

The InterPARES

International Research Project

LUCIANA DURANTI AND KENNETH THIBODEAU

AT THE CORE

THIS ARTICLE EXAMINES

- the nature of this project and its international agenda
- durable concepts and models for electronic records
- practical benefits of the research

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) is a large, multinational, collaborative research endeavor that aims to formulate principles and criteria for the development of international, national, and organizational policies, strategies, and standards for the long-term preservation of authentic electronic records. (See Figure 1 for information about research participants and funding.) The research project investigates issues related to the authenticity,

retrievability, and accessibility – over the long term – of records made, received, and stored in digital form. Its area of inquiry is divided into four domains.

Domain I

Domain I aims to identify requirements for preserving electronic records whose authenticity can be verified for an indefinite period of time, regardless of the number of migrations that have occurred from an obsolete technology to a newer one. The research questions being addressed are directed toward

- categorizing electronic records based on the elements that must be preserved to ensure their authenticity, including identifying the elements that all electronic records share
- differentiating between different types of electronic records
- verifying their authenticity over time

- verifying their authenticity in time (i.e., at the point at which they are originally used)

- ascertaining whether records can be removed from where they are currently found to a place where they can more easily be preserved – but still maintain the same degree of value

Domain II

Research in Domain II seeks to establish whether, to satisfy the requirements for authenticity identified in Domain I, the appraisal criteria and methods for electronic records need to be revised or even radically changed by considering

- the influence of digital technology on appraisal criteria
- the ways appraisal differs, depending on the type of system prevalent in each computing phase
- the influence of the records' medium and extrinsic elements on appraisal
- how retrievability, intelligibility, functionality, and research needs influence appraisal

Editor's Note: The name "InterPARES" is based on the Latin term inter pares meaning "among equals."

- the matter of whether restraints can be imposed onto the modification of systems at the time of appraisal
- whether the life cycle of electronic records differs from that of traditional records
- when electronic records should be appraised
- whether electronic records should be appraised more than once in the course of their existence and, if so, when
- how electronic records are scheduled
- identifying who should be responsible for appraising electronic records

Domain III

The purpose of Domain III is to develop methods, procedures, and rules for the preservation of electronic records according to the requirements identified in Domain I and to define the responsibilities for implementing them, considering

- what methods, procedures, and rules of long-term preservation are in use or being developed
- which of these methods meet the conceptual requirements for authenticity identified in Domain I
- which methods of long-term preservation need to be developed
- which methods are either mandated or subject to standards, regulations, and guidelines in specific industry (e.g., biochemical) or institutional (e.g., university) settings
- the procedural methods of authentication for preserving electronic records
- whether archival description can be a method of authentication for electronic records
- if appraisal and acquisition/accession reports can be constructed to allow for authentication of electronic records
- determining procedures for certifying electronic records when they cross physical or technical

FIGURE 1: InterPARES Participants and Funding

The InterPARES research effort, under the overall direction of Luciana Duranti (University of British Columbia), is multinational, regional, and interdisciplinary. The InterPARES research is being carried out by national and multinational research teams from various countries, including Canada, the United States, England, Ireland, Sweden, Finland, The Netherlands, Germany, France, Italy, Australia, Hong Kong, and China.

The national teams are the Canadian, American, Australian, and Italian research teams. The multinational teams are European and Asian. A global industry team includes multinational companies in the pharmaceutical, biochemical, health, and computer fields.

The national and multinational research teams are funded by their national granting agencies and by institutional and organizational contributions. For example, the Canadian team is funded by Social Sciences and Humanities Research Council of Canada (SSHRC), and the American team is supported by the National Historical Publication and Records Commission (NHPRC). Funding also comes from institutional sources such as the Hampton Fund of the University of British Columbia (UBC) and the UBC Vice President Research Fund.

boundaries (e.g., refreshing, copying, migrating) to preserve their authenticity

- technical methods of authentication for preserved electronic records
- the principles and criteria for both media and storage management required for the preservation of authentic electronic records
- the responsibilities for the long-term preservation of authentic electronic records

Domain IV

Domain IV focuses on developing a framework for the formulation of strategies, policies, and standards by examining

- what principles should guide the formulation of *international* policies, strategies, and standards related to long-term preservation of authentic electronic records
- the criteria for developing *national* policies, strategies, and standards

- the criteria for developing *organizational* policies, strategies, and standards

The Project's Theoretical Roots

The theoretical foundation of the research comes from diplomacy, archival science, and law. (*Diplomatics* is a science developed in the 17th century for determining the authenticity and legal validity of medieval charters and deeds; in the 19th century, historians adopted its analytical techniques as a tool for assessing the authority of medieval records as historical sources.)

From the theoretical foundations, the basic concepts to be examined are those of "record," "electronic record," "reliability," "authenticity," and "authentication." A *record* may be understood as any document (i.e., recorded organizational information) created (i.e., made or received and set aside) in the course of a practical activity as an instrument and as residue of it.

An *electronic record* is any record that, in the ordinary course of business, is used and set aside or stored in digital form regardless of whether it was made or received in such form. (Note that, according to this definition, a record received on paper, scanned into a computer, used, and set aside in digital form is an electronic record. Conversely, however, a record received electronically, printed out, inserted into a paper file, used, and set aside in paper form, is a paper record.)

Authenticity refers to the fact that a record is what it purports to be and has not been tampered with or otherwise corrupted since its creation.

Reliability refers to the ability of a record to stand for the facts it contains (i.e., the trustworthiness of the record's content). Reliability depends on two factors: the degree of completeness of the record's form and the degree of control exercised over its procedure of creation. Therefore, reliability is of greater concern to the record's creator than to its preserver.

Authenticity refers to the fact that a record is what it purports to be and has not been tampered with or otherwise corrupted since its creation (i.e., the trustworthiness of the record as a record). Records' authenticity depends on (1) their mode, form, and state of transmission as drafts, originals, or copies (at this stage, authenticity is still a concern of the record's creator) and (2) on the manner of their maintenance, preservation, and custody (at this stage, authenticity is a concern of the record's preserver, whether as a part of the creating body or a separate entity). The mode of transmission of the record is the means used to transmit a record across space or time. The form of transmission is the physical carrier on which a record is received (e.g., paper, film, disk).

Authentication is a declaration of authenticity at a given point in time,

resulting either by the insertion or the addition of an element or a statement to a record. The rules governing it are established by legislation. Although *authenticity* is a quality of the record that is to be constantly protected over the long-term, authentication is external to the record itself and is temporary. Neither declaration guarantees a record's reliability – both an *authentic* and an *authenticated* record are, strictly speaking, as reliable as they were when first issued by their creator.

Because electronic records can be preserved only by reproducing them, the authenticity and the authentication of concern refers to copies. This statement implies that an organization might have both authentic and authenticated copies of nonauthentic records. This situation could certainly be the case if the originals were corrupted or tampered with either during transmission, when first saved to a file, or when maintained by their creator in active state.

An *original record* is always either authentic or nonauthentic. It either is or is not the record that was produced and set aside in the course of activity. However, the original record may be copied for different purposes. The authenticity of the copy depends on the purpose for which it is produced. For example, if someone wants to use the contents of a record to assist in making a decision about a similar matter, a simple transcription of the text may be sufficient to consider the copy to conform to the original.

The InterPARES project is investigating the hypothesis that the requirements for authentication of a copy of an electronic record are directly related to the function for which the copy is made. In particular,

this project is concerned with making copies for the purpose of preservation (copies that can stand for the originals). To achieve this aim, the preservation copy must be able to serve the same record function as the original.

Relation of Action and Record

With respect to the action it participates in, a record may have four possible functions. The record may

1. be the essence and substance of the action that comes into existence through the written form of the record (i.e., *dispositive* function)
2. be the proof of the action, which comes into existence and is complete before the compilation of the record but requires a written form to be proven (i.e., *probative* function)
3. provide support to an action and be procedurally linked to it through its compilation and be entirely discretionary (i.e., *supporting* function)
4. relate nonprocedurally to actions that do not require its compilation, which therefore fulfills individual purposes (i.e., *narrative* function)

This categorization is fundamental for the identification of requirements for authenticity. This identification is the outcome expected of the work in Domain I and is the responsibility of the Authenticity Task Force.

Project Work To Date

The work of the Authenticity Task Force encompasses three stages: (1) developing a template capable of guiding the analysis of electronic records, (2) carrying out such analysis through four rounds of case studies of electronic records and systems, and (3) establishing a typology, or classification, of types of electronic records based on authenticity requirements.

A "Template for Analysis" (available at www.interpares.org). [accessed

20 December 2000]) was developed for identifying the elements of an electronic record relevant to a consideration of its authenticity. The template breaks electronic records into their constituent parts, defines them, explains their purpose, and indicates to what extent they are instrumental in verifying authenticity.

The template has been developed from the integrated perspectives of diplomatics and archival science, and it is based on the concept of a record described earlier. The validity of the template has been, and continues to be, tested through case studies to define conceptual requirements for authenticity that can be translated into specific methods and procedures by the task forces responsible for the Appraisal and Preservation research Domains I and II. These conceptual requirements include baseline requirements applicable to all electronic records and specific requirements associated with distinct types of electronic records. To construct the latter kind of requirements, a typology of electronic records is being developed.

Traditional records typologies are based on documentary form. Identification of the specific elements of electronic records that must be preserved over time and across technologies for the purpose of verifying authenticity cannot be linked to current record forms or applications because the technologies change.

The typological framework chosen, then, is based on the function of the record with respect to the **action** in which it participates. The records observed in the case studies will be categorized as dispositive, probative, supporting, and narrative. Each of these four types contains subtypes, which contain variants. For each type, the Authenticity Task Force will determine which elements of an electronic documentary context and form must be protected over time and the degree of rigor necessary. This determination will allow for the identification of the descriptive metadata and procedural documentation that must be brought forward with the electronic records so that decades from now their authenticity can be verified.

Prospects for the Project

The value of having a great deal of data about real electronic records and a needs analysis of that data as well is obvious. A case in point involves the U.S. National Archives and Records Administration (NARA). The collection-based persistent-object preservation method being developed for NARA by the National Partnership for Advanced Computational Infra-structure offers great promise for preserving electronic records in a way that is both immune to the vagaries of technological obsolescence and attuned to the benefits offered by technological advances (Thibodeau, Moore, Baru, 113-118). The method depends on archivists' ability to express abstractly the properties of electronic records that must be preserved. It requires explicit models both of individual records and of bodies of records.

Preservation models are needed for all types of records and also any arbitrarily organized and complex body of records. The InterPARES case studies will contribute substantially to addressing these needs both through the data about electronic records that are collected and in the norms concerning authenticity produced by that data's analysis.

However, the collection and analysis of data about electronic records will need to continue to be carried out regularly. Innovations yet to come will be at least as great as those that have occurred so far. Continuous change in information technology means that the challenges posed by electronic records are inherently dynamic. One of the most important premises of the information management architecture that underlies persistent-object preservation is precisely that it must be independent of the infrastructure on which it is built. Although users must rely on specific hardware and software, preservation systems should be constructed in a manner that

- enables any component of hardware or software to be replaced with minimum impact on the system

- allows such replacement with no significant impact on the records being preserved
- enables the preservation system to deliver the preserved records to other systems – even systems not yet invented

The possibilities that this technological strategy offers of building inherently dynamic solutions to preservation problems only heightens the need to express authenticity requirements in a way that is not bound to the limitations of users' knowledge of electronic records now or in the future. As a rich store of data and analyses in the InterPARES case studies is building, knowledge that is both greater than the sum of its parts and open-ended is also developing.

The need to expand users' knowledge of the emerging types of electronic records and the ways of organizing them is important. The "Template for Analysis" is such an important product of the InterPARES project for this reason. Without abandoning well-established principles, it allows users to be open to new realities and to avoid distorting them through the constraints of parameters rooted in a specific time, space, and technological knowledge. In fact, the template is based on the fundamental assumption of a principle from diplomatics:

- the administrative, provenancial, procedural, and documentary context of a record's creation is made manifest in its form
- this form can be separated from the record's content and examined independently of it
- that by comparing different record forms generated in different times and contexts, users can discover both the attributes they share and the nature and purpose of those attributes they do not share

For example, by comparing the digital signature with traditional means of validation, the InterPARES researchers can establish that it fits the definition of an authenticating

"seal" by allowing the addressee to verify the origin and integrity of the record. However, the digital signature is different from a seal in that a seal is associated exclusively with a person or organization; the same seal is used to authenticate every record issued by the same entity (e.g., seal of a registrar on a student's transcript).

A digital signature, however, will vary for every record even when it belongs to the same person. This kind of understanding – gained from examining the function of records elements in their context – will allow users to recognize elements generated by new applications according to their purpose and to establish how vital they are to verification of a record's authenticity over time.

Electronic records differ from traditional paper records in basic ways – in how they exist as durable objects, in how the parts of the record coalesce into the whole, and as to the boundaries of the record. Traditionally, seen from an empirical perspective, a record is a specific aggregate of information, structured in a specific way, with that structure materialized in the inscription of the information on a durable medium. An electronic record has no necessary connection between the structure of the record and its inscription on a storage medium.

Therefore, in addressing the research questions of Domain III, the InterPARES Preservation Task Force is building a preservation model on the recognition that preserving an electronic record is literally impossible; only the ability to reproduce an electronic record is possible to preserve. One can store the contents of the record, along with special bit strings that indicate how it should be structured and presented, but the sum of those bits is not itself the record. The application of some software is needed to put the bits into a state recognizable as part of a record. Moreover, the aggregate of information that comprises a record's contents might not be stored together (e.g., separation of "notes" from the rest of a database record). The assembly of the contents into what

constitutes the record may require sophisticated and complex processing. This processing may be transparent – even unknowable – to those who create or use the record.

Such a simple thing as the appearance of an electronic record is not necessarily an attribute of the record itself. The hardware and software used to present the record may be more important than what is in the stored bits. The appearance of the bits depends on screen size and resolution and on the specific software used, and it can be changed easily and often as a result of user options such as window size, zooming, or switching between draft and page mode. Therefore, the characteristics of how an electronic record is physically inscribed to a medium are not essential to its character either as a record or as an authentic record.

The first way, then, to certify the authenticity of a preserved electronic record is to demonstrate that none of the essential attributes of the record identified in the conceptual requirements for authenticity have changed. This certification can be done only if the preservation function is exercised in such a careful way that any changes that do occur are identified and documented. If this procedure can be accomplished, the need to justify that none of the changes affected essential attributes of the record is still present. That is to say, the authenticity of a preserved

for authenticity. This process would ensure that a system built for preservation of electronic records has capabilities for satisfying the prerequisites. For example, although characteristics of the physical inscription of records on media are not essential attributes or elements of the records as such, users may lose records or lose their authenticity both in physical storage and in the processes of writing and reading physical media. Therefore, any preservation system should include safeguards and procedures to protect against media damage or deterioration.

The system must also be reliable enough to copy files completely and correctly to both store and output authentic electronic records. Based on the inevitable necessity to reproduce an electronic record, demonstrating the authenticity of electronic records depends on verifying that

- the right data was properly stored
- either nothing happened in storage to change this data or any changes are insignificant
- all the right data and **only** the right data were retrieved from storage
- the retrieved data was subjected to an appropriate process, and the processing was executed correctly to output an authentic reproduction of the record

...preserving an electronic record is literally impossible; only the ability to reproduce an electronic record is possible to preserve.

electronic record can be certified only if users are able to show that none of the specific authenticity requirements applicable to the record were violated.

Generic preservation measures should be prerequisites for both the baseline and the specific requirements

Ensuring that these technical requirements were satisfied is necessary – but not sufficient – to demonstrate the authenticity of an electronic record. Verification is *necessary* because if any of these conditions are not satisfied, the result of the processing of retrieved data

cannot be the same as the electronic record from which the stored data was produced. Verification is not *sufficient* because nothing in it applies specifically to a record. This technical verification is a method for demonstrating that a digital object produced from stored digital data is an authentic reproduction of a stored digital object. ("A digital object that was stored," is more precise language than "The digital object that was the source of the stored data.")

Situations or contexts in which identifiable risks of changing the records as *boundary conditions* can occur. A boundary condition is a state from which a record cannot be moved without either changing the record or taking some action either to prevent the threatened change or to counteract or compensate for it. Classes, or even class hierarchies, of boundary conditions may be defined. For example, any processing of electronic records entails some risk that the records will be altered in the process. In turn, the processing class can be seen as a subclass of technology boundaries, a subclass that comprises all dependencies of electronic records on specific technologies. A boundary condition is encountered whenever a technological dependency is altered or removed or the technology itself is changed – as in migration of data.

Another boundary class is that of custody, both physical and legal. Obvious risks are present whenever records change custody, even in cases in which records are transferred from their creator to a person committed to archival preservation. If the new custodian is what NARA calls a "successor in function," that is, an entity that acquires the records because it has a need for them in the conduct of its business, it may reorganize the records to serve its business needs. This action destroys their ability to serve as evidence of the activities of the former custodian.

From this example, one can begin to see another generic measure for preserving authenticity. Whenever custody of records is transferred to a different entity, the receiving body must be capable of and committed to preserving authentic records (Thibodeau 1997).

Conclusion

The wealth of the data being collected, the soundness of the methods being applied to collect and analyze them and to develop new knowledge from them, and the richness and depth of the multidisciplinary, international expertise brought to bear in the InterPARES case studies and on the analysis of the data constitute a tremendous opportunity both to find ways to document epochal changes in the infrastructures and superstructures of records that must be preserved and to explore the implications of such changes for the preservation of material that can be trusted.

However, one should not underestimate the practical importance of the findings of the InterPARES project for the design of reliable recordkeeping systems, for the development of rigorous security measures and control procedures that ensure the authenticity of records since creation, and for the definition of trustworthy methods of overcoming obsolescence of live systems. As it progresses in its investigation, this project continues to prove that preservation of electronic records that does not take a global view of their management cannot take place. ■■

ABOUT THE AUTHOR: Dr. Luciana Duranti is Professor and Chair of the Master of the Archival Studies Program, University of British Columbia, Vancouver, British Columbia. She has 27 years' experience in the information management field and has a focus on diplomatics (records reliability and authenticity) as well as electronic records. She is a member of the Association of Canadian Archivists and the Society of American Archivists (SAA), is a past president of SAA, and chairs its Fellows Committee. Duranti received the Academic of the Year Award from the Canadian Universities Faculty Associations/British Columbia in 1999. She earned a doctorate in archival science at Universita' di Roma (Italy). She can be reached at luciana@unixg.ubc.ca

Dr. Kenneth Thibodeau, is Director, Electronic Records Archives Program, National Archives and Records Administration, College Park, Maryland. He has 25 years' experience in information management with emphases on archival material and electronic records. He is a member and Fellow of the Society of American Archivists. He received his doctorate at the University of Pennsylvania. Thibodeau can be reached at ken.thibodeau@arch2.nara.gov.

REFERENCES

- Committee on Electronic Records. *Guide for Managing Electronic Records from An Archival Perspective*. Paris: International Council on Archives, February 1997
- Council on Library and Information Resources. *Authenticity in a Digital Environment*. Washington, DC: Council on Library and Information Resources, 2000.
- Duranti, Luciana and Terry Eastwood. *Diplomatics. New Uses for an Old Science*. Lanham, NJ: Scarecrow Press, 1999.
- Eriandsson, Alf. *Electronic Records Management, A Literature Review*. Paris: International Council on Archives, 1997.
- Office for the Official Publications of the European Communities. "European Citizens and Electronic Information: The Memory of the Information Society." In *Proceedings of the DLM Forum on Electronic Records, Brussels, 18-19 October 1999*. Luxembourg: Office for the Official Publications of the European Communities, 2000.
- Thibodeau, Kenneth. "Boundaries and Transformations: An Object-Oriented Strategy for the Preservation of Electronic Records in the European Commission." In *Proceedings of the DLM Forum on Electronic Records, Brussels, 18-20 December 1997*. Luxembourg: Office for the Official Publications of the European Communities, 1997. 161-67
- Thibodeau, Kenneth, Reagan Moore, and Chaitanya Baru. "Persistent Object Preservation: Advanced Computing Infrastructure for Digital Preservation." In *Proceedings of the DLM Forum on Electronic Records, Brussels, 18-19 October 1999*. Luxembourg: Office for the Official Publications of the European Communities, 2000.