

Should there be many libraries and archives—thought of as those who organize, catalog, and provide access to content—served by few repositories? How might such a many-few structure be established? Who would pay for what? Meanwhile, copyright considerations and plans to re-purpose content mean that significant chunks will be managed by their private sector owners. Will these commercial repositories be trustworthy? Will society inherit this content at the appropriate time and in a preservable form? Will there be so-called dark archives that keep content on behalf of society during the copyright period? Who would pay their cost? My familiarity with digital library conversations suggest that there are far fewer proposed answers to questions like these than there are to the problems in technology.

NOTES

This paper was first given at the AMIA conference in Boston, November 5, 2002.

1. Jeff Rothenberg, "Ensuring the Longevity of Digital Documents," *Scientific American* 272 (January 1995): 42–47.
2. *Ibid.*, 43.
3. *Garden and Forest* is at http://memory.loc.gov/ammem/ndlpcoop/moahtml/title/gndf_vols.html. The entire *American Memory* offering is at <http://memory.loc.gov>. Although *American Memory*'s digital files of motion pictures are not considered to be preservation copies, readers of *The Moving Image* may wish to examine these collections: select the category "Motion Pictures" from the *American Memory* "Collection Finder" at <http://memory.loc.gov/ammem/collections/finder.html>.
4. Research Libraries Group (RLG) and the Online Computer Library Center (OCLC), *Trusted Digital Repositories: Attributes and Responsibilities* (Mountain View, Calif.: RLG, 2002), 5.
5. See this Web site: <http://ssdoo.gsfc.nasa.gov/nost/isoas/>; the OAIS document proper is at <http://www.ccsds.org/documents/pdf/CCSDS-650.o-R-2.pdf>.
6. See <http://info.wgbh.org/upf/>.
7. See <http://www.loc.gov/standards/mets>.
8. Don Waters and John Garrett, "Preserving Digital Information: Final Report and Recommendations" (1996), at <http://www.rlg.org/ArchTF/index.html>.
9. Jeff Rothenberg, "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation" (1999), at <http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>.
10. In reference to newspapers, see Nicholson Baker's *Double Fold: Libraries and the Assault on Paper* (New York: Random House, 2003), reviewed in *The Moving Image* 2, no. 1 (spring 2002): 195–98. Baker argued that the microfilm reformatting of books and newspapers altered their content in damaging ways. He cited examples where inferior film images presented

illegible texts but, more to the point, called attention to the loss of color and clarity in illustrations and to the transformation of look and feel after the printed page had been converted to a frame of microfilm, now only viewable with a special device. The final blow, Baker argued, was that the originals were often discarded, robbing society of any form of access to them. Baker's argument is sobering to those who wish to preserve "born-digital" content, which is very likely to be transformed in the process, and should lead those planning the reformatting of "analog" content to choose approaches that minimize future transformations, as well as motivating archives to retain original items whenever this is practical.

11. Paul Wheatley, "Migration—a CAMILEON discussion paper," at <http://www.ariadne.ac.uk/issue29/camileon>. See also <http://www.personal.leeds.ac.uk/~issprw/camileon/migration.htm>; see section 2 for Wheatley's discussion of various levels of preservation.
12. Margaret Hedstrom and Clifford Lampe: "Emulation vs. Migration: Do Users Care?" at <http://www.rlg.org/preserv/diginews/diginews5-6.html#feature1>.
13. Raymond Lorie: "A Project on Preservation of Digital Data," *RLG DigiNews*, June 15, 2001, at <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>.
14. See <http://www.npaci.edu/Thrusts/DI/index.html>.
15. See documents at <http://www.sdsc.edu/NARA/Publications.html>.

InterPARES

The Search for Authenticity in Electronic Records

YVETTE HACKETT

The research of the InterPARES project has focused, for a number of years, on textual electronic records created in institutional settings. As a result, its work to date may not have attracted the attention of the archival moving image community. The InterPARES 2 project, currently in its initial stages, is the first to turn its attention to audiovisual documents. This paper attempts to give a brief overview of the earlier work in order to offer an introduction to the research methodologies used and the findings that form the starting point for the current project's research.

InterPARES 2 is actually the third project in a series of archival investigations of digital materials. The first two projects were the Preservation of the Integrity of Electronic Records (1994 to 1997) and the InterPARES project (1999 to 2001). All three projects are firmly rooted in the School of Library, Archival, and Information

Studies at the University of British Columbia (UBC). This academic starting point has, from the beginning, set the research within a time-tested set of assumptions and definitions from the records management and archival disciplines, developed over several centuries. By enlisting the participation of records creators, archivists, and information technologists, the research continually tests traditional knowledge and requirements against current record-keeping practices, current operational realities in archival institutions, and the ability of the current state of technology to support the needs of these groups.

THE PRESERVATION OF THE INTEGRITY OF ELECTRONIC RECORDS

This first project is referred to by a number of names, most commonly the UBC Project, the UBC-MAS Project, or the UBC-DoD Project. Luciana Duranti and Terry Eastwood, who both teach archival studies at UBC, served as principal investigator and coinvestigator, respectively. Heather MacNeil, a doctoral student at UBC at that time, joined the project as a research assistant.

The goal of the first project, to quote MacNeil, was “to identify and define conceptually the nature of an electronic record and the conditions necessary to ensure its reliability and authenticity during its active life, based on the concepts and methods of diplomatics and archival science.”¹ In other words, the project focused on records while still in the hands of the records creator.

Diplomatics provided an important starting point for the researchers, but because it is a field of study little known outside European textual archival studies, it deserves a brief introduction. Diplomats emerged in the seventeenth century as an analytical technique for determining the authenticity of records issued by sovereign authorities in previous centuries. The tenets and methods of diplomatics were first laid out in 1681 by Jean Mabillon, a Benedictine monk. Mabillon examined, among other things, the language of documents, their characteristic parts, their seals, and the systems of chronology used in dating them. For over four centuries, diplomatics has been used to help determine a record’s authenticity for legal pur-

poses and to assess medieval records as historical sources.

In the early 1990s, Duranti began to explore the possibility of adapting essentially paper-based diplomatics to address the growing questions about the reliability and authenticity of electronic records. The results of this study appeared in six articles, originally published in *Archivaria*, the journal of the Association of Canadian Archivists. They have now been published as a book, *Diplomatics: New Uses for an Old Science*.²

The UBC-MAS Project attracted the attention of the U.S. Department of Defense Records Management Task Force (DoDTF), who joined the research in 1995 in search of improved methods for managing the department’s traditional records and the growing volume of records being created in electronic form. The DoDTF participants introduced the UBC researchers to IDEF modeling, an Integrated DEFINITION language and methodology, which was used to define records management functions in an operational environment and break them down into smaller and more precise units of analysis.

The researchers concluded that to maintain the reliability and authenticity of electronic records, the creator should

- establish agency-wide control over the creation, handling, and preservation of all its records by embedding procedural rules in its records system;³
- integrate business and documentary procedures;
- institute procedures such as classification and profiling that make explicit the relationships between and among records participating in the same activity; and
- integrate the management of the electronic and nonelectronic components of the records system.

The researchers also concluded that the reliability and authenticity of active and semi-active electronic records are best ensured by the creator, while the authenticity of inactive records is best ensured by the “preserver,” defined in the “InterPARES Glossary” as the “juridical person whose primary responsibility is the long-term preservation of authentic records.”⁴

These findings largely confirmed the continued validity of traditional records management practices, while translating them into an electronic work environment. In practical terms, this translation resulted in a U.S. Department of Defense standard: DoD 5015.2, the Design Criteria Standard for Electronic Records Management Software Applications.⁵ This software specification describes the essential characteristics and functionality for a software application designed to control digital objects. In its simplest form, it involves three parts:

- metadata (frequently stored in a database structure) to record information about the document or digital object;
- the document or digital object itself; and
- a link between the two.

INTERPARES

The first project had left the researchers eager to pursue the study of electronic records, but this time from the perspective of the records preserver. It was a logical continuation of the first project, which had restricted itself to the viewpoint and activities of the records creator and ended its investigation at the moment of transfer of inactive records to an archival institution.

The results of the first project attracted interest from other universities, from archives in a number of countries, and from the business community. By the time the second project concluded in December 2001, over one hundred researchers and research assistants from twelve countries had participated. Major funding contributions were received from Canada's Social Science and Humanities Research Council, the American National Historical Publications and Records Commission, the National Archives and Records Administration of the United States, and the Italian National Research Council, as well as universities and national archival institutions from around the world. Business interests were primarily represented by CENSA, the Collaborative Electronic Notebook Systems Association, which represents pharmaceutical and other industry interests.⁶ Luciana Duranti served as project director.

Originally referred to as "UBC Part 2," the project members selected a new name that reflected the international nature of the undertaking: International Research on Permanent, Authentic Records in Electronic Systems, known by its acronym InterPARES. *Inter pares* is also Latin for "among peers," an appropriate reflection of the collaborative nature of the project.

The research was organized into four task forces, dealing with authenticity, appraisal, preservation, and strategies. The first three task forces would examine the impact electronic records might have on traditional archival principles and procedures, while the Strategies Task Force would formulate principles to guide the development of strategies, policies, and standards for the long-term preservation of authentic electronic records, based on the findings of the first three task forces.

The Authenticity Task Force relied on case studies of systems currently operating in primarily large institutions in both the public and private sectors in a number of countries. The Appraisal and Preservation Task Forces returned to IDEF modeling to define archival functions and then test the models against the systems examined in the case studies. It is not possible to do justice to the scope and complexity of the appraisal and preservation models in a brief overview. Both are available on the InterPARES Web site as appendices to the groups' final reports.⁷

APPRAISAL TASK FORCE

The Appraisal Task Force was chaired by Terry Eastwood. The researchers' primary objective was to "determine whether the evaluation of electronic records for permanent preservation should be based on theoretical criteria different from those for traditional records [and] . . . how digital technologies have affected the methodology of appraisal."⁸ They were also interested in how the concept of authenticity being developed by the Authenticity Task Force would exhibit itself within the archival appraisal function.

The resulting "Model of the Selection Function" started, at its highest level, with the function "Select Electronic Records" (see Figure 1). By the time the task force members completed their analysis, the model contained

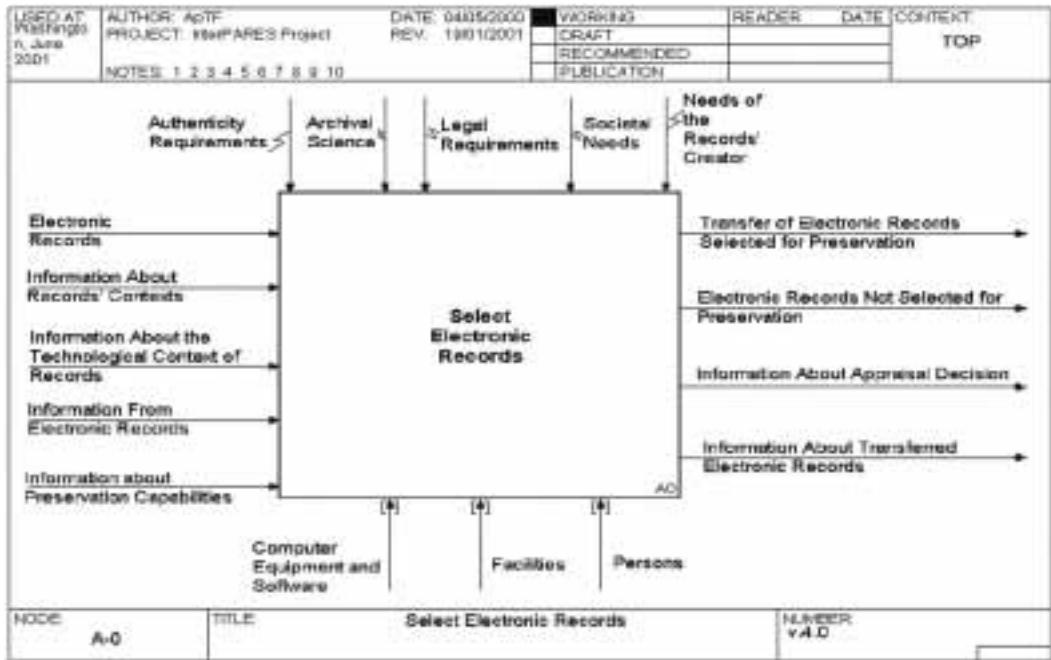


Figure 1. Appraisal Model diagram showing the “Select Electronic Records” function and the inputs (left arrows), outputs (right arrows), constraints (top arrows), and mechanisms (bottom arrows) that affect that function.

eighteen additional subfunctions in four levels, each level of decomposition offering greater precision and detail about the constituent activities of the archival function of appraisal (see Figure 2).

The model is based on the premise that appraisal of electronic records is best conducted when records are still active, and gives greater visibility to steps such as “Determine Feasibility” and “Assess Authenticity,” concepts that have always existed in the paper world but may not have been emphasized or codified in the past. The IDEF modeling methodology also requires the rigorous definition of all the activities represented in the various boxes, as well as of the inputs and outputs represented by the arrows.⁹

Task force members concluded that the medium of records affects the process of appraisal but not the fundamental task of assigning value, and that information compiled during appraisal must be “packaged” and preserved, as it is crucial to enable the eventual transfer of records, their preservation, and their description.

PRESERVATION TASK FORCE

The Preservation Task Force, which was chaired by Ken Thibodeau from the National Archives and Records Administration, focused on what

“preservation” means for electronic records. The task force’s findings offer a clarification of the language used when discussing the preservation of electronic records. The final report concludes that “it is not possible to preserve an electronic record: it is only possible to preserve the ability to reproduce the record.”¹⁰ As a result of this finding, the task force extended the preservation process to include access to the records because this process of reproduction has taken on additional degrees of difficulty in a digital environment.

The analysis also separates an electronic record into “physical and intellectual components, which do not necessarily coincide.” The final report also suggests replacing the traditional term “chain of custody” with “chain of preservation,” a more complex concept that attempts to represent the preservation of authentic electronic records as a continuous process that begins with records creation and acknowledges the need to thoroughly document the preservation process as a primary means for protecting and assessing authenticity over the long term.

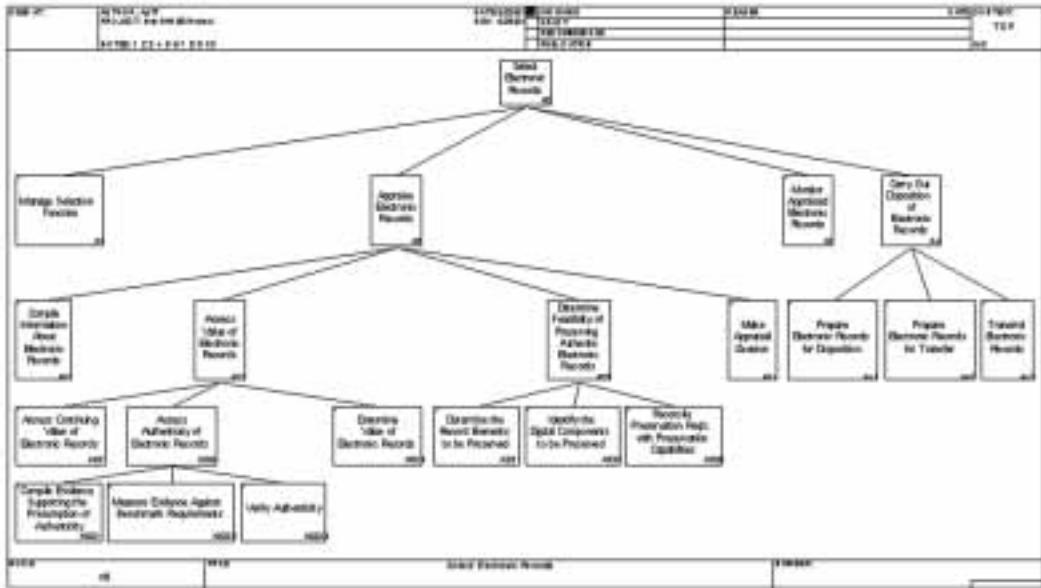


Figure 2. Appraisal Model diagram showing the overview of the complete model. Each subfunction is analyzed in more detail in the model.

THE AUTHENTICITY TASK FORCE

The Authenticity Task Force, chaired by Heather MacNeil, began its investigations within a conceptual framework for authenticity based on traditional archival theory and jurisprudence that can be summarized as follows: if records are relied on by their creator in the usual and ordinary course of business, they are presumed to be authentic. But with electronic systems, the presumption of authenticity must be supported by evidence.

Any discussion of the findings of the Authenticity Task Force must be grounded on a number of key terms, including the following three word pairs: document/record, reliability/authenticity, and identity/integrity. While a *document* is simply “recorded information,” a *record* is “any document made or received and set aside in the course of a practical activity.”¹¹ To be considered reliable, a record must be able to stand for the facts it is about (such as a marriage license, as a record of the exchange of vows at a ceremony, or a land title as proof of ownership of a certain property). In other words, *reliability* refers to the trustworthiness of a record as a statement of facts. A record’s *authenticity* relates only to evidence that it is what it purports to be and has not been tampered with or otherwise corrupted, in other words, to the trustworthiness of a record as a record. A record, therefore, can be “unreliable” (i.e., it misrepresents the facts to which it at-

tests) and, at the same time, authentic (i.e., it was written by the person who claims to have written it and it has not been inappropriately modified since its creation). Authenticity is rooted in the identity and integrity of the record. *Identity* is established by those attributes of a record that uniquely characterize it and distinguish it from other records, while *integrity* involves the intact articulation of the record’s content and any required elements of documentary form.

A record cannot be “sort of” or “more or less” authentic. But the presumption of authenticity moves on a sliding scale. A weak presumption of authenticity would not prevent an archives from acquiring something, particularly if it was unique, but the reasons why the presumption of authenticity for a given set of records is weak would need to be explained in the appraisal documentation before the records are transferred as well as in the archival description of the records that is prepared following their transfer. If, during the process of assessing the authenticity of electronic records, it is found that the presumption of authenticity is weak, the Authenticity Task Force has identified some alternative means for verifying their authenticity, such as comparison to other existing copies of the record, textual analysis, or a study of audit trails.

Having established this framework and tested its assumptions against the case studies and traditional archival science, the Authenticity Task Force developed two sets of requirements for authenticity.¹² The first are the “Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records.” These relate to the environment in which the records are created. Although there are eight requirements identified, the number that need to be met in order to support a presumption of authenticity will depend on the nature of the records, the purpose for which they were created in the first place, and the purpose for which they are being preserved over the long term.

The second set, the “Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records,” apply to the preserver. Unlike the benchmark requirements, all three of the baseline requirements are considered mandatory. In other words, all three requirements need to be met in order for the preserver to be capable of attesting to the authenticity of the records in its custody.

The first benchmark requirement is the most detailed and is expressed as fifteen data elements required in order to establish the identity and the integrity of a record. These elements must be expressed in the record, be inextricably linked to the record, or be consistently recorded in the context of the record’s creation. This concept has been implemented through the “profile” in electronic records management systems, which are compliant with DoD 5015.2.

1a. Record attributes (Identity of the record)

- Names of the persons concurring in the formation of the record, that is, name of author, name of writer (if different from the author), name of originator (if different from name of author or writer), name of addressee
- Name of action or matter
- Date(s) of creation and transmission, that is, chronological date, received date, archival date, transmission date(s)
- Expression of archival bond (e.g., classification code, file identifier)
- Indication of attachments

1b. Record attributes (Integrity of the record)

- Name of handling office
- Name of office of primary responsibility (if different from handling office)
- Indication of types of annotations added to the record
- Indication of technical modifications

The balance of the benchmark requirements are procedural:

2. Access privileges
3. Protective procedures: loss and corruption of records
4. Protective procedures: media and technology
5. Establishment of documentary forms
6. Authentication of records
7. Identification of authoritative record
8. Removal and transfer of relevant documentation

The case studies showed that while the controls relating to access, loss and corruption, and media and technology tended to be at least partially implemented within automated systems, the balance, if they were present, relied primarily on traditional procedures outside the system. For example, access (2) is frequently controlled by passwords, but also in some circumstances by physical security. Regular backup procedures are probably the most familiar method used to protect against the loss and corruption of records (3), while technological obsolescence (4) is guarded against by regularly upgrading hardware and software and migrating applications to new platforms.

The establishment of documentary forms (5) involves the specification of formats for specific types of documents, such as certificates or official correspondence, established either by the juridical system or by the creator. Authentication (6) is sort of a subcategory of authenticity in that it is a statement that a record is what it purports to be at a given moment in time, for example, a method to declare a certified true copy of a record for use in court. The ease of duplication of electronic records requires an established method to identify which copy of a record within an organization is the authoritative one (7). Finally, the dependency

between electronic records and the automated system in which they were created will require an established procedure to determine what documentation must be removed and transferred to the preserver when records are removed from their system of origin (8). The degree to which the creator meets all these benchmarks can be evaluated during appraisal.

The "Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records" are intended for the preserver and as such are more stringent. They include controls over records transfer, maintenance, and reproduction expressed, for example, as unbroken custody of the records, and adequate security for records in the archival repository; adequate documentation of the impact of the reproduction process on the form, content, structure, accessibility, and use of the records in archival custody; and finally, information about significant changes made to the records since their creation, whether performed by the creator prior to transfer or by the preserver.

In addition to these three baseline requirements, the eight benchmark requirements "for the creator" also apply to the "preserver" when the preserver is creating records detailing the acquisition and preservation of archival records.

In the last year of InterPARES, participants began turning their attention to nontextual records. This was primarily through the participation of John Roeder, a music professor at UBC, and of the project's postdoctoral fellow, Brent Lee, a composer and musician now teaching at the University of Windsor. His tale of the loss of "Mr. Crumb,"¹³ one of his earliest digital compositions, illustrated that nontextual electronic records shared many of the same longevity problems being experienced by textual electronic records, such as media obsolescence, lack of backward compatibility, and proprietary software formats. But this artistic environment has also been shaped by a much different concept of authenticity and authentic performance, more complex hardware and software relationships extending to MIDI interfaces and synthesizers, and the concept of "interactive" pieces where a computer improvises a musical accompaniment using algorithmic software to a performance by a musician or dancer.

INTERPARES 2

This growing interest in nontextual records eventually grew into InterPARES 2, a five-year project that began in January 2002 and is scheduled to end in December 2006. The new project is organized into six teams (Canadian, American, European, Australian, Asian, and African). A seventh "global industry" team is again coordinated by CENSA. Most of the researchers from InterPARES 1 remain involved, while participation from the academic sector has increased substantially. Current participants represent five continents, twenty-one countries, eleven national archives, thirty-one universities, several state, provincial, and municipal organizations, and over one hundred researchers.¹⁴

While InterPARES 2 started with a growing interest in music and the arts, the full scope of the project developed into a much larger undertaking, which can be divided into three sets of three: specifically, three areas of activity, three types of documents, and three perspectives.

First, the project will examine documents created in the course of artistic activities, scientific activities, and government on-line. The scientific community has dealt for many years with the concept of accuracy, particularly as it applies to machine-readable data. Government on-line will examine governments' current push to load data, information, and records of all kinds to Web sites and to interact with citizens in this environment. As for artistic activities, the project will examine as many varied artistic forms as possible, subject to the limitations of time and the expertise available to the project. To date, the artistic forms include moving images, music, photography, museum installations, new media, dance, theatre, and performance art.

Second, the project is deliberately seeking documents that it describes as "dynamic, interactive and experiential." Developing adequate definitions for each of these terms forms part of the research. We hope this emphasis will enable us to push research forward into the newest developments in technology and records creation; in InterPARES 1, the research confined itself mainly to the examination of databases, imaging systems, and records created with standard desktop software packages.

Finally, the project will adopt three separate perspectives: the creation and maintenance of digital "information objects," their preservation and access, and concepts defining their accuracy, reliability, and authenticity. As such, the research agenda encompasses the scope of both the original UBC-MAS project and InterPARES 1 and has moved beyond the strict definition of "record" to include a wide range of textual and nontextual documents.

Initial projects include the joining together and harmonization of the functional models created by the Appraisal Task Force and the Preservation Task Force to create a consistent description of the complete archival appraisal and preservation cycle that can be used to develop a more detailed activity model for this integrated view. This could eventually become a concrete basis for system development for archives.

Case studies will again form an important part of the research. To date, twelve case studies have been approved. These include two performance artists (one focused on dance and one on street theatre), an on-line magazine, and the CyberCartographic Atlas of Antarctica, a multiple-media Web site. Another case study will focus on film production, allowing a comparison between large-scale studio production and the independent filmmaking environment.

An early glimpse of the direction the research is taking should be available in the fall of 2004 on the InterPARES 2 Web site.

NOTES

This paper was originally presented at the Association of Moving Image Archivists annual conference held in Boston, Massachusetts, November 21, 2002. The author would like to thank Heather MacNeil for her advice and assistance in the preparation of this paper for publication.

1. Heather MacNeil, *Trusting Records: Legal, Historical, and Diplomatic Perspectives* (Dordrecht: Kluwer, 2000), 90. For a detailed description of the UBC-MAS Project and its findings, see Luciana Duranti and Heather MacNeil, "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project," *Archivaria* 42 (fall 1996): 46–67.
2. Luciana Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, Md.: Scarecrow Press, Society of American Archivists, Association of Canadian Archivists, 1998).
3. In the UBC-MAS Project, the records system is

understood to comprise the creator's records, and its record-keeping and record-preservation systems. The *record-keeping system* is a set of rules for making, receiving, setting aside, and handling active records, along with the tools and mechanisms necessary to implement those rules. The *record-preservation system* is a set of rules for exercising physical and intellectual control over the semiactive records, along with the tools and mechanisms necessary to implement them.

4. InterPARES Project, "The InterPARES Glossary," available on the InterPARES 2 Web site, at http://www.interpares.org/book/interpares_book_q_gloss.pdf.
5. U.S. Department of Defense, Assistant Secretary of Defense for Command, Control, Communications, and Intelligence Design Criteria, *Standard for Electronic Records Management Software Applications* (DoD 5015.2-STD) (Washington, D.C.: U.S. Department of Defense, 2001).
6. A complete list of project participants is available on the InterPARES 1 Web site, at <http://www.interpares.org/participants.htm>.
7. The complete appraisal model is available as an appendix to the Appraisal Task Force's final report on the InterPARES 1 Web site, at http://www.interpares.org/book/interpares_book_m_appo4i.pdf. The complete preservation model is available as an appendix to the Preservation Task Force's final report on the InterPARES 1 Web site, at http://www.interpares.org/book/interpares_book_n_appo5ii.pdf.
8. InterPARES Project, "Research Plan," available on the InterPARES 1 Web site, at <http://www.interpares.org/researchplan.htm>.
9. Arrow definitions are available as an appendix to the Appraisal Task Force's final report on the InterPARES 1 Web site, at http://www.interpares.org/book/interpares_book_m_appo4iii.pdf.
10. InterPARES Project, "Preservation Task Force Report," p. 6, available on the InterPARES 1 Web site, at http://www.interpares.org/book/interpares_book_f_part3.pdf.
11. InterPARES Project, "The InterPARES Glossary," available on the InterPARES 2 Web site, at http://www.interpares.org/book/interpares_book_q_gloss.pdf.
12. InterPARES Project, "Authenticity Task Force Report," Appendix 2, available on the InterPARES 1 Web site, at http://www.interpares.org/book/interpares_book_k_appo2.pdf.
13. Brent Lee, "The Growing Complexity of Music Preservation," published in *Preserving Authentic Electronic Records: Preliminary Research Findings*, proceedings of an international symposium held at the University of British Columbia, February 17, 2001, pp. 36–44, available at http://www.interpares.org/documents/interpares_symposium_2001.pdf.
14. A complete list of project participants is available on the InterPARES 2 Web site, at http://www.interpares.org/ip2_participants.cfm.