for producing, sequencing and processing sounds, and usually the sounds themselves, and they assert that their community uses a common "bedrock" of sound-processing procedures that should be migrated to any new technology. They also agree that a recording of the sound patterns does not preserve the work. No recording is "exact" or "precise," because it cannot manifest all the essential features of the work, because it records mistakes in performance, and because it cannot present the balance of sounds the composer has conceived for a live presentation of the music. Nonetheless, the MUSTICA interviewees regard recordings as essential to preservation, as the only substitute for the composer's authority after he or she has died.[118]

Conclusions and relevance of this analysis outside of the artistic sector

The various creative and performing arts converge in digital media works that combine physical objects, text, audio and moving and still visuals, all interacting with performers and audience. Preserving such "multiple" works means preserving the ability to perform (display) them. Not all artists embrace this conception or accept the limitations that digital media impose. Many artists are not concerned with preservation at all. But for those who are, the challenge is clear: creators need to take effort to specify and preserve the identity and integrity of the instructions and instruments, including their functionality, interoperability and accuracy of content, across technological change.

This requires an understanding of all objects and their relations, along with the interdependencies of authors, performers and technology. The modeling activities conducted by InterPARES 2 were helpful in exposing these. Considering the experience the Domain 2

[117] See, for example, Rodger, "Case Study 09(2) Final Report," op. cit., 16.

[118] The French researchers affiliated with the MUSTICA project later proposed a detailed typology of the musical works, investigated the suitability of various metadata standards and proposed some methods of preservation (see Xavier Sirven (2004), "Authenticité et accessibilité des archives électroniques - MUSTICA, Le cas de la création musicale numérique," Technical Report, Université Technologique de Compiègne. Available at http://polaris.gseis.ucla.edu/blanchette/papers/RapportSirven.pdf.

researchers had in recreating one of these works,[119] however, it is important to caution that there can be many subtle aspects to interactive systems that are only manifest when the creator evaluates the re-creation.

Artists' understanding of authenticity varies widely, and is often conflated with concepts more closely allied to reliability or accuracy. Nevertheless, many of them take at least some actions to identify the digital components of their works. Archival notions of authenticity are consistent with preservation intents and actions of the artists studied by the Domain 2 researchers. But these notions need to be nuanced in light of the disciplinary conceptions of authenticity and accuracy exposed in the Domain 2 analysis, as much in science as in the arts.

Multiple artworks, even those involving paper instructions and physical instruments, provide a model for how to regard the ephemeral "records" displayed by other digital information systems. These displays are performances of fixed instructions, using the instruments of the computer hardware.[120] Thus, methods for ensuring authenticity and reliability of multiple artworks can stand as a model for how those qualities can be preserved in digital record systems outside of the arts.

Relevance of the Benchmark Requirements of InterPARES 1

The focus of InterPARES 1 on digital records in administrative and legal systems, noted in the introduction, directed the derivation and content of the benchmark requirements it proposed. Since these documents are created in the context of well-defined procedures and function like paper documents with well-defined documentary forms, the benchmark requirements reflected long-established ideas about the authenticity of paper records. They assumed that recorded actions can be classified into types and that a record with a characteristic documentary form is associated with each type of action. Moreover, since InterPARES 1 focused on how to assess and maintain the authenticity of digital records once they become inactive and are selected for permanent preservation, it did not investigate how to create reliable digital records and maintain their authenticity during their active and semi-active life. That was the subject of a previous study, the "UBC Project,"[121] which was a collaboration between UBC researchers and the U.S. Department of Defense that produced the DoD Standard 5015.2 for recordkeeping systems.[122]

There are some difficulties, then, in applying the InterPARES 1 results to interactive and dynamic digital documents created by individuals or small collaborative groups in the arts and sciences. Although many such documents could be called "inactive," not all are records, and few of them (in particular, none in the case studies) have been selected for long-term preservation by an archival institution. This would suggest that the findings of the UBC Project might be more applicable, but the documents studied by InterPARES 2 also differ from those in the systems regulated by the DoD Standard. For instance, it is not clear how to classify the actions signified by the digital entities that are created as components of artworks, since the actions are steps in a generative process that may vary considerably from work-to-work and artist-to-artist. Concomitantly, it is not clear whether such various entities have any consistent documentary form that could be examined to determine whether they did participate in the creation of an artwork. Lastly, the UBC Project did not contemplate the special problems of interactive and

[119] See the section later in this report titled "A strategy for preventing technological obsolescence of an artistic work."
[120] Duranti and Thibodeau, "The Concept of Record," op. cit.
[121] See http://www.interpares.org/UBCProject/index.htm.
[122] Since the DoD Standard was adopted, it can be said that InterPARES-related guidelines have been validated by their regular use in some government recordkeeping activities.

dynamic systems. Informed by InterPARES 2 research, Duranti and Thibodeau's rethinking of the concept of a record deals with many of these theoretical issues.[123]

Nevertheless, some evidence for the relevance of the InterPARES 1 findings can be seen in the fact that even creators furthest from recordkeeping bureaucracy show an awareness of authenticity requirements in the way that they create and organize their digital objects. For instance, benchmark requirement A.1 asserts that authenticity can be presumed if certain identifying attributes are explicitly expressed and inextricably linked to every record. For the digital entities analyzed diplomatically by InterPARES 2, many of these attributes are at least implied, and often standard, as in some of the scientific datasets.[124] Even when the attributes are not explicit, it seems like a small step to include them (for instance, an historical trace of provenance) as part of the objects' metadata.[125]

Consider, moreover, that in several of the InterPARES 2 case studies the creators attempted to maintain their documents and encountered various difficulties. To the extent that those difficulties can be attributed to violations of the requirements proposed by the UBC Project and InterPARES 1, the requirements can be understood as relevant. For example, the Domain 2 researchers observed that various problems of archival bond (one of the attributes of record identity required to be explicit and linked by benchmark requirement A.1) can beset the digital entities associated with Web sites. In some cases, such as HorizonZero, they are not set aside in a recordkeeping system with other records with which they could form an archival bond.[126] In others, such as on the Legacoop of Bologna's site, the "entities on the Web site do not possess an archival bond beyond a chronological record of their posting."[127] Without these bonds, a creator may be able to maintain a publication as a final product, but the traces of its creation will be obscure. In contrast, the data files representing transactions with the Irish Revenue On-Line Service are structured to form natural aggregations "wrapped" together by addressee.[128]

Dynamic systems that draw information from constantly changing sources naturally run afoul of the benchmark requirements. If their displays are to serve as records of actions, the data they display must be fixed, or at least bounded,[129] and dated to enable redisplay. For example, in cases such as VanMap, the lack of date-stamping can prohibit preservation.[130] Another problem is exemplified by Stelarc's Web site, which the artist intends as a record of his work. One of its pages involves an interactive interface that simulates a performance of his work "Ping Body"[131] by requesting the "ping" (response) time from a server in Australia[132] to a randomly selected remote Web server. The returned value controls the motion of wire-frame body limbs displayed on the screen, simulating the actual performances, in which Stelarc's own limbs are controlled by electrical shocks proportionate to the ping values. However (at least in July 2006), the Australian

[123] Duranti and Thibodeau, "The Concept of Record," op. cit.
[124] See, for example, Underwood, "Case Study 08 Final Report," op. cit.; and Ballaux, "Case Study 26 Final Report," op. cit.
[125] For such a proposal for digital artworks, see Alena Williams, "Rhizome.org," in *Permanence Through Change: The Variable Media Approach*. Alan Depocas, Jon Ippolito, and Caitlin Jones, eds. (New York: Guggenheim Museum Publications, 2003), 39–41. Online reprint available at http://variablemedia.net/pdf/Permanence.pdf.
[126] Tracey Krause (2006), "InterPARES 2 Project - Case Study 03 Diplomatic Analysis: *HorizonZero*/Zero Horizon Online Magazine and Media Database." Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs03_diplomatic_analysis.pdf.
[127] Carolyn Petrie (2006), "InterPARES 2 Project - Case Study 25 Diplomatic Analysis: Legacoop of Bologna Web Site," 4. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs25_diplomatic_analysis.pdf.
[128] Tracey Krause (2005), "InterPARES 2 Project - Case Study 20 Diplomatic Analysis: Revenue On-Line Service (ROS)," 3. Available at http://www.interpares.org/display_file.cfm?doc=ip2_cs20_diplomatic_analysis.pdf.
[129] See discussion of the concept of "bounded variability" in Duranti and Thibodeau, "The Concept of Record," op. cit., 47–48.
[130] McLellan, "Case Study 24 Final Report," op. cit., 31.
[131] See http://www.stelarc.va.com.au/pingbody/ping.html.
[132] http://www.merlin.com.au.

server no longer responds to the request, so the interface no longer accurately simulates the performance.[133] Another case study that involves such dynamic documents is the Cybercartographic Atlas (case study 06), and the general study of scientific data portals (general study 10) reveals that many of them have analogous external dependencies.[134] None of the benchmark requirements directly addresses this situation. Stelarc's site itself has not undergone technical modification (benchmark requirement A.1). In some senses, the technological context has not changed (benchmark requirement A.4)—the Australian server still exists, and the simulation (a Shockwave movie) still runs. In another sense, however, it has been modified to the extent that the Australian server administrator has removed the software routines from which the simulation requests the ping values. This is a rather subtle change (indeed, Stelarc's Web site administrator has not noticed it) and demonstrates the need for careful analysis of the inputs to dynamic documents.

Of course, providing for changes in technological context, as demanded by benchmark requirement A.4, is the most pressing problem for preserving all sorts of digital systems. Most proposals for preservation in the literature deal principally with this issue, which is also considered in the following section of this report. In InterPARES 2's studies of artworks, such as *Obsessed Again...* (case study 13) and *Waking Dream* (case study 15), and for the musical works studied by MUSTICA (general study 03), the creators are not truly maintaining their original works, but are essentially creating new versions—that differ in essential ways from the originals—by rewriting software for the latest technologies. They have not specified their works in ways that minimize or eliminate dependence on custom, proprietary or obsolescent instruments. Analogously, in case study 19 (Preservation and Authentication of Electronic Engineering and Manufacturing Records) from the science focus, the methods that the creators experimented with to verify the identity and functionality of machine parts specified by CAD (computer-aided design) documents could not be successfully realized without dependence on a proprietary reasoning engine that could not itself be preserved.[135]

Thus, even though the concepts employed by InterPARES 1 and the UBC Project are not entirely adequate for the systems studied by InterPARES 2, the benchmark requirements seem relevant, because a failure to follow them prohibits preservation, and because efforts to preserve include some of the actions they specify. There are, however, indications that they are necessary, if perhaps not sufficient, so a more thorough review of this issue seems warranted.

Experience with a Possible Maintenance Strategy

Issues

In activity that produces records, the identity and integrity of the records are not in question as long as the creators are still actively referring to them, because records that the creator relies on in the usual and ordinary course of business are presumed authentic. In some cases, however, the digital objects that are the components of these records are set aside and left inactive long enough that technological change renders them unusable. In effect, even if their actual file structure and content may have been physically preserved, their integrity is undercut by the disappearance of the technological context needed to display them.

[133] There are other outdated/nonfunctional links, as well, that detract from the integrity of the site.
[134] See Lauriault and Craig (2007), "General Study 10 Final Report," op. cit.
[135] Hawkins, "Case Study 19 Final Report," op. cit., 8.

Technological obsolescence, as remarked above, is one of the primary concerns for creators, users and preservers of digital records. When such objects are merely backed up (which some confuse with "archiving"), a change of technological context is not evident. The loss of integrity will only be evident to the extent that records keepers and preservers monitor the authenticity of records when they are transmitted across space or time. What procedures of creation and transmission would ensure that these records will continue to be recognized as authentic?

Some proposed strategies, and their relative advantages and disadvantages, are summarized concisely by Heslop, Davis and Wilson.[136] These authors' suggestion—to require all digital components to be expressed in public-domain formats—is a good one, but does not address the special problems of custom-formatted entities like the ones often encountered in artistic activities.[137] It would be futile to insist, for example, that artists restrict their means of expression to the lowest-common-denominator formats. And the authors do not consider how or if such a strategy could maintain the interactive and dynamic attributes of records, for which there are no standard representations.

A strategy for preventing technological obsolescence of an artistic work

To consider more fully the problems of preserving documents with these special attributes, Domain 2 researchers attempted to resurrect a work that had already fallen victim to technological obsolescence: Keith Hamel's *Obsessed Again…* for bassoon and interactive electronics (1992), the subject of case study 13. The instructions and instruments specified originally by the composer are represented schematically in Figure 1. It was assumed that the musical score (the instructions for the bassoon, in portable document format) can be preserved, and that an accurate and reliable bassoon, microphone and amplification system will exist in the future. A recording of a performance of the work (in a format with freely available specifications, so presumably preservable) was also available.[138] However, the other instruments shown in the centre right of the figure are now obsolete. The sounds that the computer causes to be played during a performance—including their timing and their interaction with the sounds that the bassoonist plays—are encoded in the instructions symbolized as "code" in the figure. But the interactions are nowhere explicit; they can only be deduced by analyzing the code and listening to the recording. To be realized, they require a functioning software environment (the proprietary MAX 2.0 running on a proprietary operating system) to interpret them. Outside of that specific technological context they are inoperative, and it is difficult to discern what they are supposed to do without an intimate knowledge of the technical specifications of the hardware and of the syntax and semantics of MAX.

[136] Heslop et al., "An Approach to the Preservation of Digital Records," op. cit.
[137] See, for example, Nicola Bernardini and Alvise Vidolin (2005), "Sustainable Live Electroacoustic Music," *eContact!* 8(3). Available at http://cec.concordia.ca/econtact/8_3/bernardini_vidolin.html; and Joel Chadabe (2001), "Preserving Performances of Electronic Music," *Journal of New Music Research* 30(4): 303–305.
[138] For an analysis of the terminology used to characterize various levels of software "openness," see Evelyn Peters McLellan (2006), "InterPARES 2 Project - General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation." 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.

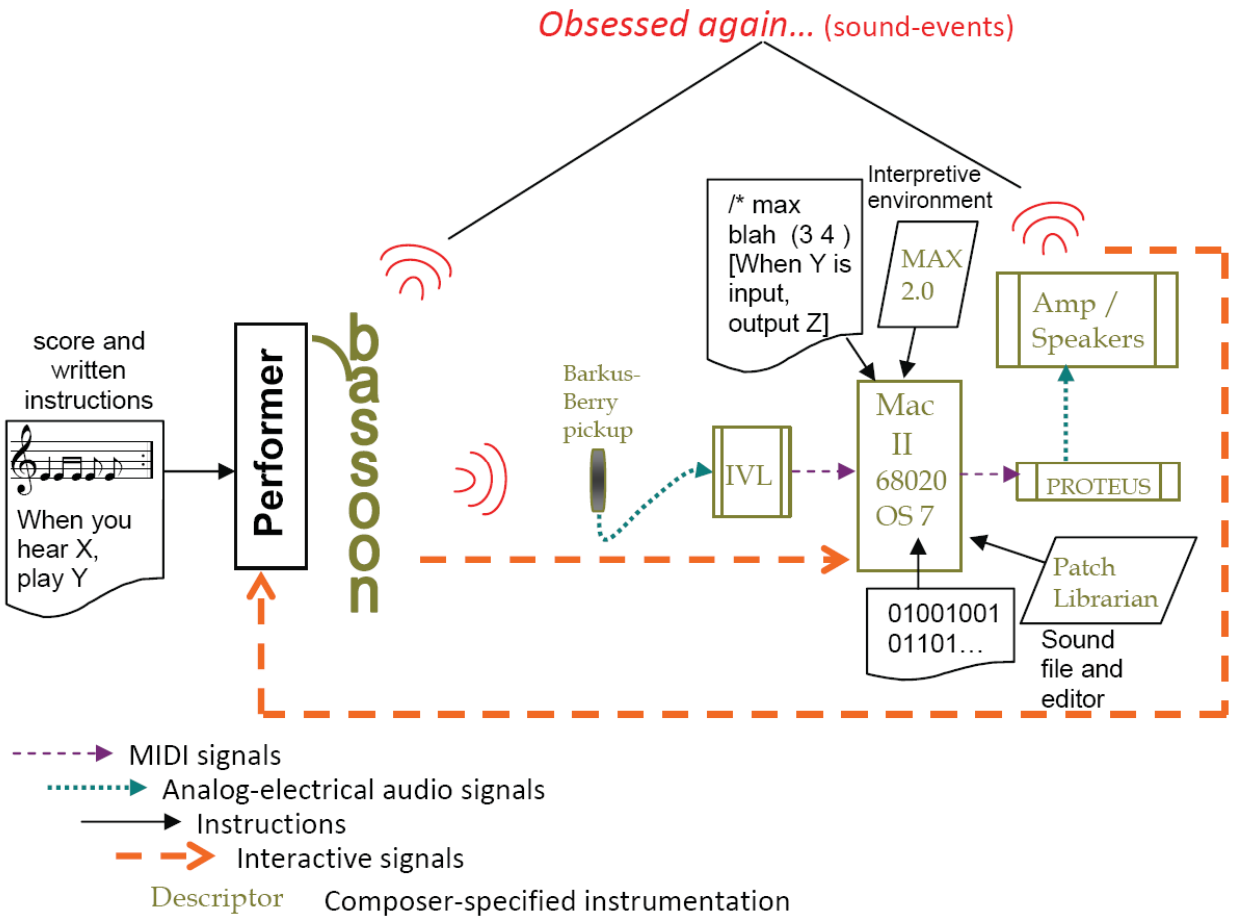

Figure 1. Schematic of Composer Instructions and Instrumentation Specifications for *Obsessed Again…*

The following paraphrased remark by one of the interviewees of the MUSTICA study summarizes the problem succinctly: 'The death of a patch [that is, the hardware instructions] means the death of the composition.' Ironically, at IRCAM, one of the institutions participating in MUSTICA, preservation efforts have produced meticulous documentation of how to perform works[139] but not enough information about content; thus, if hardware instructions no longer function, the integrity of the digital components is lost. Similar difficulties prevent the preservation of many interactive artworks. They have been articulated, for ephemeral art, by the Variable Media Initiative,[140] and the Electronic Literature Organization produced a substantial study of related issues in e-literature.[141] But the problems clearly extend to any system threatened by software or hardware obsolescence, including cases studied in the science focus (such as case study 19). A so-called "open" format can be proprietary and thus become obsolete if the proprietor ceases to support the format or asserts intellectual property rights that impede preservative transformations. Finally, even open, non-proprietary formats may become obsolete if future technology works differently than that of today.

[139] See, for example, Andrew Gerzso, "Performance Handbook: *Anthèmes 2* [by Pierre Boulez]," (Paris: IRCAM, 2005). Available at http://mustica.ircam.fr/mustica_1.2.0/rendu/pdf/output/Anthemes_2.pdf.
[140] See http://www.variablemedia.net.
[141] Liu et al., "Born-Again Bits," op. cit.

Resurrection of *Obsessed Again…* involved an exercise in controlled migration that simulated the transmission of records across space and time. One researcher, familiar with the technical details of the original instruments, translated the instructions from code into technologically neutral natural language stored in a word-processing document in non-proprietary format. Another researcher, armed only with these new instructions and the recording and with no other knowledge of the work or contact with the composer, wrote software that would control modern instruments to produce the same sounds and interactions that the original instructions and instruments did. Lastly, the Domain 2 researchers asked the composer, who was otherwise absent from the exercise, to judge the authenticity of a performance that employed the new instructions and instruments. By these means the researchers sought to establish a set of records and observe the effects of hardware and software evolution on them to determine whether it is possible to represent all that is essential to the work's identity in a technologically neutral way.

The responses of the composer confirmed the successes of this exercise while clarifying its limitations. He acknowledged that the machines and software in the new version interacted correctly with the bassoon's music, but he pointed out certain deficiencies that made the result somewhat different than he intended. That they affected authenticity was evident from his comment: "I like it, but it's not my piece." First, the sensitivity of the devices that receive input—in this case, the part of the system that detects the bassoon's sounds and translates them into digital inputs to the computer—were crucial to achieving the intended interactions. If the representation is too coarse-grained or fine-grained, the system may not respond when it should, or it may respond when it is not intended to. This sensitivity needs to be made explicit in the instructions for the artwork; it involves timing as well as other measurable aspects of the input. Also, the resolution of the output was crucial. In this case, the sounds that the electronic devices produced were all encoded in a proprietary format that could not be described in a technologically neutral way, the modern sound-producing devices could not be made to match those on the recording exactly, and the original instructions gave no indication of how accurately those sounds needed to be reproduced. Not surprisingly, the resulting sounds did not match the composer's intentions, and, in fact, this was the only reason he gave for not acknowledging the new performance as authentic.

For this work, it is not hard to imagine a solution. Recordings of the necessary sounds could be stored in a format with freely available specifications so that the migration would only involve reprogramming the interactions, which the researchers successfully did. And the very exercise confirms the intuition of the Variable Media researchers of how important it is to get the creator's feedback on attempted migration. In terms of the conceptual analysis above, it can be concluded that by making inauthentic performances one can discover how to provide instructions that can be preserved authentically and that can produce authentic performances. This supports the finding of Domain 3 that "preservation begins at creation."[142] Creators, 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.

[142] Domain 3 Task Force Report, 4. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_4_domain3_task_force.pdf.

Analogies to a mechanical engineering case[143]

This experience has an interesting parallel in a very different case study, that of the Preservation and Authentication of Electronic Engineering and Manufacturing Records (case study 19). This case study, in fact, was also an experiment, initiated by the records creator to find a method of preserving active records to meet the creator's needs. In this case, the creator's need is to be able to use the records for the same purpose for which they were created: to manufacture piece parts for physical equipment. The equipment is often maintained for decades after its manufacture. At any point over this time it may be necessary to produce replacement parts if an existing part is damaged or wears out. The replacement part must fit into the piece of equipment exactly as the original part did. Piece parts are manufactured according to specifications produced as computer-assisted design (CAD) records using computer-assisted manufacturing (CAM) records that control the processes executed by robotic machine tools to manufacture parts with the right size, shape and configuration.

CAD/CAM systems today are proprietary and subject to obsolescence. The experiment was designed to test whether the CAD records could be translated from their proprietary format into a persistent format and preserved for use in some future, unknown CAM system to produce identical replacement parts. The formats chosen for preservation were independent of any specific hardware or software, freely available, standard and self-describing.

The experiment consisted of the records creator producing persistent format versions of its original records, transmitting them to a trusted digital repository as a surrogate for an archives, retrieving them from the repository and determining whether the preserved records could be used to produce the piece parts they described. The experimental design intentionally included several potential points of failure: the translation from propriety into persistent formats, transmission of the persistent records to the surrogate archives, ingest into these "archives," preservation, retrieval and return of the records to the creator and their use in production of replacement parts.

In fact, the experiment encountered failure at the first point. Even though the records creator employed two different types of freely available, standardized, self-describing formats to capture the piece parts, the persistent format records were not adequate to enable manufacture of replacement parts. This failure obviously entails inability to use the persistent format records to produce replacement piece parts. The intermediate steps in the experiment were executed without problems.

Unlike the *Obsessed Again…* case, there was no element of subjective judgment in the determination that the CAM experiment failed. However, there are parallels between the situation in the arts and that in engineering: the lack of an adequate language for expressing the specifications or instructions in the original records in a preservable format that could be used to perform or produce something that satisfied the original intent. Note that both failures were in specific cases. They do not amount to a failure of transformation to persistent formats as a preservation methodology. Rather, they identify specific areas where additional efforts are required.

Connections to the goals of the Project

Both of these experiments highlight the importance of the revisionary conceptual work that was the principal activity of all of the research domains in InterPARES 2. Considering the

[143] Kenneth Thibodeau contributed the content of this subsection.

interactive and dynamic environments exemplified by the various case studies, researchers were led to propose expansions to the traditional conceptions of record and metadata. Records in such environments as *Obsessed Again…* and case study 19 often encompass discrete components distributed across systems, while their behaviour, operations and even their authority to reside within computing environments may depend on the messages or instructions their metadata communicate to those same environments.

The experiments also confirmed the need, suggested by the conceptual analysis, for expansions to the traditional conceptions of authenticity, reliability and accuracy. The attempted resurrection of *Obsessed Again…* suggested that, in principle, a performance of a born-digital composition could be authentically and accurately performed sometime later using newly supplied elements (new samples, etc.). Case study 19 actually proposed an expansion of the underlying basis of presumed authenticity, saying that it depended not only on reliably populated attributes evidencing identity and integrity, but also on the conduct of "proofs" involving the semantic relationships of those attributes within a domain-specific ontology.

Toward Guidelines for Creating and Maintaining Authentic and Reliable Digital Records

Although Domain 2's bibliographic research found many theoretical discussions of the challenges posed to the authenticity and reliability of digital objects, it also found, as noted by the study of the digital recordkeeping practices of photographers who operate in artistic, scientific and governmental environments, that "documentation of procedures to create and preserve [records] in the digital environment for the long term has been sparse."[144] Given the urgency of the preservation problems identified in the introduction to this report, it seemed imperative that InterPARES 2 issue guidelines to assist creators in creating and maintaining preservable digital materials, especially records. It fell to Domain 2 to produce the *Creator Guidelines*[145] a document designed to accompany the *Principles for Records Creators*[146] developed by the Policy Cross-domain.

During development of the *Guidelines*, it became evident, from the conceptual analysis, case studies and general studies of the Project, as well as from the experiments described in the previous section, that certain principles would need to guide the content, form and presentation of the guidelines:

- They should reflect the concepts and practice of archival science; for example, distinguishing backups or repositories from archives.
- They should specifically address records of interactive and dynamic systems. For example, it is not sufficient simply to require documents to be in a format that is non-proprietary or that has freely available specifications, because no freely available description standard for interactivity yet exists.
- They should avoid using the terms authenticity and reliability, while still clarifying what the records must have to be authentic and reliable. This is because Domain 2 found that

[144] Bushey and Braun, "General Study 07 Final Report," op. cit., 3.
[145] See Appendix 20. The Guidelines also are available in booklet form at http://www.interpares.org/display_file.cfm?doc=ip2(pub)creator_guidelines_booklet.pdf.
[146] See the *Policy Framework* in Appendix 19. Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_19.pdf.

these terms, although precisely defined in archival science, mean different things to different creators, and that (if they are used at all) they are often confused or conflated.
- They should be worded so to make it clear (if not simple) what is required to satisfy them, even to such disparate creators as artists, scientists and bureaucrats.
- They should reflect the finding that, for records to be preserved, information and processes must be incorporated into their creation that will allow their identity and integrity to be ascertained in the future.
- They should be consistent, as far as possible, with guidelines issued by professional organizations, curatorial institutions and standards organizations.
- They should facilitate respect for cultural differences, freedom of expression, freedom of inquiry and right to privacy.

The *Guidelines* were worked out through an iterative method. On the basis of bibliographic research that exposed previous attempts at guidelines, candidate guidelines were proposed and considered by the InterPARES 2 International Team, considering the principles articulated above. The results reflect a consensus of archival scholars, practicing archivists and specialists in the arts, science and government focuses. Although it is presumed these guidelines apply to a large class of record-making and recordkeeping activities, the InterPARES researchers do not claim that the guidelines exhaust all of the preservation-related issues and concerns that may be associated with, or impacted by, records creation and maintenance activities. Thus, although the requirements for record-making and recordkeeping derived from them seem necessary, it cannot be claimed that they are sufficient for all cases; only experience will tell.

Other products of InterPARES 2 are also intended to assist in the creation, maintenance and long-term preservation of authentic digital records. The Metadata and Archival Description Registry and Analysis System (MADRAS) supports and eases the tasks of identifying, registering, describing and evaluating existing standards for the intellectual control of records from the moment of their creation throughout their appraisal and preservation.[147] InterPARES has also produced frameworks for the development of policies, strategies and standards regarding creation, maintenance and preservation of digital records; one framework is for organizations creating digital materials, and the other is for archival institutions or programs.[148] The Project's two models of records preservation—one reflecting a record lifecycle point of view (Chain of Preservation Model) and the other reflecting a record continuum point of view (Business-driven Recordkeeping Model)—can help organizations clarify needed procedures and resources.[149] Finally, the Terminology Database, which defines the terms used in the InterPARES Project, also includes a comparison with terms in existing dictionaries of all disciplines involved in the Project, thus fostering communication among creators and preservers of our digital legacy.[150]

[147] MADRAS is discussed at length in 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.

[148] The context for this framework, known as 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*), is discussed in the Policy Cross-domain Task Force Report (available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_7_policy_task_force.pdf), while the framework itself is provided in Appendix 19, op. cit.

[149] Narratives for both models are provided in 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), while the model diagrams and definitions can be found in Appendices 14 (COP Model, available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_14.pdf) and 15 (BDR Model, available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_15.pdf). Both models are also available on the InterPARES Web site at http://www.interpares.org/ip2/ip2_models.cfm.

[150] More detailed description about the Terminology Database and each of its components is provided in the Terminology Cross-domain Task Force Report (available at http://www.interpares.org/display_file.cfm?doc=ip2_book_part_8_terminology_task_force.pdf, while the Database itself is available on the InterPARES Web site at http://www.interpares.org/ip2/ip2_terminology_db.cfm.

Appendix 12

Domain 2 Research Questions

- What does record reliability mean in the context of artistic, scientific and governmental activities? To what extent can the electronic records created in the course of each type of activity be considered reliable and why? What requirements on their form and controls on their creation would make us presume that they are reliable?

- What does record accuracy mean in the context of each activity? To what extent can the electronic records created in the course of each type of activity be considered accurate and why? What controls on their creation would make us presume that these records are accurate?

- What does authenticity mean in the context of each activity? To what extent is the definition of record authenticity adopted by InterPARES 1 relevant to the records resulting from each type of activity and from the use of increasingly complex digital technology?

- On what basis can the records created in the course of each activity be presumed authentic? How, in the absence of such presumption, can their authenticity be verified?

- How is the authenticity of these records affected by their transmission across space and time? What controls on the process of transmission would ensure that these records will continue to be recognized as authentic?

- Are the conceptual requirements for reliability and authenticity developed by the UBC-MAS project and InterPARES 1 for administrative and legal records generated within databases and document management systems applicable to the records studied by InterPARES 2?

- Do the participants in electronic transactions have shared access to reliable and accurate information about the terms and effects of the transactions? What would constitute reliable and accurate records of transactions in current electronic service delivery initiatives?

- What would be the consequence of issuing guidelines for record creation on the nature of the records of each activity?

- How can cultural differences, freedom of expression, freedom of inquiry, and right to privacy be reflected in those guidelines?

- What technological and intellectual tools would assist creators to generate records that can be authentically preserved over time?

- What legal or moral obligations exist regarding the creation, use and preservation of the records under investigation?

Appendix 20

CREATOR GUIDELINES
Making and Maintaining Digital Materials: Guidelines for Individuals[1]

Introduction

Most information today is created and stored in digital form. The advantages of the digital medium are by now familiar to everyone. Documents can be created quickly and edited and revised with ease. Thanks to the Internet, they can be distributed globally with lightning-like speed. They can be manipulated in ways that allow them to be used for multiple purposes. The digital medium also solves the longstanding storage problems associated with large files of paper records.

The blessings of the digital era, however, are not without their costs. Only in recent years have people begun to fully grasp the many problems inherent in the digital medium. For example, there is the fact that digital information can only be accessed using a computer. Furthermore, the computer must be equipped with the necessary software to be able to read the bit strings contained on the disc or tape. Ease of reproduction and the proliferation of copies make it more difficult to identify a complete or final version of a digital document. Easy distribution of information on the Internet makes the preservation of intellectual property rights difficult. Finally, all digital materials are vulnerable to viruses and simple technology failure, as well as to the rapid developments in software and hardware that risk making them inaccessible very quickly.

With all of these problems, it is little wonder that some people yearn for the comforting tangibility of paper. Yet although our systems for creating and maintaining information will likely continue for some time to be hybrid systems—that is, containing both paper and digital materials—there is clearly no turning back from the digital revolution. Consequently, everyone should be aware of the risks faced by digital materials and know how best to minimize these risks.

These guidelines have been developed for individuals who create digital materials in the course of their professional and personal activities to help them make informed decisions about making and maintaining these materials in ways that will help ensure their preservation for as long as they are needed. They may also be useful for small organizations or groups of individuals, such as medical offices, consulting groups or teams of research scientists.

Although these guidelines can be applied to various kinds of digital publications, documents and data, they are especially important for digital records. Records are the documents that you make, receive and use in your activities, and that you keep because you may need them later or because you want to have reliable evidence of what you have done. Therefore, you need to be especially careful in maintaining and preserving them. These guidelines are applicable to records that need to be maintained for only a short period of time as well as to those that require long-term maintenance. Adherence to these guidelines will help ensure that records that merit long-

[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)creator_guidelines_booklet.pdf.

term preservation in an archival repository will be accessible when they are turned over to the care of a trusted custodian.

Definitions

Before presenting recommendations to guide you in making and maintaining digital materials, it will be both necessary and helpful to clarify the meaning of some of the terms used in this document.

For the purposes of these guidelines, a *record* is defined as any document created (i.e., made or received and saved for further action or reference) by a physical or corporate person in the course of a practical activity as an instrument and by-product of that activity. A *publication* is defined as a document intended for dissemination or distribution to the public at large. All records and publications are documents and contain data. A *document* is information affixed to a medium in a fixed form; *information* is an assemblage of data intended for communication over time or space; and *data* are the smallest meaningful and indivisible pieces of information.

These guidelines aim at providing recommendations for the creation and maintenance of reliable digital materials in general, and records in particular, that can be accurately and authentically maintained and preserved over time. To facilitate their application, however, the terms "reliability," "accuracy," "authenticity" and "authentication" need to be defined.

For the purposes of these guidelines, *reliability* is the trustworthiness of digital materials as statements of fact or as content. It is the responsibility of the author of the materials, be that author an individual or the corporate person in whose name an individual is writing, and is assessed on the basis of the material's completeness and accuracy and of the degree of control exercised on the process of its creation.

Accuracy is the degree to which the data in the materials are precise, correct, truthful and free of error or distortion. To ensure accuracy, one must exercise control on the processes of creation, transmission, maintenance and preservation of the materials. Over time, the responsibility for accuracy shifts from the author to the keeper of the materials and later to the long-term preserver of the materials (if applicable).

Authenticity refers to the fact that the materials are what they purport to be and have not been tampered with or otherwise corrupted. Thus, with respect to records in particular, authenticity refers to the trustworthiness of records as records. To ensure that authenticity can be presumed and maintained over time, one must define and maintain the identity of the materials and protect their integrity. Authenticity is at risk whenever materials are transmitted across space and time. Over time, the responsibility for authenticity moves from the keeper to the long-term preserver of the materials.

Authentication is a declaration of authenticity, resulting either from the insertion or the addition of elements or statements to the materials in question, and the rules governing it are established by legislation. Thus, it is a means of proving that materials are what they purport to be at a given moment in time. Digital authentication measures, like the use of digital signatures, only ensure that the materials are authentic when received and cannot be repudiated, but not that they will stay authentic afterwards.

Recommendations

1. Select hardware, software and file formats that offer the best hope for ensuring that digital materials will remain easily accessible over time.

Accessing digital materials depends on having the appropriate software. Software that is not compatible with previous versions (backward compatibility) or with future versions (forward compatibility) makes it difficult to access records over time. Software for one application also needs to work well with that of other applications and systems (interoperability). Paying attention to the following six factors can help ensure that your software and hardware maintain accessibility.

Choose software that presents materials as they originally appeared. Ideally, materials should keep the same look over time to be fully intelligible and accessible. Be sure that new software will be able to read your older materials in the software format in which you kept it and display it on the screen in the same documentary form in which it was originally displayed. In other words, new software should be backward compatible with older software.

Choose software and hardware that allow you to share digital materials easily. Software should be able to accept and output files in a number of different formats. The ability to interact easily with other technology is called *interoperability*. It will make it easier to access your materials and also to move them to other systems.

Use software that adheres to standards. This is one of the best things you can do to ensure your material will last. Standards endorsed by national and international organizations are best. These are called *de jure* standards.[2] If these do not exist for your material, you can help ensure longevity by adopting software that is very widely used. In the absence of an official standard, such software is often referred to as a *de facto* standard.[3] Open source software; that is, freely available non-proprietary software, is preferable (see subsection G on the next page).

Keep the specifications of software. This kind of documentation (e.g. the owner's manuals or any other more detailed description of the software you might have) will be essential in the future to access the materials or to migrate them to a new computer environment as technology advances. It is particularly important to fully document any software that you build yourself.

If you customize software, make sure you document the changes you make. Give detailed information about the changes and describe clearly the characteristics and features of the material these changes produce, as well as the outcomes you are trying to achieve by customizing the software. A good way to do this is to include the information as comments in the software code. The information will not get lost, as it is part of the file, and it will be very helpful to those who need to make adjustments later, as technology advances.

Document the construction of your system as a whole to help ensure its accessibility. You should document your system's structure and functions. This means identifying its

[2] Defined as: A standard adopted by an official standards-setting body, whether national (e.g., ANSI), multi-national (e.g., CEN) or international (e.g., ISO). For computer file formats, two recent de jure standards are PDF/A (PDF standard for archiving) and ODF (OASIS OpenDocument Format).

[3] Defined as: A standard not adopted by any official standards-setting body, but nevertheless widely used and recognized by its users as a standard. Well known and widely used computer file formats that are considered de jure standards include PDF, TIFF, DOC and ZIP.

hardware and software components, including peripherals, its operating system and software packages. Such documentation will identify how the software packages represent information, and how they process it and communicate it to each other and to users. These basic specifications will ensure that those who come after you understand the context in which you are working now. They will provide the information necessary to update the system as hardware and software evolve.

Choose widely-used, non-proprietary, platform-independent, uncompressed formats with freely available specifications where possible. These are often called "open formats," which means that their specification is published and freely available. However, it may also mean that the format is free of patent or royalty fees or the possibility of such fees being applied in the future, and/or that it is widely adopted. It should be noted that "open" formats are not necessarily the same as formats produced by *open source software*, as the latter term describes software for which the code is made freely available and can be modified. Open source software does not always produce non-proprietary formats. Distinguish between file formats, wrapper (or container) formats and tagged formats such as XML-tagged files, and ensure that version, encoding and other characteristics are clear and fully specified. For XML files, make sure that the files are well-formed and valid and accompanied by the relevant DTDs or schemas. If it is not convenient for you to follow this recommendation, consult with an archives that accepts digital materials and choose among the formats that it recommends for long-term preservation. You should not compress your digital materials, if at all possible, since this can lead to problems for their long-term preservation. If you need to compress them, choose lossless compression techniques that conform to accepted international standards.

2. Ensure that digital materials maintained as records are stable and fixed both in their content and in their form.

One of the great advantages of digital materials is the ease with which information can be edited, revised or updated. But this also means that important information can be changed or even lost, accidentally or on purpose. This is a particularly important problem for records, because one of the characteristics of a record is that its content is unchanged and unchangeable. This implies that the information and the data in the record cannot be overwritten, altered, deleted or expanded. A system that contains fluid, ever-changing information or data does not really contain records until someone decides to make them and save them with *fixed form*[4] and *stable content*.[5]

Although the idea of stable content is fairly simple, the concept of fixed form is more complex. Essentially, it means that the message conveyed by a digital record (or other digital object) can be rendered with the same documentary presentation it had on the screen when it was made or received and first saved. The bit streams that compose the digital record and determine its digital presentation (i.e., its file format) may change, but its documentary presentation must not change. A simple example is when a document created in Microsoft Word is later saved as an Adobe PDF file. Although the document's digital presentation has changed—from a Microsoft Word .doc file format to an Adobe .pdf file format—the documentary presentation of the

[4] Defined as: The quality of a record that ensures the documentary appearance or presentation is the same each time the record is retrieved.

[5] Defined as: The quality of a record that makes the information and data contained in it immutable, and requires changes to be made by appending an update or creating a new version.

document—also called its *documentary form*[6]—has not changed, and therefore we can say that the document has a fixed form.

In some cases, digital materials can be presented in several different ways—in other words, the information they convey can take different documentary forms. For example, statistical data can be presented as a pie chart, a bar chart or a table. However, the possible variations of these displays are usually limited by the system. In such cases, we can regard each documentary presentation as having stable content and fixed form, since the information is selected from a fixed store of data within the system and the system's rules govern the form of its documentary presentation(s).

A similar situation occurs when the selection of both content and form is from a large store of fixed information that is only partially accessed every time a user queries the system. If the same query always produces the same output as to content and documentary form, the output can be regarded as having stable content and fixed form. Thus, if you, as the author of the record, establish fixed rules for the selection of its content and of its documentary form that only allow for a known and stable range of variability— that is, endow it with *bounded variability*[7]—then you can claim that your material has stable content and fixed form.

The concern for the documentary presentation of digital materials is particularly important for maintaining and assessing the reliability and accuracy of records. Future upgrades, conversions or migrations of data may result in changes to the documentary form. Therefore, you would be wise to first establish the documentary form of records associated with each activity or procedure and then identify the essential characteristics (i.e., the essential *intrinsic* and *extrinsic* elements[8]) of each documentary presentation or form. This will help alert you to any changes in the future that would imply a loss of identity and integrity of the record, especially if you are active in the sphere of digital art, where a certified description of those essential characteristics by the artist would help support the recognition of the intellectual property rights linked to work so described.

3. Ensure that digital materials are properly identified.

Giving a meaningful name to a computer file helps identify its content and makes it easier to find. The full identification of records is more complex than just naming files, however. Full identification is essential in distinguishing records from each other, in distinguishing different versions of a single record and in providing evidence of the identity of a record from the moment of its creation through its long-term preservation.[9]

[6] Defined as: The rules of representation according to which the content of a record, its administrative and documentary context and its authority are communicated. Documentary form possesses both extrinsic and intrinsic elements.

[7] Defined as: The quality of a record that ensures that its documentary presentations are limited and controlled by fixed rules and a stable store of content data, form data and composition data, so that the same user activity, query, request or interaction always generates the same result.

[8] *Intrinsic Elements* are defined as: The elements of a record that convey the action in which the record participates and its immediate context, including the names of the persons involved in its creation, the name and description of the action or matter to which it pertains, the date(s) of creation and transmission, etc. *Extrinsic Elements* are defined as: The elements of a record that constitute its external appearance, including presentation features such as font, graphics, images, sounds, layouts, hyperlinks, image resolutions, etc., as well as digital signatures, seals, and time stamps and special signs (digital watermarks, logos, crests, etc.).

[9] In this context, *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. (See also Recommendation 4)

The information about digital materials that supports their identification and retrieval is commonly referred to as *metadata*.[10] Most software applications automatically tag all digital materials with some data about their identity because this information is necessary to locate documents effectively. Without metadata, it would be nearly impossible to find a document without opening and reading through a folder or several directories. Metadata describe the properties or attributes of digital materials. In the case of records, however, these properties or attributes are also necessary to maintain and assess their authenticity, and that is why it is important to ensure that all the essential ones are recorded and that they are correct.

The properties or attributes conveying the identity of digital materials are referred to as identity *metadata*.[11] These include:

a. *Names of the persons involved in the creation of the digital materials*. These include:
   - the *author*—the physical or corporate person(s) responsible for issuing the materials;
   - the *writer*—the physical person(s) or position(s) responsible for articulating the content of the materials;
   - the *originator*—the physical person, position or office responsible for the electronic account or technical environment where the materials are generated and/or from which it is transmitted;[12]
   - the *addressee*—the physical or corporate person(s) for whom the materials are intended; and
   - the *recipient*—the physical or corporate person(s) to whom the materials may be copied or blind copied.

b. *Name of the action or matter*—in other words, the title or subject.

c. *Documentary form*—in other words, whether it is a report, a letter, a contract, a table, a list, etc.

d. *Digital presentation*—in other words, format, wrapper, encoding, etc.

e. *Date(s) of creation and transmission*. These include:
   - the *chronological date* written on the materials or on which the materials were compiled;
   - the *dates of transmission and/or receipt*; and
   - the *archival or filing date*—in other words, the date when the materials were associated with a computer folder or directory, or other classification scheme or filing plan (see Recommendation 5).

f. *Expression of documentary context*—for example, a classification code, or the name of the computer folder or directory, or comparable filing unit within the classification scheme or filing plan to which the materials are associated, and the name of the broader group of records in which the materials belong (see also Recommendation 5).

g. *Indication of attachments*, if applicable.

h. *Indication of copyright or other intellectual rights*, if applicable.

[10] Defined as: Information that characterizes another information resource, especially for purposes of documenting, describing, preserving or managing that resource.

[11] Defined as: The properties or attributes conveying the identity of a digital object that is to be kept as a record. (See also Recommendation 5.)

[12] Identification of the originator is only important in cases where the person, position or office responsible for physically creating and/or transmitting the materials is neither the author nor the writer, and when the presence of the originator's name appearing on, or in association with, the materials calls into question the actual author and/or writer of the materials. This is most commonly associated with e-mails in instances where the name of the originator appears in the header of an e-mail and/or its attachments that were in fact authored and/or written by another person, but physically manifested and/or transmitted on behalf of that person by the originator.

i. *Indication of the presence or removal of a digital signature*, if applicable (see Recommendation 6, Technology-dependent Authentication section).
j. *Indication of other forms of authentication*, if applicable. This could include, for example, the presence of a corroboration (i.e., an explicit mention of the means used to validate the record); an attestation (i.e., the validation of a record by those who took part in the issuing of it, and by witnesses to the action or to the 'signing' of the record); a subscription (i.e., the name of the author or writer appearing at the bottom of the document), or a qualification of signature (i.e., the mention of the title, capacity and/or address of the person or persons signing the record).
k. *Indication of the draft or version number*, if applicable.
l. *Existence and location of duplicate materials outside the digital system*, if applicable. If multiple copies of a document exist, you should indicate which one is the official or *authoritative copy*.[13] If the document is certified by the author as an "approved reproduction" of a work (for example, a digital work of art), indication of the existence of such certification is required. If the document comprises material copyrighted by different author(s), indication of copyright clearance (or lack thereof) with related dates is necessary.

4. Ensure that digital materials carry information that will help verify their integrity.

Although the identity metadata help distinguish digital materials from one another, another set of metadata allows users to infer that the materials are the same as when they were created (although not to verify or demonstrate it, because this would require comparison with a copy of the materials kept elsewhere). These metadata can be referred to as *integrity metadata* (see below). Digital materials have *integrity*[14] if they are intact and uncorrupted, that is, if the messages that they are meant to communicate to achieve their purposes are unaltered. This means that the physical integrity of digital materials, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and its required elements of *documentary form* (see Recommendation 2) remain the same. The content and the data in it are considered to be unaltered if they are identical as to the value and presentation (i.e., position on the screen) of the content and data in the first saved manifestation of the material. The attributes that relate to the integrity of digital materials have to do with the maintenance of the materials, including the responsibility for their proper handling, such as overseeing and documenting any technological transformations or transfers of the materials to other systems. The integrity metadata include:

a. *Names of handling person/office*—the person or office using the materials to carry out business.
b. *Name of person or office with primary responsibility for keeping the materials*—this may be the same as the handling person/office.
c. *Indication of annotations added to the materials*, if applicable.
d. *Indication of any technical changes to the materials or to the application(s) responsible for managing and providing access to the materials*—for example, change of encoding, wrapper or format, upgrading from one version to another of an application, conversion from several linked digital components to one component only (e.g., by embedding

[13] Defined as: The instance 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 instances.
[14] Defined as: The quality of being complete and unaltered in all essential respects. With identity, a component of authenticity.

directly in the materials digital components that were previously only linked to the materials, such as audio, video, graphic or text elements like fonts).

e. *Access restriction code*—indication of the person, position or office authorized to read the materials, if applicable.
f. *Access privileges code*—indication of the person, position or office authorized to annotate the materials, delete them, or remove them from the system, if applicable.
g. *Vital record code*—indication of the degree of importance of the record to continue the activity for which it was created or the business of the person/office that created it, if applicable.[15]
h. *Planned disposition*—for example, removal from the live system to storage outside the system; transfer to the care of a *trusted custodian* (see Recommendation 10); scheduled deletion.

5. Organize digital materials into logical groupings.

The management and retrieval of your digital materials can be enhanced if you can handle them in large sets, rather than one by one. Therefore, it is important that you group your digital materials in some logical manner. The categories chosen may reflect the way you work, your activities, procedures, thematic areas, or some sort of structural organization. Separating your records from other digital materials is an important first step. The organization of your records may be based on the different types of records or the length of time for which certain kinds of records need to be kept. These groupings can be related to each other in a hierarchical or flat way, as best suits your needs. Generally, this structure should be consistent with the organization of any paper records you have (or records in other media), so that all records related to the same activity or subject, or of the same type, can be easily identified and retrieved as part of one conceptual grouping, as needed. Your organization scheme should be recorded in a document that shows all the groupings of materials, describes them in a brief sentence and indicates how they are related. In this document, which is called a *classification scheme*[16] or filing plan, each grouping of records can be assigned a code or a name that should be linked to each record belonging in the same grouping no matter what the medium or location: thus, the records assigned to each grouping will share such code or name, followed by a number that indicates their sequence. This identifier should be recorded among the *identity metadata*[17] of your digital records and on the face of your paper records belonging to the same grouping and should be unique for each record.

Identifying how long groupings of records need to be retained will facilitate their management while they are regularly needed and help ensure that records that need or merit long-term preservation are tagged early and given proper protection to ensure their survival.

You will find it easier and more efficient to assign a retention period—the length of time you want or need to keep materials—to a grouping of materials, rather than to individual items. Trying to ensure that some things are kept as long as needed while weeding out things that are no longer needed is simply too cumbersome at the individual item level. Although you may think

[15] The *vital record code* only pertains to specific communities of practices, such as legal and medical offices, who must identify the records that are vital to the continuance of their business in case of disaster and who would therefore exercise special protection measures on those records.
[16] Defined as: A plan for the systematic identification and arrangement of business activities and records into categories according to logically structured conventions, methods and procedural rules. (See also Recommendation 3.)
[17] Defined as: The properties or attributes conveying the identity of a digital object that is to be kept as a record. (See also Recommendation 3.)

that within a grouping some records should be kept longer than others, not only will you save time if you keep the whole grouping, but you will also have more complete information when you need to refer to the records. However, for some types of records, you can create subgroups within each given grouping on the basis of the retention period.

6. Use authentication techniques that foster the maintenance and preservation of digital materials.

The authenticity of digital materials is threatened whenever they are transmitted across space (i.e., when sent to an addressee or between systems or applications) or time (i.e., either when they are in storage, or when the hardware or software used to store, process or communicate them is updated or replaced). Because the acts of setting aside digital materials for future action or reference and of retrieving them inevitably entail moving them across significant technological boundaries (from display to storage subsystems and vice versa), the inference of the authenticity of digital materials must be further supported by evidence that they have been maintained using technologies and administrative procedures that either guarantee their continuing identity and integrity or at least minimize risks of change from the time the records were first set aside to the point at which they are subsequently accessed.

Technology-independent Authentication

Presumption of Authenticity. A presumption of authenticity is an inference that is drawn from known facts about the manner in which a document has been created and maintained. Adoption and consistent application of the recommendations presented in this document provide the best evidence to support such a presumption. The recommendations are cumulative: the higher the number of satisfied recommendations and the greater the degree to which an individual recommendation has been satisfied, the stronger the presumption of authenticity.

Successful implementation of the recommendations presented in this document is predicated on establishing and continuously applying effective administrative policies and procedures.[18] Ideally, you should strive to implement authentication techniques supported by administrative policies and procedures that are as technology-independent and/or neutral as possible.

Technology-dependent Authentication

Technology-dependent authentication techniques, such as cryptography, are used to provide a technological mechanism to guarantee the authenticity of digital materials. One such cryptographic technique is the digital signature, which can be used when transmitting documents between persons, systems or applications to declare their authenticity at a certain point in time. Such technologies have been given legal or regulatory value by some bodies, like the European Commission and the Securities and Exchange Commission.

Caution! Digital signatures are subject to obsolescence themselves and, by virtue of their purpose and inherent functionality, cannot be migrated to new or updated software applications together with the documents to which they are attached. In fact, 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. Unless or until further development of digital signature technology enables such encrypted authentication information to be preserved over time with the

[18] See Appendix 19, "A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records." Available at http://www.interpares.org/display_file.cfm?doc=ip2_book_appendix_19.pdf.

document, you should, when you receive a document with an attached digital signature, detach the signature whenever possible and add information to the integrity metadata to indicate that the document had an attached digital signature when received and that the signature was verified, detached and deleted.

7. Protect digital materials from unauthorized action.

The accuracy and authenticity of digital materials cannot be presumed if there is any opportunity for modifying them without leaving a trace. You need to be able to demonstrate that it was impossible for anyone to tamper with or manipulate your digital materials without that person being identified. Security includes restricting physical access to places where computers are kept, as well as restricting access to the digital materials 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.

It is also important to set up a structure of access permissions (also called access privileges—see discussion of *integrity metadata* in Recommendation 4) for all users of the system. For example, some users may only be able to read materials, while others may have permission to modify them. In any case, it should be impossible to modify any record once it has been filed according to the *classification scheme* or filing plan (see Recommendations 3 and 5), and only the person who has been given responsibility for recordkeeping and maintenance should be able to transfer or delete materials from the system. In addition, the system should maintain an audit trail to track access to the materials to control the administration and use of access privileges.

This recommendation may appear to be a tall order for individuals who may be working out of their homes, or even for those working in very small offices or communities of practice. But it is important to remember that if you cannot demonstrate that it was impossible for anyone to tamper with and manipulate your digital materials without being identified, your assertion that your records are de facto accurate and authentic becomes irrelevant. In this regard, it might be useful to keep copies of at least the most important digital materials offline and to establish some routine by which materials stored offline are randomly compared with their counterparts online on a periodic basis.

8. Protect digital materials from accidental loss and corruption.

Computers are not foolproof, and any of a number of factors can cause corruption or other accidental loss of records or data. The best way to ensure against accidental loss or corruption is to make backup copies regularly and often. If you store such copies off-site, additional protection is obtained against fire or theft of equipment. Many backup techniques, software packages, and services are available, including ones that automatically create the backup materials and then transmit them to a secure off-site location.

a. *Develop a rigorous policy or routine that ensures your system is backed up daily*. Your system is only as good as its last backup, so you need to make sure it is backed up often, at least once daily, using proven methods that will ensure that if something goes wrong, you and/or your business will be able to recover quickly. Such regular backups should be destroyed on a rotational basis according to a strategy or schedule that is most appropriate for your requirements, since they do not contain records but only exist for recovery of the system if it fails. Note that we are talking here about a comprehensive *system backup*, which includes the operating system, the software applications and all the digital materials in your system. If, in addition to a system backup, you need to have a security copy of your digital materials in case your computer is stolen or some of your records

become corrupted, then you should backup those materials only on another computer, an external hard drive or other portable digital media and store these security copies in an off-site location away from the computer with the "original" copies.

b. *Choose and install the best backup technology for your situation*. Study the technology and services available, and choose what works best for your particular situation. Many different systems are available, ranging from those covering one-person operations to those able to back up very large systems. The backup system needs to include an audit trail, in case the system fails between backups and you need to recover the records or other digital materials created during the time for which there is no backup.

9. Take steps against hardware and software obsolescence.

The speed with which hardware and software become obsolete poses severe challenges to the maintenance and long-term preservation of digital material. One strategy to address this problem is to eliminate dependence on hardware by transferring hardware functionalities to software (i.e., use a software application to simulate the actions of a piece of hardware). This provides a more stable way to retain the function when the hardware becomes obsolete.

The rapidly changing technology environment means that both individuals and offices should regularly upgrade their digital systems as well as all the records within these systems and those that have been moved to another storage medium, such as CD, DVD or tape. In other words, when parts of the technological environment in which you are working begin to become obsolete, they should be upgraded to the most advanced technology available according to your particular requirements and constraints, and all digital materials inside and outside the system should be migrated to the new technology. When replacing hardware, it is important for the replacement hardware to have capabilities at least equal to the hardware it is replacing. For example, a new monitor needs to display a graphic record in a way that retains the documentary form of the original record. Planning for regular technology upgrades on a rotational basis will help ensure that your technology does not become out of date and also help prevent large and unexpected technology expenses.

Sometimes digital records produced by or maintained in systems that are becoming obsolete need to be retained for a long time, but they are not expected to be accessed often. If such records are textual records and need to be read sequentially rather than randomly, you could convert them from their digital form to computer output microfilm. This will protect them from accidental loss or corruption better than any other measure. Another good protective measure is duplication—creating a second copy of groups of vital records and keeping it on another computer, on a second hard drive, on DVD, with another office or individual or in remote storage. When digital records or other entities are removed from a live system, for storage on magnetic or optical media outside the system, for example, it is essential that documentation about the system and about the digital materials (for example, the records' metadata) is also removed and kept with them. For more detailed information about the types of documentation in question here, see Recommendation 1, subsections D, E and F.

10. Consider issues surrounding long-term preservation.

Although the focus of this document has been on the creation and maintenance of all kinds of digital materials while they are needed on a regular basis by their creators, it is important to consider how best to preserve important digital materials for the long term. Typically, only a small percentage of materials need to be preserved for the long term, but the ability to provide ongoing, long-term care for materials, especially digital materials, is often beyond the capability

or interest of individuals and small organizations. There are real costs—both financial and human—in retaining materials for the long term, but such preservation efforts are essential for establishing and maintaining our cultural heritage, for accountability purposes and for informing managerial decision-making.

To begin this process, you should identify someone who will take charge of your digital materials once they are no longer needed for regular personal or professional purposes. This person would take the role of *trusted custodian*.[19] A trusted custodian is a professional—or a collection of professionals, as in an archives or a community historical society—who is educated in recordkeeping and preservation, and who ideally has no stake in the content of the records and no interest in allowing others to manipulate or destroy the records.

In the case of small organizations or offices, this person could be the one responsible for keeping the records and organizing and storing them during their active use. In the case of individuals who manage their own recordkeeping, the person fulfilling the preservation function may be an archivist or a librarian in a documentation centre, or simply themselves. In either case, a preservation strategy should be established as soon as possible, because digital materials that have not been targeted for preservation early and taken care of in a proactive way will not be preserved. Close adherence to these guidelines will therefore facilitate long-term preservation.

Conclusion

This document has outlined a series of activities for individuals and small organizations to carry out to create and maintain digital materials that can be presumed to be authentic, accurate and reliable. For individuals the burden may seem great, but the alternative—loss of records or the emergence of corrupt and unverifiable data—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 maintenance 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; you should be able to select and adopt the measures that address your particular problems in the specific context in which you operate. There may also be cases in which additional measures are necessary because of legislative or regulatory requirements specific to your field, or because of the characteristics of the activity and hence of the records that it produces. In such cases, consultation with experts may be required. Among such experts are the archivists of city, provincial, state or national archives, as well as local archival associations. Individuals, offices and small organizations should not hesitate to contact such experts for advice on any issues relating to the creation and maintenance of their digital materials.

Finally, this set of guidelines is but one of the documents issued by the InterPARES Project, an international research project studying the long-term preservation of authentic digital records. Additional material that will support the understanding of the nature of digital records and the development of methods for their reliable creation and accurate and authentic maintenance and preservation can be found on the InterPARES Web site at www.interpares.org.

[19] Defined 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.