

D-Lib Magazine July/August 2000

Volume 6 Number 7/8

ISSN 1082-9873

Preserving the Authenticity of Contingent Digital Objects

The InterPARES Project

Anne J. Gilliland-Swetland, Assistant Professor
Department of Information Studies
University of California, Los Angeles
swetland@ucla.edu

Philip B. Eppard, Dean and Associate Professor
School of Information Science and Policy
University at Albany, State University of New York
pbe40@csc.albany.edu

1.0 Introduction

In the development of digital libraries and of digital information systems in general, increasing attention is being given to issues relating to the preservation and authenticity of digital objects in order to assure their long-term accessibility and physical and intellectual integrity [[Lynch](#) 1994, [Duranti and MacNeil](#), 1996, [Bearman and Trant](#), 1998, [Rothenberg](#), 1999, [Council on Library and Information Resources](#), 2000]. Different types of digital objects have varying preservation and authenticity requirements, however, depending upon the contexts of their creation and use. Furthermore, these requirements are also subject to differing degrees of stringency. The most basic requirements for establishing the authenticity of a digital object may be very similar to the heuristics that information literacy programs seek to inculcate in end users working with of any type of information -- that is, establishing the who, what, when, where, how, and why associated with that information. The most stringent requirements for digital objects are arguably those imposed by legal warrant and business processes upon records of organizational or personal activity that are made or received and set aside for further action or reference in electronic form [[Duff](#) 1998].

Demonstrable integrity of preserved *electronic records* is critical to ensuring the accountability of the parent organization as well as its ability to rely on its records in the conduct of its business -- issues of increasing concern with the rise of e-commerce. However, while records are created primarily for such purposes, they have other uses and values that often cause them to be exploited for other purposes within digital information systems -- they can be managed and mined as active corporate knowledge assets, or preserved and made available as archival sources for historical scholarship and popular use. How such records are understood, used, preserved, and verified over time is highly contingent upon the juridical-administrative, procedural, provenancial, documentary,

and technological contexts. As a result, archival and recordkeeping approaches to the management of electronic records have been focused on the functions, processes, and uses associated with the records, rather than on physical object control. In digital information systems, however, where electronic records may be subjected to a range of uses and actions by both the original creators and secondary researchers, both approaches will have to be facilitated and the same information objects (i.e., the electronic records) will need to be both fixed and mutable when accessed for different purposes.

Several specific issues arise when addressing the preservation of authentic electronic records:

- Records are heterogeneous distributed objects comprising selected data elements that are pulled together by activity-related metadata such as audit trails, reports, and views through a process prescribed by the business function for a purpose that is juridically required. Identifying the boundaries of such intellectually complex objects and then moving those objects forward through time and through migrations without compromising their authentic status is a significant issue.
- Records are temporally contingent -- they take on different values and are subject to different uses at different points in time. Records are also time-bound in the sense that they are created for a specific purpose in relation to a specific time-bound action.
- The degree to which a record can be considered reliable is dependent upon the level of procedural and technical control exercised during its creation and management in its active life. Authenticity, by contrast, is the responsibility of archival management of inactive records, and is an absolute concept [[Gilliland-Swetland, 2000](#)].

Issues such as these that relate to the preservation and authenticity of record and archival materials are being addressed from several perspectives by current research projects, including CAMiLEON (Creative Archiving at Michigan & Leeds: Emulating the Old on the New, investigating the viability of emulation as a preservation strategy that maintains the "look and feel" of a software-dependent document); Cornell University's Prism (focusing on policy enforcement for ensuring information integrity in the areas of preservation, reliability, interoperability, security, and metadata) [[Prism](#)]; and the San Diego Supercomputer Center's Collection-Based Persistent Archives (deriving XML information models from collections of software-dependent data objects and developing tools that can be used to ensure preservation and access to those objects over time) [[Moore, et al. 2000](#)]. This paper reports on the ongoing work of InterPARES [[International Research on Permanent Authentic Records in Electronic Systems](#)], a multi-disciplinary collaborative archival research project that is taking a record-centric approach to the development of a typology of requirements for maintaining the authenticity of records over time, and analyzing appraisal and preservation processes in order to establish the extent to which they meet those requirements.

2.0 The InterPARES Project

Issues of authenticity and long-term preservation are central to the work of archivists, and so it is appropriate that researchers from the archival community should engage in efforts to address issues surrounding the accessibility to authentic electronic records over time. Professor Luciana Duranti of the School of Library, Archival and Information Studies at the University of British Columbia is the director of the international research team participating in InterPARES. The research builds on an earlier project at UBC, "The Preservation of the Integrity of Electronic Records," [[UBC](#)] which addressed issues surrounding the creation and maintenance of authentic and reliable electronic records in their active, pre-archival state [[Duranti 1995](#), [Duranti and MacNeil 1996](#)]. One of the products of that research was the U.S. Department of Defense's 5015.2 standard for records management applications <<http://jitc.fhu.disa.mil/recmgt/index.htm>>. The current project seeks to extend this work by considering the problems of maintaining the authenticity of electronic records

that must be preserved for extended periods of time.

The InterPARES project is organized into national, multi-national, and industry-based research teams. There are research teams in Canada, the United States, Italy, Northern Europe (United Kingdom, Ireland, Sweden, France, and the Netherlands), Australia, and Asia (China and Hong Kong) as well as a global industry group that includes CENSA (the Collaborative Electronic Notebook Systems Association). The national and multi-national teams include academic researchers, representatives of the national archival institutions in the various countries, and industry. Funding for the project has been provided by the Social Sciences and Humanities research Council of Canada, the National Historical Publications and Records Commission in the United States, the Italian National Research Council, and the U.S. National Archives and Records Administration, as well as by other funding agencies and institutions in the countries represented in the projects. In addition to archivists, the research teams include members who are computer scientists, preservation experts, lawyers, and media specialists.

Much of the work of the research is being carried out through a series of task forces that correspond to four research domains: authenticity; preservation; appraisal; and policies, strategies, and standards. A glossary committee oversees the compilation of a glossary of all of the terms used in the InterPARES project. The glossary, currently under development, will ultimately be a multi-lingual glossary that will also take account of variations in usage between different national and professional communities. While this glossary supports full understanding of the products of the research, it is hoped that it will have a much broader utility to the archives, preservation, and digital library communities.

3.0 Identifying Requirements for Preserving the Authenticity of Electronic Records

The theoretical framework within which InterPARES is operating is that of contemporary archival diplomatics. Diplomats was first developed in Europe in the eighteenth century as an analytical approach to the identification of the authenticity of medieval ecclesiastical documents, and its principles influenced the development of both modern history and theories of legal evidence. Diplomats studies the genesis, forms, and transmission of archival documents; their relation to the facts represented in them; and their relation to their creator in order to evaluate and communicate their true nature [[Duranti 1998](#)]. In recent years, this approach has been adapted by archival theorists for application to contemporary archival records. The theory underlying contemporary archival diplomatics has continued to be developed and tested with reference to understanding electronic records, first through the UBC Project and now through the InterPARES Project.

A major goal of InterPARES is to use contemporary archival diplomatics to analyze the elements of documentary form that occur in records associated with different types of actions and the juridical-administrative, procedural, provenancial, documentary, and technological contexts within which they occur. From this analysis, a typology of requirements for authenticity for records is being created.

3.1 Template for Analysis

Based on the prior work of the UBC Project and assessment of what is known about the characteristics of existing paper and electronic records, the Project has developed a Template for Analysis as a working hypothesis about the necessary and sufficient elements of a record. The template is a model of an ideal record that, based upon prior archival knowledge of record types, contains all the possible known elements that a record may contain. However, where diplomatic typologies and other analytical methods have in the past been developed retrospectively based upon what is known about existing records, this template is being developed as a predictive model that

will assist archivists in identifying future record types and their associated requirements for maintaining their physical and intellectual integrity over time.

The basic premise of the diplomatic approach is that recordkeeping functions and processes endure even if the physical manifestation of the record changes because of technological implementations. The template provides indicators that might allow archivists, and society more broadly, to identify when and how specific types of records have changed, are being re-invented, or where totally new forms are emerging; and hence to begin to understand the extent to which recordkeeping in the digital world exhibits continuity or discontinuity with what we know of past and present record functions, processes, forms, and implementations.

The Template for Analysis identifies and defines all the possible elements that a record may contain, explains the purpose of each element, and whether, and to what extent, it plays a specific role in ensuring the record's authenticity. The elements are organized into five categories:

- the medium, i.e., the physical carrier of the content;
- extrinsic elements of documentary form, i.e., the elements that determine a record's material make-up and its appearance -- including language; presentation features; seals (including digital signatures and authentication certificates of a trusted third party); special signs identifying one or more of the persons involved in the compilation, receipt, or execution of the record and which are distinct from a signature or seal (such as digital watermarks or the logo or crest of an organization); and other possible extrinsic elements of form, such as digital time stamps and digital signatures;
- intrinsic elements of documentary form that convey the action in which the record participates and its immediate context. These elements include the names of the author, originator, addressee, and receiver; the chronological date (and potentially exact time); the place of origin of the record; indications and description of the action (e.g., subject line or caption); and validation mechanisms, such as corroboration, attestation, and qualification of signature;
- annotations, i.e., additions, made to a record after its compilation or receipt in the course of its management.
- context or framework in which the action in which the record participates takes place.
 - the Juridical-Administrative Context, i.e., the legal and organizational system in which the creating body belongs.
 - Provenancial Context, i.e., the creating body, its mandate, structure, and functions.
 - Procedural Context, i.e., the business procedure in the course of which the record is created.
 - Documentary Context, i.e., the documentary aggregation to which the record belongs and its internal structure.
 - Technological Context, i.e., the hardware and software environment in which the record exists. ([InterPARES Authenticity Task Force](#), 2000).

3.2 Grounded Theory Development and Case Studies

To refine the Template for Analysis further, as well as to construct the electronic records typology that will be based on it, a form of grounded theory is being used. Four successive rounds of case studies of electronic information and recordkeeping systems are being used to identify and describe

phenomena, and to develop and test the Template for Analysis. Because a grounded theory approach is being used, theoretical, rather than statistical, sampling is being applied in the selection of case studies. In other words, we are identifying the cases that will best elucidate the aspects that the research is seeking to understand. In order to inform theory development, the case study data are coded for inter-related themes and concepts by means of an instrument called a Template Element Data Gathering Instrument that then is used to populate and refine elements contained in the draft Template for Analysis. The case studies are, therefore, interpretive and are directed towards not only understanding the elements of form of electronic records but also the situatedness of those records within their various contexts as well as the relationships of those contexts to each other. While identifying the intellectual components that comprise the record is fundamental, it is only by examining electronic information and recordkeeping systems through the lens of these contexts that we can really identify what is the appropriate unit of examination for the diplomatic analysis. The case studies conducted so far include large-scale databases (such as student registration systems and genetic databases), geographic information systems, and web-based applications (such as online interactive sites). Case studies are also being conducted of systems performing similar functions but in different national, institutional, and technological contexts.

4.0 Modeling the Preservation Process and the Appraisal of Electronic Records

Both the Preservation and the Appraisal Task Forces are using IDEF0 modeling to develop unambiguous high-level models and functional decompositions of the records preservation and appraisal functions. The preservation modeling addresses the management of the preservation function, the ingestion of electronic records, the maintenance of electronic records, and the delivery of electronic records in terms of their reproduction, assessment of preservation strategies to identify the extent to which they address authenticity requirements, certification of authenticity, information about electronic records, and information about the preservation process.

The determination of which records merit long-term retention in an archives (i.e., appraisal) is one of the most challenging aspects of archival work, one made even more difficult by the contingent nature of electronic records. The Appraisal Task Force, therefore, is examining questions surrounding the influence of digital technology on the criteria for appraisal, the timing of appraisal, and the responsibility for appraisal. A literature review of appraisal methods for electronic records was conducted and is available on the InterPARES website. The Appraisal Task Force has begun the process of modeling the appraisal function using the IDEF0 modeling methodology with the purpose of defining the activities involved in the selection of authentic electronic records for long-term preservation. The task force has considered appraisal as part of a larger function, which we are calling "Select Electronic Records."

In the modeling exercise, the task force is viewing the selection process from the standpoint of the entity responsible for the long-term preservation of electronic records, without any presumption that the entity will necessarily be an archival agency. It is clear from the work so far that a central part of the appraisal of electronic records for long-term preservation relates to the feasibility of preservation both from the standpoint of institutional resources and from an understanding of what precisely needs to be preserved in order to maintain authenticity. Therefore the modeling exercise for appraisal is integrating into its work the research of the Authenticity Task Force by incorporating into the appraisal process an analysis of how the record elements necessary to maintain authenticity are related to the various components of the technological context in which the records exist. Although the IDEF0 models produced by the Appraisal Task Force and the Preservation Task Force are being developed separately, the two groups are sharing information with each other with the understanding that we will need to produce models that can be easily integrated with each other.

5.0. Conclusions and Areas of Ongoing Research

The real issues then become what are the indicators that help us to see when true change is happening in functionality, forms, and implementation of records; how do we move that intangible intellectual construct of the record forward through time while maintaining its integrity; what are the events or other triggers that warn us that the record entity is losing its "recordness" over time; how do we recreate the original record upon demand regardless of whether it is maintained in an archives or in an active business system, and what form might that recreated record take?

The research has already identified several key areas that will demand closer study:

- **Affixedness** - The notion of a record needing to be physically affixed to a medium in order to be a record (concept of the physical carrier of the record). The case study data so far indicate that the medium is incidental and transparent and does not play a significant role in assuring authenticity, except in the immediate moment of rendering the record, e.g., in a screen display.
- **Fixity** - Intellectual fixity is more critical than physical fixity and is generally absent, at least conceptually. How is it to be achieved? The "setting aside" of a record (e.g., through processes such as capture, registration, and storage) needs to be triggered by some intellectual event that represents the intellectual closure of that activity, or some other indication that the record has achieved the consequences it was created to achieve.
- **Temporal Views** - Can they be reconstructed? Completed records kept in live systems without being physically segregated or otherwise set aside are generally still subject to retrospective updating or reformatting when the system's data structure is changed or the system is migrated.
- **Annotations** - When annotations are made to a record after its compilation or receipt or in the course of its management, they are not readily identifiable.
- **Juridical-Administrative Context** - It is difficult to identify juridical persons involved in the creation of electronic records because they are frequently not readily visible but are inferred or implied based on the context and other intellectual elements of form in the record; are inherited values from other elements; or are inserted automatically through the presentation or display.

These are some of the issues identified in the early stages of the research. All of the data from the first two rounds of case studies has not been fully analyzed yet, and a complete diplomatic analysis will take place over the next several months. The findings of this analysis will be used to refine the Template for Analysis and thus inform the later rounds of case studies. As it continues its research, the InterPARES team will also be studying existing strategies for digital preservation, such as migration, emulation, and persistent object preservation, as well as any new strategies that might be developed. Obviously research in this area cannot be conducted in a vacuum, and the centrality of records to business, government, and society at large makes the ability to maintain the authenticity of these electronic records, which by their very nature are contingent digital objects, an area of growing importance. By using the record, i.e., the contingent digital object itself, as the unit of study, and diplomatic analysis, which has been used to demonstrate authenticity of records in the past, the InterPARES project seeks to understand better the nature of electronic records and the elements necessary for ensuring their authenticity over time.

Acknowledgments

The authors gratefully acknowledge the funding support of InterPARES by the United States National Historical Publications and Records Commission, the Social Sciences and Humanities Research Council of Canada, the National Archives and Records Administration of the United

States, and the Italian National Research Council.

References

Bearman, David and Jennifer Trant. "Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process," *D-Lib Magazine* June 1998

<<http://www.dlib.org/dlib/june98/06bearman>>

Council on Library and Information Resources. *Authenticity in a Digital Environment* Washington, D.C.: Council on Library and Information Resources.

<<http://www.clir.org/pubs/abstract/pub92abst.html>> .

Duff, W. 1998. "Harnessing the power of warrant." *American Archivist*. 61:88-105.

Duranti, L. 1998. *Diplomatics: New uses for an old science*. Lanham, MD: Society of American Archivists, Association of Canadian Archivists, and Scarecrow Press.

Duranti, L. and H. MacNeil. 1996. "The protection of the integrity of electronic records: An overview of the UBC-MAS Research Project." *Archivaria*. 42:46-67.

Duranti, L. 1995. "Reliability and authenticity: the concepts and their implications." *Archivaria*. 39:5-10.

Eastwood, Terry. *Appraisal of Electronic Records: A Review of the Literature in English*.

<<http://www.interpares.org/documents/AppraisalLiteratureReview.doc.html>>

Gilliland-Swetland, A.J. 2000. *Enduring paradigm, new opportunities: The value of the archival perspective in the digital environment*. Washington, D.C.: Council on Library and Information Resources. <<http://www.clir.org/pubs/abstract/pub89abst.html>>

International Research on Permanent Authentic Records in Electronic Systems (InterPARES).

<<http://www.interpares.org>>

InterPARES Authenticity Task Force. *Template for Analysis Version 2.0*, May 22, 2000.

Lynch, C. A. 1994. "The integrity of digital information: Mechanics and definitional issues." *Journal of the American Society for Information Science*. 45:737-44.

Moore, R., C. Baru, et al. 2000. "Collection-based persistent digital archives." *D-Lib Magazine*. 6, nos. 3-4. <<http://www.dlib.org/dlib/march00/moore/03moore-pt1.html>> and

<<http://www.dlib.org/dlib/april00/moore/04moore-pt2.html>>

Prism. Digital Libraries Initiative Phase 2. Cornell University. <<http://www.prism.cornell.edu>>

Rothenberg, J. (1999). *Avoiding technological quicksand: Finding a viable technical foundation for digital preservation*. Washington DC: Council on Library and Information Resources.

<<http://www.clir.org/pubs/abstract/pub77.html>>

UBC (University of British Columbia). Preservation of the Integrity of Electronic Records Project (UBC Project). <<<http://www.slais.ubc.ca/users/duranti/>>>

US-InterPARES. <<http://is.gseis.ucla.edu/US-INTERPARES>>

Copyright © 2000 Anne J. Gilliland-Swetland and Philip B. Eppard

[Top](#) | [Contents](#)
[Search](#) | [Author Index](#) | [Title Index](#) | [Monthly Issues](#)
[Previous story](#) | [In Brief](#)
[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

[DOI](#): 10.1045/july2000-eppard