Authenticity as a Requirement of Preserving Digital Data and Records

by Shelby Sanett and Eun Park [1]

Abstract
Assuring continued authenticity is an essential but intransigent preservation consideration for digital data and records. Several key issues need to be addressed: Which intellectual and technical elements of data and records are essential for assuring authenticity; how should these be maintained and represented over time; and how are authentic data and records used in various systems of practice? The authors will address these questions in light of case studies and interviews being conducted with government agencies, academic institutions, and various organizations in America, Canada, Europe and Asia by the InterPARES (International Research on Permanent Authentic Records in Electronic Systems) Project. This article will also discuss initial project findings as they relate to the specific characteristics of authenticity in the preservation of digital data and records.

I. Introduction

Why is it important to know that preserved digital data and records[2] are authentic? How do we define authenticity? How do we know that received digital data and records are authentic? How are we assured that the digital data and records are as authentic when we retrieve them as they were when they were first stored and preserved?

These questions are large in scope. Our presentation explores the notion of the significance of authenticity in the management of records and data and reports upon the work of the InterPARES project currently underway.

The records generated by society, whether in the course of government, business or private activity, need to be maintained and preserved as a mechanism for accountability; as evidence of individual and corporate rights; and as a form of long-term memory. In the paper world, documentary forms and procedures have developed over time to ensure that records are capable of serving as evidence of activity - to be so, the records must be both reliable and authentic. *Reliability* can be defined as the trustworthiness of the content of the record, which is ascertainable through an examination of the completeness of the record and of the procedures exercising control over its creation[3]. Charles M. Dollar states that "Archival science defines authentic records as being what they purport to be — reliable records that over time have not been altered, changed or otherwise corrupted."[4] *Authenticity* guarantees that the record is not changed or manipulated after it has been created or received or migrated over the whole continuum of records creation, maintenance and preservation[5]. In the context of records as legal evidence, authenticity is an absolute concept in that it either exists or does not. There is no relative degree of authenticity, while there may be for reliability. The status of being authentic, however, can change at any moment as a result of residual effects of an action or migration that has been performed on the record over time. This is the case for digital data as well. By contrast, *authentication* is the process of guaranteeing the authenticity of a record[6]. If authenticity is the status of being authentic, then authentication is the action or set of activities that demonstrate that something is authentic.

When it is created, a record has two indispensable components: its content and the medium to which that content is affixed. With traditional paper records, the content of a record could not be separated from its medium. In the case of an electronic record, however, its content can be separated from the original medium and transferred to another medium or even to multiple other media. Even maintaining the same type of medium, an electronic record can be migrated to another hardware and software environment, thus effectively breaking the bond between content and medium. Due to the physical separation of the content from the media, as well as the various ways in which the integrity of the record's content can be compromised during the migration processes, the authenticity of the record is vulnerable. To address and overcome this vulnerability, increasing emphasis is being placed in many communities on the development and implementation of authentication processes to ensure and demonstrate the authenticity of the record. Authentication processes have always included both methodological and procedural techniques for assuring authenticity, although with traditional records, these techniques have tended to be more implicit than explicit, for example, through demonstration of an unbroken chain of custody for a record and through archival description[7]. There has been

increasing concern, therefore, about understanding (i.e., identifying and defining) the quality and processes associated with authenticity and authentication of information objects within the digital environment.

II. Authenticity and the InterPARES Project

In recent years, along with the rapid growth of electronic communications and information systems, digital records and data have presented new challenges and opportunities to a variety of communities of records. For example, the legal community is concerned that digital records are legally reliable as evidence. Although attorneys on either side of a case may interpret the record differently, the records themselves must somehow be demonstrated as being as authentic when we retrieve them, as they were when they were first stored and preserved. In healthcare, it is critical that digitally stored x-rays retrieved perhaps, in connection with a court case, or to evaluate a treatment decision, are identical in resolution and color when we retrieve it, as when they were stored and preserved. In computer network communication systems, it is important to establish the security of a transmission, a message, a station, or an originator, by ensuring that the sender transmits a message only to an intended receiver and that the message has not been altered in route. To the archival community, the significance of the description, identification and preservation of digital materials is increased as a result of the evidence-based approach to the management of records.

These needs and concerns raise several research questions concerning the establishment of the authenticity of digital records and data:

> Which intellectual and technical elements of data and records are essential for ensuring authenticity in different communities of practice?
>
> Can these the requirements for ensuring the authenticity of data and records be applicable across jurisdictional and technological boundaries?
>
> How should authentic data and records be maintained over time?

By identifying the requirements for ensuring authenticity, the InterPARES project also hopes to answer these questions: What is a record and what is data?

The overall focus of the project is the long-term preservation of vital organizational records and critical research data created or maintained in electronic systems and which must be preserved permanently for administrative, legal or cultural reasons. The InterPARES Project is a collaborative effort among fourteen countries to develop strategies, policies and standards of authenticity and preservation of electronic records within archives.

Research is divided into four interrelated investigative domains: (1) the conceptual requirements for preserving authentic electronic records; (2) appraisal criteria and methodology for authentic electronic records; (3) methodologies for preserving authentic electronic records; and (4) development of policies, strategies and standards to ensure preservation of the authenticity of those records. The goal of the first research domain, which is concerned with authenticity, is to identify the elements of electronic records which are necessary to maintain the authenticity of those records over time through an analysis of the elements of physical and intellectual form which may affect the authenticity and nature of an electronic record. Task forces in each domain are using methodologies including diplomatic analysis, structured interviews, literature reviews, systems analysis and design, and activity and entity modeling. The four task forces each focus on Authenticity, Appraisal, Preservation and Policy Development.

III. Preservation and the InterPARES Project

The importance of determining and analyzing the preservation function, institutional needs and long-term expectations of use and accessibility of electronic records underlies the research questions driving the InterPARES Preservation Task Force. The first goal of the Preservation Task Force is to identify and develop the procedures and resources required for the implementation of the conceptual requirements and criteria identified in the first two research domains. Broadly put, responses to the research questions will incorporate an examination of the present state of long-term preservation either in use, or in development; articulate an understanding of procedural and technical methods of authentication for preserved electronic records; provide data about the principles and criteria for media and storage management required for preservation of authentic electronic records; and lastly, enable the development of a statement of responsibilities for the long-term preservation of authentic electronic records.

The second goal of the Preservation Task Force is to model the preservation function and implementation, which will be based on information gathered from responses to the research questions. The institutional investigators working at the various national archival institutions will test models. Through an iterative process, results will be brought back to the International Team and will be used to further refine the models, which will then be re-tested. This process is expected to reveal certain basic principles upon which strategies, policies and standards for the preservation of authentic electronic records can be drafted.

IV. Method and Findings to Date

The project uses case studies to analyze requirements for authenticity based on an analysis of features of records and their genesis, using a research methodology, which is derived from diplomatics. Diplomatics is an analytical

method developed in Europe in the seventeenth and eighteenth centuries to determine the authenticity and reliability of historical documents. In the process of its introduction into most European countries, diplomatics grew into a very sophisticated system of ideas and methods about the nature of records, their creation and their relationships with the actions and persons connected to them and with their organizational, social, and legal context[8]. The concepts and principles of contemporary diplomatics have been applied in ongoing electronic record-keeping research[9] including InterPARES and have proven effective in identifying technical and procedural requirements for ensuring the reliability and authenticity of electronic records[10].

The InterPARES project has developed a typology of the conceptual requirements for authenticity for different types of electronic records: a Case Study Interview Protocol (CSIP) and the Template Element Data Gathering Instrument (TEDGI). These are the tools used to gather the empirical data, and perform diplomatic analysis of electronic records and systems to create the electronic record typology. The CSIP is the primary instrument to gather the empirical data. The CSIP will then provide the data that researchers will need to populate the Template for Analysis elements for each case study. These protocol instruments have been devised by the Authenticity Task Force to ensure that interviews carried out by the InterPARES case studies are conducted under comparable conditions at each institution. Currently, case studies are being conducted with a variety of institutions in Canada, the United States, Europe (Italy, 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). Additional information may also come from supporting documentation provided by the interviewee, additional comments made by the interviewee, external documentation from or about the case study system or organization or other identifiable sources.

To date, twelve case studies for round 1 and nine case studies for round 2 have been completed or are underway. The analysis of case studies focuses on the specific characteristics and function of ensuring authenticity in the preservation of digital data and records. Among the case studies, there are the multiple case studies that have the similar function and purposes with different situational contexts. For example, there are six registration systems being conducted in six different institutions in five different countries. There are five student records systems in five universities in three countries. The case studies with the same function are examined to identify whether the requirements for ensuring authenticity are applicable across juridical, technological, functional, and cultural contexts.

IV. Implications for Further Research

The results of the InterPARES project will be used as the basis for developing further research on electronic record-keeping systems. A methodological typology derived from a variety of case studies in real-life settings will be the basis for developing further data collection instruments and refining data analysis methods, which can then be applicable across electronic record-keeping systems. An in-depth analysis of different communities of practice would yield more insight into the ways that authentic data and records can be understood, used and managed and how common requirements of ensuring authenticity can be shared across jurisdictional and technological boundaries. As a result of the InterPARES project findings, it is hoped that standards establishing authenticity of electronic records will be developed that will be applicable across many communities of practice now and in the future.

Findings from each investigative domain are expected by December 2001. However, given the depth of the problem domains and the ongoing iterative process of designing, testing and analyzing the various requirements and methodologies, it is anticipated that research will continue beyond this date. Stay tuned; the results should be quite exciting.

Acknowledgments

The authors gratefully acknowledge Dr. Anne J. Gilliland-Swetland and Dr. Michéle V. Cloonan for their comments and encouragement while preparing this article. They also 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

1. The authors are Ph.D. students in the Department of Information Studies at the University of California, Los Angeles and participants in the InterPARES Project.

2 Records are recorded information in any form, including data in computer systems created or received and maintained by an organization or person in the transaction of business or the conduct of affairs and kept as evidence of such activity. Digital records are created or received and maintained in digital form by individuals or agencies in the course of conducting business.

3 Duranti, Luciana. "Reliability and Authenticity: The Concepts and Their Implications." *Archivaria,* 39 (Spring 1995): 5-10.

4 Dollar, Charles M. *Authentic Electronic Records: Strategies for Long-Term Acces*s. Cohasset Associates, Inc. Chicago, Illinois: 1999, p. 54.

5 Duranti, 1995.

6 Bearman, David and Jennifer Trant. "Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process." *D-Lib Magazine* (June 1998). Available from http://www.dlib.org/dlib/june98/06bearman.html, August 1, 2000.

7 Gilliland-Swetland, Anne J. *Enduring Paradigms, New Opportunities: The Value of the Archival Perspective in the Digital Environment* (Washington, D.C.: Council on Library and Information Resources, 2000).

8 Duranti, Luciana and Eastwood, Terry. "Protecting Electronic Evidence: A Progress Report on a Research Study and its Methodology", *Archivi & Computer,* 3 (1995): 213-250.

9 These are: the University of British Columbia's *The Preservation of the Integrity of Electronic Records Project*. Available http://www.slais.ubc.ca/users/duranti, August 1, 2000; the InterPARES Project international Research *on Preservation of Authentic Records in Electronic Systems.* Available http://www.interpares.org, August 1, 2000; and the U.S. Department of Defense's *Records Management Task Force Project.* Available http://jitc-emh.army.mil/recmgt, August 1, 2000.

10 Duranti, Luciana, MacNeil, Heather, and Underwood, William. "Protecting Electronic Evidence: A Second Progress Report on a Research Study and Its Methodology." *Archivi & Computer*. 6 (1996): 37-70; Gilliland-Swetland, Anne J. and Eppard, Philip B. "Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project. *D-Lib Magazine* (July/August 2000). Available from http://www.dlib.org/dlib/july00/eppard/07eppard.html, August 3, 2000.

* Paper presented at IASSIST 2000 (Chicago, 7-10 June 2000).