# The Authenticity of Electronic Records: The InterPARES Approach

*Luciana Duranti and Jean-François Blanchette*
*School of Library, Archival and Information Studies*
*University of British Columbia*
*Vancouver, Canada*

## Abstract

The InterPARES Project began in 1999 with the purpose of developing the theoretical and methodological knowledge essential to the permanent preservation of authentic digital records and to formulate model policies, strategies, and standards capable of ensuring that preservation. Its first phase, which was concluded in 2001, produced requirements and methods for the creation, maintenance, selection, and preservation of authentic digital records, mostly generated in the course of administrative activities. The second phase, which began in 2002, studies records created in the course of artistic, scientific, and e-government activities. Images are thus an important focus of the on-going research, which involves specialists from the film and computer industries, archivists, as well as photography and film scholars. This session will present the findings of the first phase of InterPARES that are relevant to the preservation of the authenticity of digital images and will describe case studies that are being carried out in the second phase of InterPARES to meet the challenge of the authentic preservation of records generated in the course of artistic, scientific, and e-government activities.

## I. Introduction

The International Research on Permanent Authentic Records in Electronic Systems, known as the InterPARES project, began in 1999 and completed in 2001.[1] Its goal was to develop the theoretical and methodological knowledge essential to the permanent preservation of authentic records generated and/or maintained electronically, and, on the basis of this knowledge, to formulate model policies, strategies and standards capable of ensuring that preservation. The first phase of InterPARES was set out to deal with digital records mandated for accountability and administrative needs, which are usually created in very large databases and document management systems. The creation, maintenance and use of this type of records by the organization producing them are highly controlled, thus the InterPARES research has focused on the preservation of authenticity after the records are no longer needed by the creating body.

The research used concepts and methods from a variety of disciplines, including diplomatics, archival science, law, computer science, computer engineering, and statistical sciences. The team included co-investigators from the public and private sectors of Canada, United States, United Kingdom, Ireland, Sweden, Netherlands, France, Portugal, Italy, Australia, China, and Hong Kong. The intellectual mediation and integration that occured among disciplines and cultural traditions are expressed in the project's glossary of terms.

## II. Fundamental Concepts and Assumptions

The work began with a definition of the fundamental concepts, in consideration of the interdisciplinarity of the project and the tendency of the disciplines involved to borrow terms from each other attaching them quite different meanings. The terms on the use of which the researchers needed to agree at the outset were "record" and "authenticity".

*Record* was defined as any document made or received in the course of activity as a means and instrument for it, and set aside for action or reference. An electronic record was defined as a record maintained for use in electronic form. In order to distinguish records among all other kinds of information that may reside in a digital system, the research team named several identifiable characteristics, deriving from the fact that a record can be viewed as a complex of elements and their interrelationships. First and foremost, a digital entity is a record if it has a fixed form. A record is considered to have a fixed form when its binary content, including indicators of its documentary form, is stored in a manner that ensures it remains complete and unaltered, and when the message of the record is capable of being rendered with the same documentary form it had when it was first set aside. In addition to a fixed form, a digital entity that is a record must have an unchangeable content, an explicit linkage with other records within or outside the digital system through a classification code or some other unique identifier, an identifiable administrative context, an author, an addressee and a writer. Finally, a digital entity that is a record must participate in or support an action either procedurally or as part of the decision making process.

*Authenticity* was defined as the trustworthiness of records as records, as distinct from reliability, which is the trustworthiness of a record as a statement of fact and the exclusive responsibility of the record creator, rather than of its preserver. An authentic record is a record that is what it purports to be, immune from corruption or tampering.

Authenticity is not to be confused with *authentication*, which is a declaration of authenticity, resulting either by the insertion, the addition or the attachment of an element, a statement or a component to a record, which allows one to verify that the record is what it purports to be at that point in time. Authentication does not establish authenticity over time.

In both archival theory and jurisprudence, records that are relied upon by the creator in the usual and ordinary course of business are presumed authentic. However, digital information technology creates significant risks that electronic records may be altered, inadvertently or intentionally. Therefore, in the case of records maintained in electronic systems, the presumption of authenticity must be supported by evidence of it. In order to assess the authenticity of an electronic record, one must be able to establish its identity and to demonstrate its integrity.

The *identity* of a record comprises the names of its author, addressee, writer and originator, its date, the name of the action or matter, its status of transmission (i.e. original, draft or copy), its relationship to the other records of the same creator, and the indication of attachments. Knowledge of all these attributes is essential to establish the identity of any record.

The *integrity* of a record relates to its wholeness and soundness: a record has integrity when it is essentially intact and uncorrupted. This does not mean that the record must be precisely the same as it was when first created for its integrity to exist and be demonstrated. Even in the paper world, with the passage of time, are subject to deterioration, alteration and/or loss. In the electronic world, the fragility of the media, the obsolescence of technology and the idiosyncrasies of systems likewise affect the integrity of records; therefore, there is not such a thing as an uncorrupted record. *When we refer to an electronic record, we consider it essentially intact and uncorrupted if its identity is clear and the message that it is meant to communicate in order to achieve its purpose is unaltered.* This implies that its physical integrity, such as the actual value of its constituting bit strings, may be compromised, provided that the articulation of the content and its required elements of form remain the same. The attributes that constitute the identity and integrity of a record may be explicitly expressed in an element of the record or in metadata related to the record, or be implicit in its various contexts.

The application of the two fundamental concepts of record and authenticity was based on two assumptions. The first is that the authenticity of records in live systems is threatened during transmission across space (i.e., person-to-person communication) and time (i.e., maintenance by the creating body for future reference), especially when this involves migration from an obsolescent to a new technology. The second is that it is not possible to preserve an electronic record, but only the ability to reproduce it. Further, it is virtually impossible to deliver any preserved electronic record in such a way that none of its elements have changed.

To attest the authenticity of a preserved electronic record, then, is to demonstrate that no essential element of the record has changed. This requirement can be satisfied only if the preservation function is exercised in such a way that any changes that do occur are identified and documented. This can only be accomplished if one knows what the elements of the record were when the record was selected for preservation. After that, one faces the need to demonstrate that none of the changes affected the ability to prove the identity and integrity of the record.

## III. Methodology and Outcomes

To achieve the project goal and to address the complex variety of issues that affect the permanent preservation of authentic electronic records, the investigation was divided into four interrelated domains, each representing a research objective, supported by a dedicated interdisciplinary and multicultural task force, and including a specific set of research questions. The domains were:

1) *Authenticity:* conceptual requirements for the preservation of authentic electronic records;
2) *Appraisal:* appraisal criteria and methods for selecting authentic electronic records to be permanently preserved,
3) *Preservation:* methods, rules and procedures for the permanent preservation of authentic electronic records;
4) *Strategy:* principles that should guide the development of international strategies and standards for the long-term preservation of authentic electronic records, and criteria for developing from them national and organizational policies and strategies respecting cultural diversity and pluralism.

### Authenticity Task Force

The task force responsible for formulating conceptual requirements for authenticity established an analytical framework for understanding existing and future records in electronic systems by developing a "Template for Analysis" according to diplomatic concepts and methods. The Template is a decomposition of an electronic record into its constituent elements: it defines each element, explains its purpose, and indicates whether, and to what extent, that element is instrumental in verifying the authenticity of the record over the long term.

To populate and test the validity of the template, the task force has conducted case studies of digital systems that contain, generate, or have the potential or possibility to create electronic records. The studies include large databases used to manage, for example, student records, patent granting, securities or bank transactions; document management systems used to support agency-wide administrative functions, such as the drafting and management of procedures, as well as specific operational functions, such as the issuing of permits for the transportation of hazardous waste or the conditional release and pardon of criminal offenders; geographic information

systems, such as land data systems; and web application systems, such as trademarks systems.

The instrument for conducting the case studies was a "Case Studies Interview Protocol" developed from the template and refined after each round of case studies on the basis of the data resulting from them. The whole process was guided by grounded theory, a method for discovering concepts and hypotheses and developing theory directly from the data under observation. This means that cases were selected according to their potential for helping to expand on or define the concepts or theory that had already been developed: data collection and analysis proceeded together.

After including the case studies results in the "Template Data Gathering Instrument," which maps the responses to the interview questions to the elements of the Template for Analysis, a diplomatic analysis of each case study was conducted. If the electronic systems examined contained records and, the analysis determined:

- Whether the elements of the records were brought together and how;
- Whether they manifested themselves in a way similar to traditional records;
- Which elements the creating organization considered essential for verifying the records' authenticity;
- What kind of procedural controls exercised over the system and the records contained in it supported the organization's presumption of authenticity,
- What type of records the system contained.

From the understanding developed in the course of this work, the task force developed conceptual requirements for the preservation of authentic electronic records, which will be discussed in the next section.

**Appraisal Task Force**

The task force responsible for developing appraisal criteria and methods for electronic records that respect the authenticity requirements analyzed the methods and procedures employed by archival institutions for the appraisal of electronic records and developed activity models of the appraisal function for electronic records. A major benefit of its work is the specification of the kinds of contextual information that needs to be gathered during appraisal. However, all the other steps involved in conducting selection of electronic records, including timing, location, agents, manner and feasibility, were modeled, and several case studies were walked through the modeled appraisal process for the purpose of analyzing the outcome.

**Preservation Task Force**

The task force charged with identifying and developing the preservation procedures and resources required for implementing the outcomes of the first two domains developed a formal model of the process of preserving electronic records, a template for applying the model to specific sets of records, a model of the entities that are involved in preserving electronic records, and guidelines that institutions and organizations can use to articulate comprehensive and coherent frameworks to guide the development and operation of a preservation system specifically tailored to the records each institution is responsible for preserving. Central to the entire preservation model is the concept of what it means, at an empirical level, to preserve an electronic record.

**Strategy Task Force**

The task force responsible for developing a framework for the formulation of international standards and national and organizational policies and strategies developed a methodology and a procedure for the distillation of principles and criteria guiding the formulation of standards, policies and strategies from the findings and final recommendations of the three other task forces. The procedure heavily involved the national research teams. This represented one of the most delicate point of the research, when the universal concepts, principles and methods developed by internationally constituted task forces were brought into specific national, organizational and cultural realities and so, contextualized.

## IV. Benchmark and Baseline Requirements

As mentioned earlier, the first task force established benchmark requirements supporting the presumption of authenticity of electronic records maintained by their creator. The records affected by these requirements can be distinguished in two categories. The first category comprises those records that exist as created, having not undergone processing that has altered their documentary form, architecture or any part of their technological context. The second category comprises those records that result from a migration process from an obsolescent technology to a new one.

Once one has assessed the evidence of the identity and integrity of the records of the creator, one can make a presumption of their authenticity based upon how many of the requirements have been met and to what degree. The requirements are, therefore, cumulative: the higher the number of satisfied requirements, the higher the presumption of authenticity. The degree to which an individual requirement is satisfied also affects the degree of presumption. This is why these requirements are termed "benchmark" requirements. Where there is an insufficient basis for a presumption of authenticity, a verification of authenticity will be needed. Unlike the presumption of authenticity, which is established on the basis of the requirements, this verification involves a detailed examination of the records themselves in all of their contexts. Methods of verification include, but are not limited to, a comparison of the records in question with copies that have been preserved elsewhere or with backup tapes, textual analysis of the record's content, a study of audit trails over time, and the testimony of a trusted third party.

It is an assumption of the task force that the records are presumed or verified authentic in the appraisal process by the entity responsible for their preservation. Thus, the maintenance of their authenticity after that process is the exclusive responsibility of the preserver, who must carry

forward the records by reproducing them, and authenticating the copies so produced. The production of authentic copies is a complex endeavor, which must be regulated by a second set of requirements. Unlike the benchmark requirements for authentic electronic records, all the requirements for the production of authentic copies of electronic records must be met before the preserver can attest to the authenticity of the copies in its custody. This is why this second set of requirements is termed "baseline" requirements.

Traditionally, the official preserver of the records has been the person entrusted with issuing authentic copies of them. For a copy to be considered authentic, the preserver needed simply to attest that the copy conformed to the record being reproduced. With electronic records, the difficulties related to preservation make it prudent for the preserver to produce and maintain documentation of the activity of reproduction to support its attestation of authenticity. Thus, an electronic copy of an authentic electronic record is authentic if attested to be so by the official preserver and if such attestation is supported by the preserver's ability to demonstrate that all the requirements for the production of authentic copies have been satisfied. By virtue of this attestation, the copy is deemed to conform to the record it reproduces until proof to the contrary is shown.

The conceptual benchmark and baseline requirements apply to any type of electronic record. Among the systems analyzed as case studies, all those containing records implemented at least two of the benchmark requirements. The main concern of the research team was, however, that systems which, because of their function in the organization, are meant to contain records attesting to specific actions and transactions, such as universities' student information systems, and several government registration and inventory systems, given the fluidity of their content, did not contain records but only data, and made therefore impossible to implement the requirements.

In fact, the most significant, if not unexpected, finding of the case studies was that most large databases used in electronic governance and administration are unable to serve accountability purposes, let alone to allow for the verification of the authenticity of the information they contain. A second important finding is that the best method of ensuring ongoing authenticity of electronic records is external to the records themselves and involves a tight control on record-making and record keeping procedures and on the flow of metadata into the record's formal elements, rather than digital authentication measures, which have been shown to hamper long-term preservation of authentic records.

## V. The InterPARES 2 Project

In the course of the research, it became apparent that the solutions identified for the long-term preservation of the administrative and legal records produced in large databases and document management systems are not sufficient for ensuring the continuing authenticity of records whose creation and form are discretionary, and which are generated by more complex systems.

Therefore, the second phase of InterPARES, which began in January 2002, focuses on the reliable creation, as well as on the authentic preservation, of records in dynamic, and interactive systems, in the course of artistic, scientific, and e-governmental activities. This is especially urgent as, on the one hand, the products of such activities originate in digital forms not controlled by any existing standard; and on the other hand, both those who produce and those who use complex information systems appear to be uniquely concerned with the "here and now", showing a great disregard for the permanent preservation of a recorded authentic memory of our times.

**Structure of research**

The research has been divided into three *domains of inquiry*, each addressing a specific set of research questions:

1) The *Records Creation and Maintenance* domain will investigate the nature of the records under examination and the process of their creation;
2) The *Authenticity, Accuracy and Reliability* domain will study the concepts of reliability, accuracy and authenticity, as they are understood in the various disciplinary areas involved in the research;
3) The *Methods of Appraisal and Preservation* domain will test existing appraisal and preservation methods on the records in question and develop and test new ones.

The grouping of researchers into Domain Task Forces promotes interdisciplinary collaboration where issues of concern are common to different types of activities.

Each domain comprises three *focuses*, each addressing a type of records: (1) the records of artistic activities, (2) of scientific activities, and (3) of e-governmental activities. The simultaneous grouping of researchers into Focus Task Forces allows for conducting case studies of records creators in the context of each type of activity, analyzing their records from creation to their ultimate appraisal and preservation by archival institutions or programs.

In addition, four *cross-domains* have been established; in each a Cross-Domain Research Team addresses a key objective common to all domains: (1) the *Terminology Research Team* will control the use of terms and related definitions in all areas of the research; (2) the *Policy Research Team* will analyze the existing policies and strategies in each domain and focus of inquiry. It will distill from the findings and products of the task forces policies, strategies and guidelines for the reliable and accurate creation and maintenance of the records under examination, and for their authentic preservation within the context of each activity and culture generating them; (3) the *Description Research Team* will develop guidelines for the intellectual control and archival description of all types of records studied; (4) the *Modeling Research Team* will develop models of the records lifecycle in each of the focuses.

The project is interdisciplinary inasmuch as its goal and objectives can only be achieved through the contribution of several disciplines. One of the methods chosen to develop

ways of creating records whose accuracy and reliability can be protected overtime is to conduct exploratory case studies in each of the areas of activity identified. In order to analyze the nature, characteristics, behaviour, relationships and process of creation of the interactive and dynamic records produced in the course of artistic, scientific and electronic government activities, InterPARES researchers need a deep understanding of those activities, their purpose, their phases and the component actions, their byproducts and their structure, and their context, but also their technological environment and their use. Such case studies are conducted by multidisciplinary teams of archivists, computer engineers, and relevant specialists, e.g., musicologists, scientists, etc The results of the case studies are analyzed using methodologies developed in the context of a variety of disciplines, among which, text analysis, diplomatic analysis, statistical analysis, etc.

Examples of case studies conducted in, respectively, the artistic, scientific, and e-government focuses are described in the following sections.

**Electronic Café International**

The Electronic Cafe Intentional (ECI) is a multimedia international network for showcasing creative, multi-cultural, multi-disciplinary, collaborative telecommunications. ECI experiments have involved several museums as telepresence sites (including the Los Angeles Museum of Contemporary Art and the San Jose Museum of Art).

The first to build an international network dedicated to the creation of telecollaborative arts, ECI has accumulated a large and historically important collection of assets throughout the years. The collection of electronically stored documentation represents a large part of the history and story of how artists migrated to cyberspace (beginning decades before WWW, and continuing on into the present). After nearly three decades of aesthetic inquiry, research and development, and the production of hundreds of telecollaborative arts projects, ECI now holds thousands of hours of video and digitally stored documentation, including many of the domestically and internationally recognized "classics"" intrinsic to the telecollaboration genre. Their collections hold more than 3,000 hours of video, optical disks, audio recordings, computer back-up media, additional electromagnetic storage media, equipment, text, drawings, paper documents, photos and other types of images — documenting hundreds of artists engaged in telecollaborative works.

This case study thus deals with a wide variety of media types that now pose the problems of aging and obsolescent formats. It highlights the problems posed by interactive and dynamic records many years after they were initially created. Digital repositories are currently being set up independent of knowledge of the needs of aging records, and tend to be based upon thinking of problems beforehand. The ECI material will allow repository design using a more bottom-up than a top-down approach, with design based upon real problems that have shown up in similar aging records. It is expected that this activity will foreground issues that need to be dealt within record systems at the point of creation (now and in the future).

**Preservation and Authentication of Electronic Engineering and Manufacturing Records**

Various components of the U.S. Government are involved in the design, development, manufacture and use of complex systems. Many of these systems have very long life spans. In addition, there is an enduring requirement to retain records of the systems and their components. There may be need in the future to produce an identical system or to produce one or more parts for such systems. Such needs can only be addressed by retaining records that include both engineering specifications and manufacturing process methods. In the current environment, critical records, such as 3-dimensional models of solid artifacts and records of the process of manufacturing of such artifacts, exist and can only exist in digital form.

Representing or reproducing a highly engineered system or component requires precise, accurate, reliable and authentic records. The Electronic Records Archives Program of the National Archives and Records Administration is engaged in a collaborative research project to develop and test prototype trusted computational environments for the creation, exchange and preservation of digital manufacturing models of machined piece-part shape and machining process knowledge. These models and process knowledge are records of manufacturing whose preservation, authentication and reliability are important for long term needs derived from the original purpose for which they are created.

Because of the diverse nature of the creation, exchange and storage environments, multiple perspectives on the requirements for authenticity will be encountered. The first objective of this case study is thus to understand the conceptual, logical and physical requirements on the multiple abstract languages used to model part shape and manufacturing process knowledge.

The project is planned in two phases. The first phase will prototype the preservation, retrieval, and authentication of digital models of machined piece parts. The second phase will build on the results of the first, expanding the prototype to a significantly more complex model and to include records of the process which produced the system or part depicted in the model. The prototype for both phases will involve all basic functions of an archival system (as described in the OAIS standard), from submission by the producer to dissemination from the archives. The project will apply the findings of the InterPARES 1 project to support the authenticity of the records preserved in the prototype and will assess the reliability of the records to transmit required knowledge both of the systems and the processes over time.

**Irish Revenue On-Line System**

This case study centers on the roll out and expansion of the award winning Revenue On-line System (ROS) currently in operation by the Revenue Commissioners of Ireland (Revenue). In the move towards e-government, Revenue has created an on-line system that enables the generation,

maintenance, access and preservation of electronic based tax and other records, in a secure and appropriate environment. Under Revenue's current corporate plan, it is envisaged that sums in excess of 40 billion Euros will be collected, of which ROS will directly collect 5 billion Euros. Revenue hopes to have 50% of all tax returns and transactions occurring on-line by 2005.

ROS is an Internet based system allowing taxpayers to file a series of tax returns on-line and view details of revenue account information in a secure environment. It is also possible to transfer funds electronically to settle tax liabilities. The system allows taxpayers and agents to download ROS software to enable tax returns to be processed off line before connecting on-line to transfer the details electronically. The ROS system can also integrate with at least ten other third party software solutions necessitating sophisticated interfaces for the secure transfer and authentication of data and records.

The ROS system is increasing in functionality and complexity on an on-going basis. The first phase of ROS was released in September 2000 and an additional 7 phases have been released to date, augmenting both the range of tax forms that can be processed by the system, and the range of services available. Two further phases are expected to be released before the end of 2004.

The ROS system is one of the very few Irish government websites actually creating, managing and preserving electronic records. At the level of record creation, this study will examine metadata models and standards (if any) used for information creation and exchange, as Revenue offers XML DTDs of all its forms for developers to ensure any third party product can interface with the system. At the level of preservation, this case study will look at the use of digital signatures and public key certificates in an operational environment in addition to encryption and the associated challenges of preserving encrypted data (ROS includes a certificate management infrastructure and is itself a certifying authority for its digital certificates). At the management level, this study will examine how and if ROS provides clear and unambiguous authentication measures, guarantees regarding the integrity and confidentiality of information supplied by users, and the non-repudiation of tax returns — all aspects of the archival record.

## VI. Conclusion

Administrative transparency, historical accountability, long term legal requirements and the protection of culture require that governments, universities and industry look beyond the present and consider the political, social and economical implications of entrusting all knowledge to digital systems destined to quick obsolescence before having in place strategies and standards for their continuing authentic preservation. The most important achievement of InterPARES has been to get experts from all sectors to work together in a sustained, intense, consistent and integrated way, irrespective of differences in culture, discipline and intent. But this is only the beginning of a necessary worldwide effort.

## References


1. InterPARES, *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (2002).


## Biography

Luciana Duranti is the Director of the InterPARES project and has taught in the archival program at the School of Library, Archival and Information Studies since 1987. Professor Duranti holds a Doctorate in Arts (1973) and graduate degrees in Archival Science from the University of Rome (1975), and in Archivistics, Paleography, and Diplomatics from the School of Archivistics, Paleography and Diplomatics of the State Archives of Rome (1979). In addition to her university responsibilities, she has been the President of the Society of American Archivists for the year 1998-99, and is active nationally and internationally in several other archival associations; she publishes widely on archival history and theory, and on diplomatics.


Prof. Luciana Duranti
School of Library, Archival and Information Studies
University of British Columbia
301-6190 Agronomy Rd, Vancouver (BC) Canada V6T 1Z3
Tel.: +1 604 822-2587; Fax.: +1 604 822-6006
Email: luciana@interchange.ubc.ca
Web: http://www.interpares.org