Le projet de recherche Interpares sur l'authenticité des documents électroniques (Luciana Duranti, Professeur et présidente, Archival Studies Program, Université de Colombie Britannique, Canada)

Archivistique

Documents électroniques

Journées internationales Archivage à long terme des documents électroniques Paris, France 8-9 mars 2001

Le projet de recherche Interpares sur l'authenticité des documents électroniques

Luciana Duranti, Professeur et présidente, Archival Studies Program, Université de Colombie Britannique, Canada

The Authenticity of Electronic Records: The InterPARES Approach

Reliance on the authenticity of records is at the root of decision-making and scholarly endeavor. While everyone acknowledges that it is vital to know that the records we use for action and reference or as sources for research are trustworthy, individuals and organizations have little hesitation in adopting, for carrying out their activities, increasingly complex, fast changing computer technology, which is making the authenticity of electronic records very hard to preserve and demonstrate.

For their common transactions with citizens and businesses, governments are progressively replacing fixed-format paper forms with dynamically configured information objects, such as interactive web sites, and with document type definitions that allow the same stored content to be configured in different ways for different purposes. Similarly, self-certified digital submissions are replacing internal control procedures. Commercial businesses, as well as scholarly undertakings, use systems that convey direct experience of products and natural objects, such as their smell. Both government and business increasingly rely on systems that are shared among the participants in transactions. Creative industries and individual artists produce audio and visual materials that lack the security gained through the use of traditional media for publication and dissemination. This threatens the ability to identify the author of the records, the copyright owner, and the context of record creation, and the accessibility, intelligibility and integrity of the records.

Surmounting these threats requires a deep understanding of the nature of the new records, and the development of theory and methods capable of guiding the formulation of policies, strategies and standards for preserving their authenticity while protecting cultural diversity and pluralism. Understanding the records produced by complex systems requires a well developed and tested theory and methodology for analyzing the systems and their by-products. It demands the kind of interdisciplinary and multicultural collaboration that has been established in InterPARES (International Research on Permanent Authentic Records in Electronic Systems), a project that began in 1999 and is nearing the completion of its first phase. The goal of this project is 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 is set out to deal with electronic records mandated for accountability and administrative needs. In most countries, such records are the majority of those selected for permanent

preservation. They constitute a high priority for both the public and the private sector. They are usually created in very large databases and document management systems. The authenticity of administrative records has been a concern of most juridical systems, which have explicitly stated requirements for attesting and verifying it on records generated on traditional carriers, like paper and microform; these requirements can be used as benchmarks for developing new requirements for the same records in electronic form. 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. A previous research project carried out by the same University of British Columbia team in collaboration with the U. S. Department of Defense, had formulated the requirements for the creation, maintenance and use of these records by the body producing them. This work provided a theoretical foundation for the InterPARES project. However, it is becoming apparent that technological developments are interfering even with the procedures and forms prescribed for legal records, that increasingly decision making is based on records whose creation and form are discretionary, and that great concern is developing about the trustworthiness of record types generated by more complex systems. Therefore, the second phase of InterPARES, which will begin in January 2002, will focus on reliable records creation as well as on authentic maintenance, use and preservation of records made or received and set aside in dynamic, interactive, performance, sensory and experiential systems, including those produced in the course of creative and performing activities.

To achieve the project goal of the first phase of InterPARES, and to address the complex variety of issues that affect the permanent preservation of authentic electronic records, the research objectives were divided into four interrelated domains of investigation. Each domain represents a research objective that is supported by a dedicated interdisciplinary and multicultural task force and includes a set of research questions. The objective of Domain I is to identify the conceptual requirements for the preservation of authentic electronic records. The objective of Domain II is to develop appraisal criteria and methods for selecting authentic electronic records to be permanently preserved, respecting the conceptual requirements identified in the first domain. The objective of Domain III is to develop methods, rules and procedures for the permanent preservation of electronic records according to the conceptual requirements identified in the first domain. The objective of Domain IV is to define the principles that should guide the development of international strategies and standards for the long-term preservation of authentic electronic records and the criteria for developing from them national and organizational policies and strategies respecting cultural diversity and pluralism. To achieve the objective of this last domain, national teams support the task force. They are responsible for taking the universal principles and criteria drafted by the task force and reviewing them in relation to their own juridical/administrative/cultural circumstances and provide feedback to the task force. This process is iterative and the final document will include baseline and specific recommendations, as well as universal principles and context-based criteria.

The research uses concepts and methods from a variety of disciplines, including archival science, law, computer science, computer engineering, statistical sciences, and especially diplomatics, a science developed in France in the 17th century to ascertain the authenticity of records of unproven origin. Taught since then in the context of history and law, its concepts and methodology are consistent with those disciplines and are at the very heart of archival science.

The team includes 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 occur among disciplines and cultural traditions are expressed in the project's glossary of terms.

The work began with a definition of the basic concepts involved. This was especially important given the

interdisciplinarity of the project and the tendency of computer related disciplines, archival science, diplomatics and law to borrow terms from each other attaching them quite different meanings. The fundamental 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. Thus, every record is a document, but not every document is a record. A document was defined as recorded information, where information is any aggregation of data intended for communication across space or time, and data are the smallest indivisible units of meaning.

An electronic record was defined as a record maintained and used 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. A digital or electronic entity is a record if

* It has a documentary form. Documentary form is defined as the rules of representation according to which the content of a record, its immediate administrative and documentary context, and its authority are communicated. It includes intrinsic elements, which make up the internal composition or articulation of the record, and extrinsic elements, which make up the features of a record's external appearance.

* It has a fixed form, that is, 1) the binary content of the record, including indicators of its documentary form, is stored in a manner that ensures it remains complete and unaltered, and 2) the content of the record is capable of being rendered with the same documentary form it had when it was first set aside.

* At least three persons are involved in its creation They are the author, that is, the physical or juridical person having the authority and capacity to issue the record, the addressee, that is the person for whom the record is intended, and the writer, that is the person responsible for the articulation of the discourse. The three conceptual persons may be one physical person fulfilling different roles.

* It participates in or supports an action either procedurally or as part of the decision making process. This means that, although the record is created as a means for action, its creation can be either mandatory or discretionary.

* It has a stable content.

* It has an identifiable context. The relevant contexts are juridical/administrative, provenancial, procedural and documentary; and

* it has an explicit linkage (i.e., archival bond) with other records within or outside the system, through a classification code or some other unique identifier.

The second concept defined was authenticity. Authenticity was defined as the trustworthiness of records as records. Authenticity must not be confused with reliability, which is the trustworthiness of a record as a statement of fact, that is, as to its content, and which is the responsibility of the record creator. Rather, an authentic record is a record that is what it purports to be, immune from corruption or tampering. Authenticity is to be distinguished from authentication, which is only a means of proving that a record is what it purports to be at a given moment in time. It usually manifests itself in a declaration of authenticity, resulting either by the insertion or the addition of an element or a statement to a record, and the rules governing it are established by legislation. For example, authentication may be provided by a digital signature, which is however useless as it regards attesting to a record's continuing authenticity.

Authenticity of a record is assessed in relation to its identity and integrity. The identity of a record is provided by its provenance (i.e., creator), author, addressee, writer, date, matter or action, and relationship to other records (i.e., archival bond). The integrity of a record is its wholeness and soundness: it implies that a record is intact and uncorrupted. This quality cannot be absolute, however. Given the passage of time and the deterioration of the physical parts of the record, there is not such a thing as an intact record, even in the paper world. Thus, it was decided that an electronic record can be considered intact and uncorrupted if its identity is clear and the message that the record is meant to communicate in order to achieve its purpose is unaltered. This implies that physical integrity, such as the proper number of bit strings, may be absent, provided that the articulation of the content and its required formal elements remain the same. Both the identity and the integrity of a record can be verified on its face, and/or on metadata linked to it, and/or on components of its context.

Once the fundamental concepts were defined and agreed upon, the Authenticity Task Force (ATF) proceeded to establish an analytical framework for understanding records in electronic systems, both existing and future types, 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 ATF has conducted case studies of digital systems that either 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 is a "Case Studies Interview Protocol (CSIP)," developed from the template and refined after each round of case studies on the basis of a statistical analysis of the data resulting from them. The whole process is guided by grounded theory, a method for discovering concepts and hypotheses and developing theory directly from the data under observation. This means that cases are selected "according to their potential for helping to expand on or define the concepts or theory that have already been developed. Data collection and analysis proceed together." After including the case studies results in the "Template Data Gathering Instrument (TEDGI)," which maps the responses to the CSIP questions to the elements of the Template for Analysis, a diplomatic examination of each case study is conducted for the purpose of establishing whether the electronic systems examined contain records and, if the answer is affirmative, to determine

* whether the elements of the records are brought together and how, * whether they manifest themselves in a way similar to traditional records, * which elements the creating organization considers essential for verifying the record's authenticity,

* what kind of procedural controls exercised over the system and the records contained in it support the organization's presumption of authenticity, and

* what type of records the system contains.

From the understanding developed in the course of this work, the ATF has developed draft conceptual requirements for the preservation of authentic electronic records that will be tested and finalized in the next few months.

The conceptual requirements are 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 for future reference), especially when this involves migration from an obsolescent to a new technology. The second assumption is that it is not possible to preserve an electronic record, but only to preserve the ability to reproduce it. In addition, it is not possible to deliver any electronic record that has been preserved, in a way that none of its elements have changed. To attest the authenticity of such a record, then, involves demonstrating 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. If such knowledge exists and the changes are documented, one has to show that none of the changes that occurred affected the ability to prove the identity and integrity of the record.

In light of the above, the ATF first established requirements for the organization producing the records that will enable the preserver of the records no longer needed by the organization but destined to permanent preservation to establish a presumption of authenticity. The records affected by these requirements include both the records that exist as created, as they have not undergone processing, and any copies of them that result from a refreshing or migration process. The presumption of authenticity for these two categories of records is to be based on the ability of the preserver (usually a trusted third party-archival program or other) to verify that the elements revealing the identity of the records are either present on their face or inextricably linked to them, and that the records have been made or received, set aside and maintained in a trusted record keeping system.

A record keeping system is defined as a set of internally consistent rules governing an organization's activities of making, receiving, setting aside, handling and maintaining active and semi-active records, and the tools and mechanisms used to implement those rules. The requirements for a record keeping system that can be considered trusted are rather detailed, but they can be summarized in a few statements. A trusted record keeping system is one that:

* implements and monitors access privileges based on defined competences;

* explicitly regulates the procedures by which the records move inside and outside the organization;

* requires the creation of a profile for each record, which is inextricably linked to it as an annotation and includes fields that allow for the verification of the record's identity (persons involved, date, subject or matter, archival bond) and integrity (handling office/officers; changes-type, date, person; migrations-date, person, certification);

* maintains an audit trail of every access to the records and every transmission;

* includes explicit procedures to prevent loss or corruption of the records;

* includes explicit migration procedures;

* includes explicit procedures for taking records out of the live system for preservation purposes;

* includes explicit procedures for transfer of inactive records to the entity competent for preservation;

* has precise requirements for storage facilities and equipment for records maintained outside the live system; and

* maintains evidence of unbroken custody of each record.

Once one has assessed the evidence of the authenticity of the records of the creator, it can make a presumption of their authenticity, or may need to undertake further analysis to verify the authenticity of the records. A presumption of authenticity will be 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, 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 for the production of authentic electronic copies. Unlike the benchmark requirements for authentic electronic records, all of the requirements included in this second set must be met before the preserver can attest to the authenticity of the electronic copies in its custody. This is why the requirements for the production of authentic electronic copies are termed ‘baseline' requirements.

Satisfaction of these baseline requirements will enable the preserver to produce authentic copies of electronic records. Traditionally, the official preserver of the records has been the person entrusted with issuing authentic copies of such records. 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 this 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 it has satisfied all of the requirements for the production of authentic copies. By virtue of this attestation, the copy is deemed to conform to the record it reproduces until proof to the contrary is shown.

The baseline requirements supporting the production of authentic copies of electronic records are the following:

The preserver should be able to demonstrate that:

1. the procedures and system(s) used to transfer records to the archival institution or program, maintain them, and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that

a) unbroken custody of the records is maintained, and

b) security and control procedures are implemented and monitored.

2. the activity of reproduction has been documented, and that this documentation includes

a) the date of the records' reproduction and the name of the responsible person,

b) the relationship between the records acquired from the creator and the copies produced by the preserver, and

c) the impact of the technology chosen for those copies on their form, content, accessibility and use;

3. in those cases where a copy of a record is known not to fully and faithfully reproduce the elements relating to its identity and integrity, such elements have been documented by the preserver, and this documentation is readily accessible to the user;

4. the archival description of the fonds containing the electronic records includes-in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts-information about changes the electronic records of the creator have undergone since they were first created.

These requirements establish grounds for a presumption of authenticity of the records kept by the preserver (i.e., an archival institution or program, or a records office within an organization): until proof to the contrary is shown, records that meet the requirements are considered authentic. However, the preserver's declaration of authenticity is only as strong as the evidence on which such declaration rests. This means that, if the preserver is unable to verify at the outset the authenticity of the records transferred by the creating body, because their identity and integrity were lost or compromised while the records were in its care, the preserver cannot guarantee the identity of the copies themselves, and can only declare that the reproductions are authentic copies of records the authenticity of which could not be proven.

The baseline conceptual 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 requirements that the creating organization must respect. 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 any records, and made therefore impossible to implement the baseline 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 tend to hamper long-term preservation of authentic records. The complete findings of the ATF and its requirements will be made public by the end of 2001.

The Appraisal Task Force (ApTF) is responsible for developing appraisal criteria and methods for electronic records that respect the authenticity requirements. To work towards this aim, the task force has conducted a review of the literature on appraisal of electronic records; 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 task and a principal expected benefit of its work is the specification of the kinds of contextual information that needs to be gathered

during appraisal. However, also all the other steps involved in conducting selection of electronic records, including timing, location, agents, manner and feasibility, are modeled, and several case studies are walked through the modeled appraisal process for the purpose of analyzing the outcome.

The Preservation Task Force (PTF), which is charged with identifying and developing the procedures and resources required for implementing the conceptual requirements and the selection criteria identified in the first two domains for preserving electronic records, is addressing this charge principally by developing a formal model of the process of preserving electronic records. In addition, it is developing related products, including a template for applying the model to specific sets of records; an entity model of the things 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.

The Strategies Task Force (STF), which is responsible for developing a framework for the formulation of international standards and national and organizational policies and strategies, has 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 task forces. The procedure will heavily involve the national research teams. This represents the most delicate point of the research, when the universal concepts, principles and methods developed by internationally constituted task forces are brought into specific national, organizational and cultural realities and so contextualised. At this time, the STF is in the process of comparing international and national standards, as well as national and organizational policies that are relevant to the work of the task forces with their drafted findings, deliverables, and recommendations. In light of the results of this comparison and after receiving the final reports of the Authenticity, Appraisal and Preservation Task Forces, the Strategies Task Force will draft its recommendations and hand them to the national teams of researchers for feedback. Upon receipt of the national teams' reports, the STF will write its final report in which the work of all InterPARES units will coalesce at the end of 2001. This report will articulate a conceptual framework for standards, policies, and strategies that will allow for the development of technologies supporting the long-term preservation of the authenticity of electronic records while respecting pluralism and protecting the wealth of our cultural diversity.

The InterPARES project has already had a substantial impact on the way of thinking about preservation of electronic records, but probably its most important achievement has been to get experts from academia, government, business and industry to work together in a sustained, intense, consistent and integrated way, irrespective of differences in culture, discipline and intent. The significant outcome of this unprecedented collaboration is the design of InterPARES 2, which will begin in January 2002, and will involve additional sectors (among others, notably, the creative and performing arts) and additional countries from five continents. The danger of losing the authentic recorded memory of our times for the next generations looms large enough to warrant a worldwide effort.