

Meeting the challenge of accuracy and authenticity

Luciana Duranti

1. The Challenge

For centuries, our presumption of the authenticity of records has been premised on the presence or absence of visible formal elements, on creation and transmission control, and on an uninterrupted line of legitimate custody. The use of digital technology has not only reconfigured those formal elements, allowed for the bypassing of production controls, and made of physical custody an elusive concept, but, first and foremost, it has eliminated the original record, that is the first complete instantiation being communicated either across space or time, thereby making impossible the verification of authenticity on the record itself.

If electronic records will ever be considered authentic as those on traditional media, the practices by which they are created, maintained, made accessible and used must be analyzed, and strategies and standards for their authentic preservation must be produced and implemented. This is the mission of InterPARES (International research on Permanent Authentic Records in Electronic Systems), a research endeavor that aims 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. At the end of its first phase, which ran from 1999 to 2001 (hereinafter called InterPARES 1), InterPARES issued, in addition to methods and activity models of selection and preservation, a series of authenticity requirements for records that, although digital, were very similar to their analog counterparts. Indeed, these requirements were based on a concept of record that has at its heart the characteristics of a) a fixed form, in that its binary content is stored so that it remains complete and unaltered, and its message can be rendered with the same documentary form it had when first set aside, and b) an unchangeable content. They were also based on a concept of record authenticity that can be neatly divided into identity and integrity. Record *identity* is conveyed by the attributes of a record that uniquely characterize it and distinguish it from other records. These attributes include: the names of the persons concurring in its

formation (i.e., author, addressee, writer and originator); its date(s) of creation and transmission; an indication of the matter or action in which it participates; the expression of its linkage to other records; as well as an indication of any attachment(s). Record *integrity* refers to the wholeness and soundness of a record. A record has integrity if it is intact and uncorrupted, that is, if the message that it is meant to communicate in order to achieve its purpose is unaltered. A record's physical integrity, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and its required elements of form remain the same. Integrity may be demonstrated by evidence found on the face of the record, in one or more of its contexts,¹ or in metadata related to the record. Given the practically immediate disappearance of the original instantiation of the digital record, these metadata necessarily relate to responsibility and accountability (e.g., the handling office and the office of primary responsibility for the maintenance of the record, if different from the former, must be among them) and to the physical changes the record has undergone due either to technological upgrading or to preservation strategies.

Accordingly, the requirements that, if respected, allowed for a presumption of record authenticity focused on metadata expressing attributes related to identity and integrity, on access privileges, protective measures to avoid loss and corruption and to deal with obsolescence of media and technology, determination of documentary forms, establishment of authenticating procedures, identification of the authoritative record, and procedures of removal of records from the system and transfer to storage.² The requirements were tested on records contained in document and record management systems and in databases in a variety of legal and cultural contexts and resulted effective.

However, increasingly, organizations and individuals have been generating records of a dynamic, experiential, or interactive nature, which will need different, and perhaps type-specific, authenticity requirements. Dynamic records depend for their content upon data extracted from databases, which may have variable instantiations. The challenge they present to those who generate and access them is their lack of fixity both in form and content, but more serious issues are raised by experiential and interactive

1. Five contexts were identified: administrative/juridical, provenancial, procedural, documentary, and technological.

2. The requirements developed by InterPARES 1 can be found on the project's website at http://www.interpares.org/book/interpares_book_k_app02.pdf

records. Clifford Lynch describes experiential digital objects as objects whose essence goes beyond the bits constituting them to incorporate the behaviour of the rendering system, or at least the interaction between the object and the rendering system. He also maintains that defining the authenticity of such objects is a much more complex problem than with raw data or traditional digital records, because it is dependent not on the ability to reproduce a copy of the object's original bit-stream, but on the capacity to recreate the environment in which that object was experienced, an activity that involves issues of intellectual property, copyright, etc.³ An interactive system is one in which each user's entry causes a response from or an action by the system. To acquire and maintain control of records in such systems, we need to ascertain a) how user input affects the creation and form of digital entities, and b) if and when the interactive system and its inherent functionality need to be preserved for those entities to remain meaningful and authentic.

Whether dynamic, experiential, and interactive digital objects are indeed records depends of course on their relationship to the activity of their creator and on the use that society makes of them. The entertainment and cultural industries as well as researchers in sciences like microbiology have a long history of creating such objects, and clearly the professionals charged with their preservation may have to face the concrete challenge of taking and keeping views of dynamic systems, recreating the environment of experiential objects, and maintaining the functionality of interactive records. These issues are further compounded when individual record creators lack the knowledge and tools to generate digital material that can be maintained across technologies. This point can be illustrated by reflecting on the challenge provided by some new habits that they have acquired. The ease with which digital entities can be manipulated has in fact given those who generate them, particularly in the creative and research sectors, a new reason for keeping them: 'repurposing'. Makers and distributors of digital music and art works, as well as designers and architects, for example, often obscure the meaning and cultural value of their products by treating their form and content merely as digital data to be manipulated to generate new products, decontextualizing them and destroying their original identity. The potentially wide dissemination of repurposed materials threatens the authenticity of

3. Lynch, Clifford. "Experiential Documents and the Technologies of Remembrance." *I in the Sky: Visions of the Information Future*, edited by Alison Scammell. London: Library Association Publishing, 2000.

records, as well as their authors' rights. Another common habit is that of buying pieces of technology off the shelf and modify and combine them in such a way that they result in proprietary systems the products of which cannot be easily moved to any subsequent technology without compromising their authenticity. Indeed, just like the only way of maintaining a digital record is to maintain the ability to reproduce it, the only way of keeping it authentic is to produce an authentic copy of it, a copy, that is, which manifests both identity and integrity.

Moreover, the concept of authenticity as defined by InterPARES 1 is no longer regarded by those who create and use these new types of records as capable of addressing their concerns, which are much more focused on accuracy, either as a separate concept or as a component of authenticity. In diplomatics, the 17th century science of the records, the methodology of which has been used as the primary way of analysing case studies in InterPARES 1, authenticity used to incorporate the concept of reliability, which is similar to accuracy and may appear, on the surface, to imply it. Reliability refers to the authority and trustworthiness of a record as representation of what it is about, that is, to its ability to stand for what it speaks of. In other words, reliability is the trustworthiness of a record as to its content. It depends upon two things: the degree of completeness of a record's form and the degree of control exercised over the procedure in the course of which the record is generated.⁴ Reliability, then, is linked exclusively to record making, although one could infer the reliability of received records on the basis of their participation in the relevant business procedure of the entity receiving it. However, reliability is the sole responsibility of the individual or organization making the record, and the lack of it cannot be blamed on the recipient of the record, notwithstanding the fact that the latter is the record creator. Reliability is a relative concept: a record may be more or less reliable according to the degree to which the rules for making it have been respected.

The concept of accuracy is not used in diplomatics or archival science. Therefore, it has not been examined to date in the course of research guided by an archival and diplomatic approach. On the basis of a variety of dictionary definitions, an accurate record is one that contains correct, precise and exact information. Accuracy of a record

4. This concept was developed by the UBC Project and is discussed in its Template 7, "How is record created reliable in the electronic environment?" (see <http://www.interpares.org/UBCProject/tem7.htm>)

may also indicate or emphasize the absoluteness of the data it reports or its perfection or its exclusive pertinence to the matter in question. Accuracy is a very relevant concept in the context of scientific activities, although one could imagine it to have a similar significance in artistic and legal activities. Its meaning varies from a discipline to another, therefore dedicated research on the meaning of this concept in each discipline that considers it central to its activities needs to be carried out for the purpose of determining how to ensure that records are created and maintained which can be considered accurate by their creators and users.

Because of all the issues raised by records lacking fixity, repurposed records, and records valued for their accuracy as much as for their authenticity, it is necessary to develop an understanding of all new digital objects, not only in the later phases of their life cycle, but from the moment of their creation. In fact, it is necessary to revisit the concept of record itself, so that both the identification and the protection of the new types will be possible. We have to consider the possibility of substituting the characteristics of completeness, stability and fixity with the capacity of the system where the entity resides to trace and preserve each change the digital object has undergone. And perhaps we may look at each instance of these new digital entities as existing in one of two modes, as an entity in becoming, when its process of creation is in course (even if such creation is ongoing), and as a fixed entity at any given time the entity is viewed. There is no doubt that knowledge and strategies must be developed that are beneficial to both the creators and preservers of these complex new materials.

Technological obsolescence, which poses a continuing challenge to the accessibility, readability and intelligibility of electronic objects, is of even more concern in the context of dynamic, experiential and interactive records. Inadequate records management practices have already precipitated the disappearance of many such records that depended upon now obsolete software and hardware for their continued existence, including entities situated in virtual environments, and other performance materials whose essential parameters were insufficiently documented to allow for their recreation. This has generated enormous difficulties for creators concerned with the long-term preservation of the unique and authoritative version of their records, engendering an urgent demand for effective and tested strategies.

To meet these challenges, the international team of researchers formed for InterPARES 1, together with additional researchers with discipline-specific knowledge, decided to initiate a second phase of its research, called InterPARES 2.

2. InterPARES 2: Intellectual Framework

InterPARES 2 began in 2002 and its completion is scheduled for the end of 2006. Its goal, objectives, structure and methodological principles have been articulated in an intellectual framework on which all co-investigators agreed.

2.1 Research goal

The goal of InterPARES 2 is to ensure that the portion of society's recorded memory that is digitally produced in dynamic, experiential and interactive systems can be created in accurate and reliable form, and maintained and preserved in authentic form, both in the short and the long term, for the use of those who created it and of society at large, regardless of digital technology obsolescence and media fragility.

2.2 Research objectives

- To develop an understanding of dynamic, experiential and interactive systems and of the materials produced and maintained in them, of their process of creation, and of their present and potential use;
- to formulate methods for ensuring that these digital objects are generated and maintained by the creator in such a way that they can be trusted as to their content (that is, are accurate and reliable) and as works (that is, are authentic);
- to formulate methods for selecting among them those that have to be kept after they are no longer needed by the creator because of their social or cultural value;
- to develop methods and strategies for keeping the materials selected for continuing preservation in authentic form over the long term;
- to develop processes for analyzing and criteria for evaluating advanced technologies for the implementation of the methods listed above in ways that respect cultural diversity and pluralism; and

– to identify and/or develop specifications for policy, metadata, and automated tools necessary for the creation of an electronic infrastructure capable of supporting the creation of accurate and reliable, and the preservation of authentic digital objects.

2.3 Guiding methodological principles

2.3.1. Interdisciplinarity

The project is interdisciplinary in the measure in which its goal and objectives can only be achieved through the contribution of several disciplines and of all categories of stakeholders: individual creators, the information technology sector, the records management, archival and conservation professions, etc. are involved in the formulation and selection of case studies, gathering of empirical evidence, and analysis. Such a mode of research ensures that the project's results will find ready acceptance within the targeted communities. The scholars conducting the research come from the following fields: Archival Science, Chemistry, Computer Engineering, Computer Science, Dance, Diplomats, Film, Geography, History, Information studies, Law, Library Science, Linguistics, Media Studies, Music, Performance Art, Photography, Records Management and Theatre. The countries actively involved in the project are: Canada, United States, Australia, Belgium, China, France, Ireland, Italy, Japan, Netherlands, Portugal, Singapore, Spain and the United Kingdom. The Advisory Board also includes an archivist from South Africa.

2.3.2 Transferability

The ultimate goal of the project is archival in nature, in that it is concerned with the development of a trusted system for making and keeping digital entities⁵ and of a preservation system that ensures the authenticity of the entities under examination over the long term. This implies that the work carried out throughout the project in the various disciplinary areas must be constantly translated in records management and archival terms and linked to records management and archival concepts, which are the foundation upon which the systems intended to protect the digital entities are designed. However,

5. A trusted system comprises the whole of the rules that control the creation, maintenance, and use of the materials of the creator and that provide a circumstantial probability of the accuracy, reliability and authenticity of the digital objects within the system.

upon completion of the research, the archival systems need to be made accessible and comprehensible to records creators, organizations and institutions and disciplinary researchers. In other words, the research outcomes must be translated back into the language and concepts of each discipline that needs to make use of them. In light of the above, all researchers are committed to learning the key archival concepts that are identified by the archival scholars in the team as constituting the core of the InterPARES 2 research, so that each discipline can identify the corresponding entities within its own body of knowledge.

2.3.3. Open inquiry

InterPARES 1 had its epistemological roots in the humanities, specifically in diplomatics and archival science. In contrast, InterPARES 2, while planning as one part of its research to test some of the outcomes of InterPARES 1 in a range of applied settings, espouses no epistemological perspective or intellectual definitions *a priori*. Instead, researchers in each working group identify the perspective(s), research design, and methods that they believe to be most appropriate to their inquiry. The reason for this openness is that InterPARES 2 is conceived to work as a “layered knowledge” environment, in the sense that some of the research work will build upon knowledge developed in the course of InterPARES 1, some will take knowledge of similar issues developed in other areas of endeavour and bring it to bear on creation and preservation of digital materials, some will reconcile knowledge about records and their attributes, elements, characteristics, behaviour and qualities existing in various disciplines and develop it for records management and archival purposes, and some will explore new issues and study entities never examined before and develop entirely new knowledge.

2.3.4. Multi-method design

As stated, each research activity is carried out using the methodology and the tools that the dedicated investigating team considers the most appropriate for it. Examples of the methods used are surveys, case studies, modeling, prototyping, diplomatic and archival analysis, and text analysis.

The research is guided by detailed research questions that specifically address the records creation process in each of the examined areas of endeavour, and the characteristics, structure and interrelationships of the resulting materials; the issues related to the development of a chain of preservation for those materials that begins with creation and includes appraisal, description, and reproduction as authenticating procedures; the meaning of the concepts of accuracy, reliability and authenticity in the various disciplines; the policies, strategies and standards in each area of activity covered by the research; the descriptive schemas necessary to the identification, use and preservation of the records produced by each activity throughout their life-cycle; and the models that more appropriately represent the digital object that is investigated and the processes of its creation, maintenance, use, selection and preservation.

3. Research Progress

The need to concentrate the initial part of the research on gathering an understanding of the process of creation in dynamic, experiential and interactive digital environments has been especially encouraged and supported by the participant stakeholders. The researchers have carried out case studies and general studies. The case studies were identified according to the specific kind of activity that generated the material, and conducted by individual teams assembled in an interdisciplinary way for the purpose of investigating the entire life cycle of the digital objects that were examined. The general studies were developed to address issues relevant to each of the three types of activities producing records, but not specific to any given case.

While waiting for the results of the case studies, the research unit responsible for investigating the key concepts of accuracy, reliability and authenticity in each discipline involved with the research has produced annotated bibliographies and reviews of literature; an analysis of those concepts as discussed in the reviewed literature; and a bibliographic database for managing references in bibliographies and literature reviews. The knowledge so acquired is guiding the analysis of the findings of the case studies and the development of the related preservation strategies.

The Modeling research unit has developed an activity model of the management of the chain of preservation, the related entity model, and a methodology for walking case

studies data through the model; is testing the model by walkthroughs of selected completed case studies; has begun the development of an activity model of preservation from the creator viewpoint; has designed a protocol for representing in models the findings of case studies within the case study reporting framework; and is developing creation activity and entity models of the completed case studies.

The Policy research unit has conducted literature reviews of existing policies, strategies, guidelines, standards and legislation; has located international instruments that have been developed on freedom of expression, moral rights, etc., and examined how they have been or are being implemented in individual countries; has developed a framework for the comparison of policies and has begun comparing those of different sample countries; and has located relevant legislation and ethical codes, analyzing them in relation to the domains research questions.

The Description research unit has conducted a literature review across all focus areas in order to identify authorities addressing the accuracy, reliability, authenticity and preservation of digital materials by means of descriptive and other metadata schema and standards; has developed a database registering and describing salient features of relevant extant descriptive and other metadata schema and standards; has developed guidelines and provided training for researchers using the database; is populating the database; has established a metadata schema registry (developed specification, developed an XML-based DTD); has created the metadata schema registry database and populated it with a pilot set and is beta testing it with existing standards; and is studying the extent to which all metadata schemas and standards identified in the database and through the completed case studies address the research questions related to description.

Several more research activities are conducted both by individual research units and by the team as a whole, the latter primarily aimed to the development of methodological research instruments and to the reconciliation of findings and drawing of conclusions, but the list presented above should be sufficient to provide the flavour of the InterPARES 2 work. As to the findings, the case studies completed to date have at the same time presented very clear and consistent answers, and raised unexpected challenges, which have helped the team to develop some interesting hypotheses.

3.1 *The Hypothesis*

One hypothesis that I wish to present as an example of the challenges we are dealing with results from the case studies carried out in the context of the artistic and e-government activities, results that are remarkably similar. The records of each individual activity that has been examined both comprise and are each composed of a mix of analogue and digital entities interacting among themselves, often with the mediation of human beings, instruments and/or computer technology. This situation presents three types of issues: 1) issues related to the maintenance of each entity, be it larger, smaller or equal to a record, in a way that its accuracy/reliability and authenticity can be ensured; 2) issues related to the maintenance of these entities' ability to interact with each other, with or without human or technological mediation, both within a record and between records in precisely the same way in which they were meant to when generated; and 3) issues related to the identification of the boundaries of the entity record. The first type of issues may appear easy to solve through traditional methods for the analogue components and through migration for the digital ones. However, migration of digital interacting entities existing in different formats often makes their interaction impossible, alters the functionality of the entities, and results in partial, inaccurate, unreliable and inauthentic reproductions.⁶ The difficulty of the second type of issues derives from the fact that interaction is usually not documented in a way that makes it possible to re-enact it in a different environment or when one of more of the digital entities is upgraded. To overcome this problem requires the development of a special kind of notation for arts material, and of metadata for e-government material, that is capable of describing in an objective, detailed and standardised way the interaction between the record's digital components, a record and another, and the record's components or the records themselves with the mediating entity, so that such interaction can be accurately reproduced. The third type of issue is to be solved case by case, but on the basis of a renewed understanding of what a record is, an understanding that must be linked to our answer to the previous two issues.

6. Migration will present us with this sort of problem for a long time, at least until we have developed a sense of what change in a record is acceptable to the point that we can still say that, regardless of it, the record has preserved its identity and integrity. With paper we know, on the basis of centuries of experience, how much damage a record can tolerate to be considered intact, or how different a copy can be from the item it reproduces to be considered authentic, but with the digital medium, we have to define parameters and develop standards.

The case studies completed in the area of the performing arts are paradoxically helping us to find such answer. With music, the work is the performance, while the score is a set of instructions that allows performers in different times and places to re-produce the same work. Each performance is a little different, depending on how detailed the score is, and therefore on how much discretion is left to the performer in interpreting the piece, on the ability of the performer, on how the musical instruments have changed over time, on the acoustics of the place where the performance occurs, etc., but it is close enough to the original work to be easily identified by the audience for what it is. In other words, although the original live performance cannot be carried forward in time, the existence of a score ensures the accuracy and authenticity of the live performances that will follow. With electronic music, it is becoming quite clear that the set of instructions included in the score, when it exists, is not sufficient to reproduce the piece: we need to carry forward the computer codes, the patches, perhaps a synthesiser, and the interaction between the performer(s) and all of the above, an interaction that so far has never been described. Increasingly, both composers and researchers are arriving at the conclusion that the only way of maintaining digital music is to describe each component of it and the interactions among them, that is, to produce a set of instructions for re-creating each part of the piece and the piece as a whole.

Through case studies of visual art, we are beginning to advance the proposition that, in the digital world, every art form is becoming performing art in character, in that it can only be manifested over time by re-creating it on the basis of a set of instructions, rather than by migrating or even emulating its components and hoping that they will be able to behave as their first instantiations. And here is the hypothesis I wish to present: could it be possible that, with regard to the records resulting from e-government activities, we will be dealing with the same kind of scenario?

In the e-government case studies that we are carrying out, a few recurring features are evident. For each service digitally delivered by a government to a citizen in an interactive mode, we have a record spread across several interacting technologies, a record that has no clear boundaries and changes continuously on the basis of the input of the user (either the government officer or the citizen) and of the reaction of the system, and a record that rarely corresponds to an action and more often includes the whole

interaction between a government office and a citizen with respect to one matter (what used to be a paper file, in other words). How can we maintain this kind of record accurate and authentic for purposes of accountability? Some have tried to take views of the system at regular intervals and save them as records in fixed form in a recordkeeping system, but this does not really work when changes to the record-making system do not occur at regular intervals. Others have tried to keep a log of all changes and freeze it as the complete official record, but such record is permanently in the making and very cumbersome to use for the verification of the accuracy and authenticity of the records it relates to. Yet others have tried to keep the system with all the records it contains permanently live. This choice, regardless of the fact that it does not ever allow to close any matter, does not protect the records from tampering, and the necessary periodical upgrades put at risk the accuracy of the data and the authenticity of the record, whatever might the record be. And this is indeed the crux of the matter. In e-government interactive systems, where does the record commence and end, what is the record that we need to save in a trusted recordkeeping system?

Whenever I am confronted with a dilemma such as this I try to look for past situations that can be related to the one I am observing. Certainly we have never had in the past interactive systems such as those used in e-government activities, but in Medieval times we have kept records “attributes” in such a way that the actual complete and effective records could be created at will. I am not talking about record metadata, which we have also had for centuries in form of “*regeſtum*”, because these existed in addition to the record or as its surrogate once the record had been destroyed, in either case for the purpose of proving the existence of the record, not of producing it when it was needed some time in the future. I am rather referring to the “*imbreſiaturae*” of thirteenth and fourteenth century notaries. When the notaries became ſo powerful as a profeſſion that every tranſaction had to be recorded and preſerved by them, they decided not to go any longer through the trouble of writing out the records of the tranſactions that they witnessed. They would take a parchment, fold one corner forward, and write on it the tranſaction type, the names of the parties, the date, the deſcription of the tranſacted property or matter, and any other data ſpecific to that tranſaction. Then, they would file the blank parchment away and, at the end of each year, bind all the *imbreſiaturae* of the

year. If later on one or more of the parties to that transaction or their descendants wanted the complete record of the transaction, the notary would find the *imbreviatura* in question by date, take a new piece of parchment (or paper, if appropriate), and write out the record with the data written on the *imbreviatura*'s corner and the formulas contained in a special book, called *formularium*, which provided clear instructions for writing out a record for any possible type of transaction. Thus, what the notaries really maintained were not records of transactions, but the ability to re-produce such records upon request. In other words, they kept a record of the fact that a transaction had occurred (the data of the transaction) and sets of instructions guiding them to make the accurate and authentic record of the transaction when needed, even centuries later, as each notary left its archives to its legitimate successor. However, precisely because of this trust, almost never such a record was requested: the existence of the *imbreviatura* was sufficient.

Reflecting on the *imbreviaturae*, I came to think that perhaps the best way of keeping the interactive records of e-government activities across technologies is to separate the data of the record from its form and technological environment, fix the former and link them to the latter by generating a description of the original form and functionality. As with the *imbreviaturae*, most times this set of records of the existence of a transaction would be adequate to serve both administrative and historical accountability, as well as legal purposes. In the few cases in which the actual record of the transaction would have to be re-created, it would likely be sufficient to provide the data with the appropriate form, and accompany this record with the description of the functionality of the original environment. Of course, this hypothesis is based on the assumption that, upon completion of the interaction between the parties, the finished entity will be the exclusive responsibility of a trusted custodian like a notary, that is a person who has no stake in the content of the record, and can therefore fulfil the role of a third neutral party, and who is formally recognized competent to maintain the record because of his/her professional formation—a records manager, that is. This hypothesis has yet to be fully developed and tested, but I have to admit that it has a very special appeal to treat the records of e-government like the records of a performing art...

4. Conclusion

Meeting the challenge of accuracy and authenticity in the digital world is much more than establishing policies and procedures, developing metadata schema, or designing new types of digital signature. There is no silver bullet that can get rid of the problems presented by digital technologies. The ability to deal with them stands at the heart of our profession as records specialists. We need to re-examine the concepts that constitute the fundamental tenets of our profession and ensure that we find solutions that are consistent with the nature of the records, with their function and purpose, and with our mission as the trusted custodians of their accuracy and authenticity. In my view, the only way to do this is to remember that there is very little which is new under the sun and, most times, this is only a new manifestation of something that existed before, that our civilizations have formidable legal and record cultures developed over millennia, and that, if the solution cannot be initially found within our discipline or environment, it is always useful to look at other disciplines and environments and wonder..."what if"?