The Concept of Record in the Digital World: The View of InterPARES

Luciana Duranti

The InterPARES Project is an international research endeavor that aims at developing the theoretical and methodological knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form. This knowledge should provide the basis from which to formulate model policies, strategies and standards capable of ensuring the longevity of such material and the ability of its users to trust its authenticity. InterPARES has developed in two phases. InterPARES 1 dealt with records mandated for accountability and administrative needs that are created in databases and document management systems. InterPARES 2 has focused on the portion of society's recorded memory that is digitally produced in interactive, dynamic and experiential systems in the course, and as a byproduct, of artistic, scientific and electronic government activities.

The InterPARES research team determined at the very beginning of the first phase of the project that, in order to be able to identify among the various types of digital information contained in existing systems what corresponded to a record, it was necessary to agree on the definition and concept of record and on how such entity differed from document, information and data. Thus, the team called a **record** any document created (i.e., made or received and set aside for action or reference) by a physical or juridical person in the course of a practical activity as an instrument and by-product of such activity, thereby adopting the traditional archival definition. The team then proceeded to define **document** as recorded information, **information** as a message

intended for communication across space or time, and **data** as the smallest meaningful piece of information. Finally, an **electronic record** was defined as a record created in electronic form, meaning that a document received in electronic form but set aside for action in paper form is a paper record, while a document received on paper but scanned in the computer and only used as a digital file is an electronic record. This definition is fully consistent with the archival principle that whatever the creator treats as a record in the course of any given action is indeed a record of the action in question.

After having chosen a definition of electronic record based on traditional archival theory, the team considered essential to delve into the concept of record underlying and implied by such definition by determining what the necessary characteristics of an electronic record are on the basis of both archival theory and diplomatic theory. The support of diplomatic theory was considered necessary because diplomatics is a discipline that examines records as items rather then as part of aggregations and is therefore able to complement the knowledge provided by archival theory for the purpose of identifying the characteristics embedded in the record itself rather than in its relationships. The following necessary characteristics were identified: 1) a fixed form, meaning that the entity's binary content must be stored so that it remains complete[1] and unaltered, and its message can be rendered with the same documentary form it had when first set aside; 2) an unchangeable content;[2] 3) explicit linkages to other records within or

1 Completeness here is not mentioned as a characteristic of the record, because an incomplete record is still a record, albeit a bad one, but as a characteristic of a fixed form, according to which a form that is fixed is one that does not lose any of its original elements in the process of being stored and retrieved.
2 The stability of the record, as determined by its fixed form and its unchangeable content, is only implied in the part of the archival definition that reads that a record is a document (i.e. rather than just data or information), but it is explicitly stated in the diplomatic definition and concept of record (see Luciana Duranti, *Diplomatics. New Uses for an Old Science* (Lanham, Maryland, and London: The Scarecrow Press, Inc., The Society of American Archivists and the Association of Canadian Archivists, 1998), pp. 41-58.

outside the digital system, through a classification code or other unique identifier;[3] 4) an identifiable administrative context; 5) an author, an addressee, and a writer; and 6) an action, in which the record participates or which the record supports either procedurally or as part of the decision making process.[4]

Having specified the necessary characteristics of an electronic record, the research team decided to accept as a working hypothesis the fundamental assumption of diplomatics that, regardless of differences in nature, provenance or date, from a formal point of view, all records are similar enough to make it possible to conceive of one typical, ideal documentary form containing all possible elements of a record. On the basis of this assumption, the team itself hypothesized that, while they may manifest themselves in different ways, the same formal elements that are present in traditional records exist either explicitly or implicitly in electronic records, and that all electronic records share the same formal elements. Thus, it created a template, that is, a decomposition of the ideal electronic record, first, into its constituent parts, and then, within the part "form," into its elements.[5] In the template, the parts and elements are defined and their purpose is explained. The research team used the template as an instrument for the systematic

---

3 This characteristic corresponds to the archival bond, which is implied in the archival definition when records creation is linked to an activity, but it is made explicit by archival theorists of all times and cultures. See Luciana Duranti, "The Archival Bond," Archives and Museum Informatics 11, nos.3-4 (1997): 213-218.

4 While characteristics 4 and 6 can be deduced from the archival definition, characteristic 5 derives from the diplomatic concept of record: it was considered important in order to distinguish records from digital entities resulting from simply querying a database. The author is the person issuing the record, the writer is the person determining the articulation of the discourse in the record, and the addressee is the person for whom the record is intended. As a record must participate in an action and any action must fall on somebody, the addressee is necessary to the existence of the record. See the Appendix 2 of the book *The Long-term Preservation of Electronic Records: the InterPARES Project* on the InterPARES web site www.interpares.org [http://www.interpares.org/book/index.cfm].

5 The reason why the constituent parts of the record ended up in the template that is supposed to represent the ideal form of a record is that all identified constituent parts used to be regarded as necessary extrinsic elements of form by traditional diplomatists. It was important to show their presence, definition and purpose, and the fact that they are now independent of form.

analysis of the electronic entities contained in several different systems, for the purpose of establishing which ones are records.

The template is composed of four sections corresponding to the four necessary constituent parts of every record: documentary form, annotations, context, and medium.[6] The documentary form[7] includes, among the intrinsic elements,[8] the names of the persons concurring to the creation of the record, the chronological date, the place of origin of the record, the indication and description of the action or matter, the attestation, and a statement of validation; and, among the extrinsic elements,[9] overall presentation features (e.g. text, image, sound, graphic), specific presentation features (e.g. layouts, hyperlinks, colors, sample rate of sound files, resolution of image files, scales of maps), electronic signatures and seals (e.g. digital signature), digital time stamps, and special signs (e.g. digital watermarks, organization crest, personal logo).[10]

The annotations[11] fall into three fundamental groups: 1) additions made to the record after its creation as part of its transmission (e.g. priority of transmission, date of compilation and date of transmission in an e-mail record, the indication of attachments),

---

6 In a previous research endeavour commonly known as the UBC-DoD project, the parts constituting the records were identified as: medium, form, action, persons, archival bond, content and context. See Luciana Duranti and Heather MacNeil, "The Preservation of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project," *Archivaria* 42 (Spring 1997): 46-67; and Luciana Duranti, Terry Eastwood and Heather MacNeil, *Preservation of the Integrity of Electronic Records* (Dordrecht, Kluwer Academic Publishing, 2002: Chapter 1. In the context of InterPARES, it was decided that action, persons, archival bond and content, contrary to the other constituent parts, continue to manifest themselves in formal elements and are inextricable from them, so they do not have to be identified separately from the form. As it regards the annotations, which were not among the parts identified in the MAS-DoD project, they were added to the constituent parts because they are often linked to the record rather than embedded in it, and need therefore to be looked at separately from the record form.

7 Defined as "The rules of representation according to which the content of a record, its administrative and documentary context, and its authority are communicated.

8 Defined as "The elements of a record that convey the action in which the record participates and its immediate context.

9 Defined as "The elements of a record that constitute its external appearance.".

10 See Authenticity Task Force, "Template for Analysis," in *The Long-term Preservation of Electronic Records: the InterPARES Project*, cit. <http://www.interpares.org/book/index.cfm>.

11 Defined as "Additions made to a record after it has been created."

2) additions made to the record in the course of handling the business matter in which the record participates (e.g. date and time of receipt, action taken, name of handling office), and 3) additions made to the record in the course of managing it as a record (e.g. filing date, class code, registration number). The categorization of the contexts of the record[12] and the list of what would reveal them correspond to an hierarchy of frameworks that goes from the general to the specific: 1) juridical-administrative context (manifested in, for example, laws and regulations), 2) provenancial context (manifested in, for example, organizational charts, annual reports, tables of users in a database), 3) procedural context (manifested in, for example, workflow rules, codes of administrative procedure), 4) documentary context (manifested in, for example, classification schemes, records inventories, indexes, registers), and 5) technological context (manifested in, for example, hardware, software, system models, system administration).[13]

The medium[14] was difficult to place within the template, because, although it is still necessary for an electronic record to exist, it is not inextricably linked to the message, does not store the record as such, but a bit-stream, and its choice by the record-maker or keeper can be either arbitrary or based on reasons related to preservation rather than to the function of the record. In addition, the medium is not a relevant factor in assessing a record's authenticity—one of the primary purposes of InterPARES—at least from the perspectives of the creator and of the record preserver.[15] This was confirmed by the case studies undertaken by the research team, by the end of which the team was

[12] Defined as "The framework of action in which the record participates."
13 For details related to annotations and contexts, see the Template for Analysis referenced above.
14 Defined as "The physical carrier of the message."
15 An additional reason for the InterPARES team to dissect the concept of record was to identify what parts or elements contribute to the authenticity of the record and to the ability to verify it.

convinced that, with electronic records, the medium should not be considered a constituent part of the record but a part of the record technological context.

The analyses of the case studies conducted using the template indicated that only a half of the examined systems contained records (twelve out of twenty-two), primarily because the entities identified within the other half did not appear to possess either a fixed documentary form or a stable content. When systems did contain records, these could rarely be compared with the model represented by the template, because, although they were able to achieve their purposes, they were not good records. For example, in most systems, there was no explicit manifestation of the relationship among the records participating in the same affair or matter, and, although it was easy to identify the business processes supported by the system, it was not always possible to determine how the records participated in or supported specific actions. In addition, it was often difficult to determine the significance of the presence or absence of given elements of documentary form or of annotations.

More importantly, the case studies showed that, with digital records, a key concept to consider is that of *records attributes*, which are the defining characteristics of each given record or of a record element in it. A *record element* is a constituent part of the record's documentary form and, as seen earlier, may be either extrinsic, like a seal, or intrinsic, like the salutation.[16] An attribute may manifest itself as one or more record elements. For example, the name of the author of a record is an attribute, which may be expressed as a letterhead or a signature, both of which are intrinsic elements of documentary form, that is, record elements. In addition to attributes that manifest themselves in the form of the record, that is, on the face of the record, as record elements,

[16] A defining characteristic, or attribute, of the record element "seal" may be its legend.

every record has attributes that are implicit in other parts of the record, such as the name of the creator or of the medium, but in digital records they are also expressed, albeit outside the documentary form. Because of this, they are mostly transparent to the user, and manifest themselves as metadata included in either a record profile,[17] a topic map, or other digital entity linked to the record. Attributes made explicit outside the record as metadata demonstrating its identity are important to uniquely identify any digital record, but they become the primary means for the identification of digital objects that do not have—or at least for as long as they lack—a stable content or a fixed form. This will become clearer later on.

As if the distinction between record elements and attributes were not sufficiently complex, with electronic records we also have to differentiate elements and attributes from the record's digital components. A *digital component* is a digital object that may contain all or part of a record, and/or the related metadata, or more than one record, and that requires specific methods for preservation.[18] For example, an e-mail containing a textual message, a picture and a digital signature has at least four digital components that are stored in different part of the system, although they are linked among themselves, and require different protection measures: the header, the text of the message, the picture and the digital signature. In contrast, a report with textual attachments may consist of only one digital component. In other words, a digital component is a unit of storage, but one that needs to be identified when the concept of digital record is dissected.

[17] A record profile is an annotation inextricably linked to the record that includes several fields, which are either automatically or manually filled in with the record's metadata.
[18] See Preservation Task Force Report in *The Long-term Preservation of Electronic Records: the InterPARES Project*, cit. <.http://www.interpares.org/book/index.cfm>.

Finally, the InterPARES 1 team felt the need to point out that the relation between a digital record and a computer file can be one-to-one, one-to-many, many-to-one, or many to many, thus we should never use the terms record and file interchangeably; that the same presentation of a record can be created by a variety of digital presentations and, vice-versa, from one digital presentation a variety of record presentations can derive, thus fixed form does not imply that the bit streams must remain intact overtime; and that it is possible to change the way in which a record is contained in a computer file without changing the record, thus the name of a record form does not necessarily indicate what digital object we are dealing with.[19]

The concept of digital record presented above, with all its characteristics, parts, formal elements, attributes and digital components, has worked quite well with databases and document management systems. However, it may appear problematic when applied to the entities examined by InterPARES 2, the most salient characteristic of which seems to be, as mentioned earlier, the lack of a stable content and/or fixed form, not because they are bad records, as it was often the case with the digital entities examined during InterPARES 1, but because fluidity is part of their nature and contributes to the accomplishment of their purpose as instruments of or support for action. They are experiential, interactive and dynamic records.

Experiential records are electronic objects the essence of which goes beyond the bits that constitute the object to incorporate the behavior of the rendering system, or at least the interaction between the object and the rendering system. Defining the characteristics, parts, elements, attributes and components of such objects is much more complex than with traditional electronic records, because it is dependent not only on the

[19] Ibidem.

object per se, but on the environment in which the object is experienced. Examples of experiential digital objects range from audio and moving images embedded in a web page to virtual reality systems.

Interactive records are records made and maintained in interactive systems, where each user's entry causes a response from or an action by the system. To determine the boundaries of such records (i.e., where one record ends and another begins), when they can be considered finished rather "in progress", when they are complete rather than partial or incomplete, etc., one needs to ascertain 1) how user input affects the creation and form of each record (as is the case with much on-line commerce); and 2) if and when the interactive system and its inherent functionality are to be regarded as meaningful parts of the record. Examples of interactive systems range from web pages delivering government services online to musical performances based on human-computer interaction to commercial video games.

Dynamic records are documents whose content is dependent upon data that vary continuously and are held in several databases and spreadsheets. Examples range from simple web pages with embedded links to complex systems where information is stored and updated in order to be shared via wireless transmission by multiple mobile users in diverse ways. The increasing reliance on such documents by individuals and institutions will necessitate understanding how the information they contain is captured and saved.

Whether experiential, interactive, and dynamic digital objects are records primarily depends on their relationship to the activity of their creator. It is out of question that these objects must be subjected to same kind of scrutiny which any document undergoes in relation to the action and procedure in which they participate, the archival bond they have

with other records of the same creator, etc., in order to establish whether they are records or not. However, even if they pass such scrutiny with respect to these fundamental requirements, the question of form looms large. Is it possible to have a record in fluid form and with undetermined boundaries? If not, should an entity with fixed form be generated for the purpose of making a record to be kept in a trusted recordkeeping system and perhaps preserved over the long term? If yes, who should make it? On the basis of which criteria? When in the entity's lifecycle?

The InterPARES 1 Authenticity Task Force's report suggests the possibility of trading the record characteristics of stability of content and fixity of form (including completeness of content and form with respect to the first and to any subsequent instantiations of the record) with the ability of the system containing it to track and preserve any change to the record. In other words, the Task Force was inclined to shift the requirements of stability and fixity from the record to the log of the changes to the record once the record was no longer active;[20] in this context, the entity identified as the record and to be kept intact would then be the last instantiation of the fluid entity, plus the complete log of changes, and the metadata of both. This option is conceptually sound only if the creator uses this set of entities as its record, but this scenario is very unlikely because it would be highly impractical. Alternatively, one might look at each digital object participating in the creator's activity as an instrument and by-product of it as existing at any given time in one of two modes, as a record *in fieri*, that is, in becoming, when its process of creation is ongoing, that is, when the entity is accessed to add data or information, and as a record when the entity is accessed for use. This would imply the stabilization and maintenance of every instantiation accessed for use and its metadata.

[20] Here, active is used to mean "subject to changes or additions."

Conceptually, this option is as sound as the previous one, but it appears to be equally impractical.

One other option can be developed from the findings of 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 issues of three types: 1) issues related to the maintenance of each digital object, 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 the relationships among analogue and digital entities, and of the ability of the various digital objects 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 interact when generated; and, most important in the context of this paper, 3) issues related to the identification of the boundaries of the entity record.

Although it is theoretically improper to base the identification of an entity on the requirements for its preservation beyond its active state, it is methodologically justifiable in the context of InterPARES, which 1) aims at finding solutions to the long-term preservation of digital records (hence, the entity we identify as a record has to be preservable), 2) has determined that the only way of preserving digital records is to produce authentic copies of them (hence, the entity we identify as a record must have a definitive complete instantiation—a state of stability beyond which no change will occur-

- which is the entity of which authentic copies would be made), and 3) has stated that the chain of preservation begins at creation (hence, the entity we identify as a record at creation should be the one that we can preserve). In this context, it is important to remember that InterPARES 1 has clearly formulated the difference between the copies of record made by the creator in the course of business and for the purposes of its business (be they the reproductions generated every time a record is retrieved or the copies made as a result of system upgrade or record migration), which it has called "the records of the creator," and the copies of records made by the preserver in the course and for the purposes of archival functions, which it has called "the authentic copies of the records of the creator." This distinction is vital for identifying the entity record, because it means that, if the creator alters the form of the record in order to be able to keep it for future action or reference, the result of such alteration is a record of the creator, and the preserver will have to carry forward an authentic copy of the last instantiation of that record as produced by the creator, including the metadata attesting to the changes. Finally, to arrive at the articulation of a concept of record on the basis of conceptual requirements for preservation is justifiable because we would not be defining records, something that the archival and diplomatic sciences have already done quite satisfactorily, as aggregates the one, and as items the other:[21] What we would be doing is

[21] The archival and diplomatic traditional definitions have served us quite well through the centuries because they have enough rigour to show a clear demarcation between an entity that is a record and one that is not a record, and enough flexibility to be applicable to entities produced in all the technological, administrative and cultural environments that have existed to date. All attempts to refresh those definitions by making them more specific have in my opinion miserably failed. For example, the definition coined by the *Guide for managing electronic records from an archival perspective*, issued by the International Council on Archives Committee on Electronic Records (ICA Studies, February 1997), which, at p. 22, reads "*A record is recorded information produced or received in the initiation, conduct or completion of an institutional or individual activity and that comprises content, context and structure sufficient to provide evidence of the activity,*" is at the same

to develop a description of what entity the archival and diplomatic definitions concretize themselves into in the context of a dynamic, interactive or experiential digital environment. With this understanding, we can proceed to discuss the three types of issues identified earlier.

The first type of issues, related to the maintenance of each digital object, may appear easy to solve by using migration. 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.[22] The difficulty of the second type of issues, related to the maintenance of the relationships among analogue and digital entities, derives from the fact that those relationships, as well as the interaction among the digital entities in question are usually not documented in a way that makes it possible to re-enact them 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 schemata for e-government material that are capable of describing in an objective, detailed and standardised way the interaction between the record's digital and analog components, a digital 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, related to the identification of the boundaries of the entity record, is to be solved case by case, but on the basis of a renewed understanding of

---

time ambiguous and inflexible, and would certainly exclude from the record category all dynamic, experiential and interactive records.

[22] Migration as a method of maintenance and preservation 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.

what is implied in the definition of record, an understanding that must be linked to the answer to the previous two issues.

The case studies completed in the area of the performing arts are helping to find such answer. With music, the work is considered to be the performance, while the score is regarded a set of instructions that allows performers in different times and places to re-produce the same work.[23] 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 related 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 performance cannot exist as a live work of art longer than its manifestation, the presence of a score ensures the accuracy and authenticity of the live performances that will follow. However, 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: one needs to have also the computer codes, the patches, a synthesiser when used, and the interaction between the performer(s) and all of the above, an interaction that so far has never been described as a reproducible set of instructions. Increasingly, both composers and InterPARES researchers are arriving at the conclusion that the only way of keeping digital music as a record 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.

[23] There is some disagreement on whether the score is also a work in its own right, but this does not invalidate the option I intend to present. If anything, it supports this option.

Furthermore, through case studies of visual art, InterPARES is 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 record made up 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. This proposition is supported by several initiatives, separate from InterPARES, which have tried to solve the preservation problems presented by digital objects that are characterized by their capacity of evolution, their interactivity, their dependence for form and content from external factors, and the centrality of movement and multimediality to their meaning. The fundamental issue these initiatives are confronted with is whether to preserve these objects or maintain them live. To preserve them means to choose between fixing them in a definitive form with one final act of interpretation and representation, and treating them like a musical score, by creating a description of them, a document that opens to the possibility of generating future reiterations of the work. In the former case, it would be necessary to identify as the work a collection of, for example, installations, machines, software, Internet screens, videos of interactions among the parts, etc. This would imply loss of causality, dynamicity and artistic experience. In the latter case, the act of interpretation would be left to the performer or the user in general, accepting the fact that each act of interpretation will have a different result, although always identifiable as the same work. A project that tries to find a compromise between the two preservation alternatives is the Rhizome ArtBase. This project has developed two concepts, that of "connected art object" and that of "cloned art object." The former comprises the description of the work and its components, a thumbnail of the work,

keywords and metadata, a link to the URL of the work, the biography of the artist, and his/her certification that this aggregate of parts corresponds to the work and constitutes an adequate representation of it. The latter includes, in addition, an authentic copy of the work preserved in the server of the project.[24]

The project Rhizome ArtBase brings to light the most important issue presented by the identification of the work to be preserved, that of artistic intention. Lacking the possibility of carrying forward into the future an intact work of electronic art, one is left with the option of capturing the essence of the work, but this implies that the author and/or creator become active participant in the preservation act and state in explicit and authentically preservable way that the entity being preserved is the substance of the work in question. The authenticity of the work is ensured by the personal involvement of its author/creator in the decision of how it will be re-created in the future. Its nature of record is ensured by the fact that the author generates this surrogate, or potential work, as a regular part of his/her creative activity and for its purposes: it is a very fine line to walk, but, as long as the creator does not produce this entity for the preserver, but for his/her own benefit, it is one consistent with the concept of record. Certainly, examining the list of the entities comprising the connected and the cloned object, it appears evident that, from their sum it is not possible to generate a work identical to the original one, but its essence would be captured. In InterPARES terms, to equate the concept of work of art with the objects described above means to give preference to continuing accessibility over accuracy (obviously, reliability is maintained by the control of the author/creator on

[24] Alena Williams, "Rhizome.org," in Alain Depocas, Jon Ippolito, and Caitlin Jones eds., Permanence *Through Change. The Variable Media Approach*, New York and Montreal: Guggenheim Museum Pub. And The Daniel Langlois Foundation for Art, Science and Technology, 2003, p. 39.

the process) without compromising the spirit of authenticity, given that the identity of the work is ensured and its integrity[25] can be (potentially at least) reconstituted.

And here is a hypothesis that I wish to propose: With regard to the records resulting from e-government (and possibly e-science) activities, we might be dealing with the same kind of scenario presented by the digital works of art. With e-government records, a few recurring features are likely to occur. For each service digitally delivered by a government to a citizen in an interactive mode, there will be 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/or of the reaction of the system to such input, and a record that rarely corresponds to one action and more often includes the whole interaction between a government office and a citizen with respect to one matter (i.e., what used to be a paper file, a dossier that is). Thus, one will need to identify 1) the boundaries of the digital entity that constitutes the record which, once made or received, and repeatedly set aside in different instantiations, is to be kept for future action or reference in a trusted recordkeeping system, 2) the essence of such entity, i.e., keeping in mind the acceptable degree of change from the moment the record has reached its final instantiation, its constituent parts and digital components to be kept stable as content and fixed as form and linked among themselves, 3) its attributes to be manifested in metadata permanently attached to the record as an annotation (thereby becoming a constitutive part of the record), and 4) the necessary accompanying documentation of what is not fully preservable, that is, interactivity, connectivity, and functionality. Once this identification

[25] In the Authenticity Task Force Report, cit., integrity is taken to mean that the substance of the message is conveyed in the same intellectual form as its first instantiation.

has occurred, one could assemble the stabilized essence of the record, its metadata and system documentation and treat this entity as the record. Would such procedure be acceptable if it were the interest and responsibility of the creator to carry it out? Is such an idea contrary to the theoretical understanding of what is a record?

Although, as stated in the InterPARES 1 Authenticity Task Force report, the study of new record types seems to indicate that not always what is known can guide to the understanding of what is unknown, I believe that we should keep looking for past situations that can be related to each present situation that one is observing. Certainly, there have never been in the past interactive records such as those resulting from e-government activities, but in Medieval times offices have kept records "attributes" in such a way that, when a finished record did not exist, complete and effective records could be created at will years later. I am not thinking of record metadata, which have also been generated for centuries in form of "*regestum*", because they existed either 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 "*imbreviaturae*" of medieval notaries. When the notaries became so powerful as a profession that most transactions had to be recorded and preserved by them, they did no longer go 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 away the blank parchment with the

annotated corner[26] and, at the end of each year, bind all the *imbreviaturae* of the year in the same volume, and index the volume and or keep a separate registration of the occurred transaction in a book of *regesta*. 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 volume containing the *imbreviatura* in question by date, retrieve the document in it through the index or the register, take a new piece of parchment (or paper, if appropriate), and write out the record with the data written on the *imbreviatura* corner and the formulas contained in a special book, called *formularium*, which provided clear instructions for writing out a record for each type of transaction that occurred in a specific range of years. Thus, what the notaries really maintained was not the complete record of each transaction, but a record of the content of the transaction and of the documentary form in which it had to be manifested, and the ability to produce a complete record upon request by integrating content and form. In other words, they kept a record of the fact that a transaction had occurred (register and/or index), the data of the transaction (*imbreviatura*), and sets of instructions guiding them to make the accurate and authentic record of the transaction when needed (*formularium)*, even centuries later, as each notary left its archives to its legitimate successor. However, precisely because of this trust, almost never a party or its successors requested that a complete record be issued: the existence of the *imbreviatura* in a notary archives was sufficient evidence of the transaction.

From the observation of the *imbreviaturae*, one can imagine a similar way of keeping the interactive records of e-government activities across technologies: at the completion of each transaction, the handling office/officer would, as a matter of routine,

[26] Sometimes, rather than on a corner, they would write the data on the back of the medium.

separate the data of the record from its form and technological environment, stabilize the former and the metadata of the original record, and link them to the latter by means of a description of the original form and functionality. As with the *imbreviaturae*, most times, this set of documents of a transaction, properly registered, would constitute a record adequate to serve both administrative and historical accountability, as well as legal purposes. In the few cases in which a complete and finished record of the transaction would have to be re-produced, it would likely be sufficient to embed the data in the appropriate record form, and accompany this record with the description of the functionality of the original digital environment. The fundamental difference between e-government records and the *imbreviaturae* is that interactive records come into existence as complete and effective records at the end of the interaction between the government and the citizens, however abstract the concept of complete can be with a record that is live with its original functionality, while the *imbreviaturae* only exist as initial rough drafts of a potential record. Thus, while the record generated from an *imbreviatura* would be created as a first instantiation of a record, its original, that is, the record generated by re-producing the e-government record after it had been taken apart in order to set its components aside as fixed entities, would be created as an authentic copy of the original interactive record. However, because the creator would use it in the usual and ordinary course of business, for further action or reference, this authentic copy would be considered a record of the creator, as discussed earlier in this paper.

This hypothesis is, however, workable only on the assumption that, upon completion of the interaction between the parties, the finished entity will be the exclusive responsibility of a trusted custodian similar to 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 a person who is formally recognized competent to maintain the record because of his/her professional qualifications—a records officer, that is. This person would be the one making the surrogate of the interactive record, keeping it accessible to the competent offices in a trusted recordkeeping system, and generating the complete re-production upon request.

Thus, in the scenario depicted above, and keeping in mind the accepted definition of record, to what entity does a record correspond? I would suggest that, while the business procedure is active, the interactive digital entity constitutes the overall record of the transaction. If instantiations accessed for use by the parties at each given time are set aside with a fixed form and a stable content and linked to other records of the same transaction, they are also records of the transaction. Once the business procedure is concluded, the final record of the transaction will consist of the data contained in the last instantiation of the interactive entity[27] and its metadata, properly linked to an exemplary of its form and a description of its digital environment (i.e. record functionality and system documentation) that would already be maintained in the recordkeeping system to which such record will be transferred. The key to the reliability, accuracy and authenticity of such record will be the fact that the responsibility for generating and maintaining it as the source of future re-productions of the original interactive entity in its active state will reside with the creator (and, within the creator, with the record office):

[27] Assuming that no data would be deleted in the course of the transaction, as good record making would recommend. If data were deleted as a matter of course during the transaction, a log of changes would have to accompany the data contained in the last instantiation. If it were part of the formal procedure to delete data in the course of the transaction, the record of the transaction would be complete without the log, but the office would have to keep a description of the procedure linked to the series of records subject to it in order to account for the deletions.

the record-source will be the record of the creator just like the interactive entity was. If a re-production of the original interactive entity from the record-source will be made by the creator for its purposes, such re-production will still be the record of the creator, while, if it will be issued to an external user for other purposes either by the creator or by the preserver (if the record-source were transferred to an archives), such reproduction will be an authentic imitative copy of the record of the creator.

Of course, this option only addresses situations in which e-government creates dossier-like digital interactive records. But InterPARES 2 case studies have dealt with more complex situations. One example will suffice to show similarities and differences: the VanMap case study. VanMap is a GIS system that allows the City of Vancouver to meet the needs of internal users in providing services to Vancouver's citizens and businesses. VanMap supports the functions and activities of the following departments: Community Services Group, Engineering Services, Corporate Services Group, Board of Parks and Recreation, Vancouver Police Department, Fire and Rescue Services. The decisions on the layers and the set of data they should contain are made collectively by the departments and the VanMap Technical Team. Data are uploaded by each department directly in Oracle Spatial or taken as extracts from external offices databases (for example, permit and license data stored in PRISM or License+ are extracted to a SQL server; property tax data are extracted from the SQL Property Tax System, etc.) for inclusion in VanMap by the Technical Team, which is responsible for its administration. Engineering and CSG graphics are created in the form of CAD drawings in AutoDesk, or keyed or drawn in the Oracle Spatial database. VanMap data are overwritten at each

update and, every once in a while, existing layers are modified to receive different kinds of data sets and new layers are added.

The VanMap, as an indivisible whole and in the context of each of the numerous business processes in which it participates, perfectly corresponds to the archival concept of record in that it is treated by the creator as a record, it is linked to the other records participating in each business process by a documentary and procedural relationship, and it is the byproduct and residue of the transaction of affairs. As such, besides exhibiting all parts and characteristics identified by InterPARES 1, except of course stable content and fixed form,[28] VanMap manifests the traditional archival characteristics of naturalness, impartiality, interrelationship, authenticity, and uniqueness in context.[29] However, this conceptual recognition is logical, clear and of no consequence while one focuses on the VanMap as a live system, but it becomes very much clouded when one focuses on each business process and tries to identify the records participating in it. Would the whole VanMap be a record in each of the very numerous business processes carried out by the city departments that use it? Unless the instantiation of VanMap and, within it, the specific data layers and data sets participating in each given business process have been extracted from the system, frozen and, as a matter of course, attached in such form to the

[28]Following the VanMap diplomatic analysis, it might be argued that, although its digital components may undergo dramatic changes every so many years, the documentary form of VanMap is quite stable, because its intrinsic and extrinsic elements do not change other than in their content.

[29] A discussion of these characteristics can be found in Terry Eastwood, "What is Archival Theory and Why is it Important?" *Archivaria* 37 (Spring 1994): 122-130. As it regards uniqueness, it should be noted that, in the case of VanMap, it is also present with regard to content, as the information provided by VanMap and resulting from the combination of data sets originating in different departments and offices, does not exist anywhere else, although many of its data sets exist either as such, as part of records, or as isolated data in individual offices.

A reduced copy of VanMap accessible to the public at large is kept outside the city firewall to ensure that users do not either accidentally or maliciously see layers that are confidential within the VanMap for city staff. This digital entity is to be regarded as a publication and does not have any of the archival characteristics of the VanMap used by the city employees. Of course, if a user were to download the public VanMap to its system and use it in the course and for the purposes of its business, that specific downloaded entity could be a record in the user's fonds.

related records—a routine that nobody uses—the observed entity, which existed for the time necessary to carry out a transaction with a citizen, or to inform somebody of a given situation, or to make decisions or plans, disappears. Each instantiation was "made" at each update of the data, and "received" at each use, but never "set aside," thus, it was only a potential record that never materialized. As a consequence, while we can still regard the whole VanMap as a record of the City of Vancouver collectively generated by its staff, it is not a record in the context of any given business transaction.

This situation should be rather disturbing to the City Council though, because its ability to account to the citizens for the actions of the city staff that affect them, both individually and collectively, is greatly diminished by the inability to demonstrate the factual grounds of city decisions.[30] Thus, as archivists responsible for advising creators about proper recordkeeping practices, we could imagine a solution capable of supporting both current and historical accountability: we would recommend the VanMap Technical Team, which includes representatives of the city departments, to develop a detailed description of each business process in which VanMap is involved and of the way in which VanMap is used in each of them, thereby revealing the archival bond between the records of each business process and VanMap. It is a centuries-long tradition to embed in a code of administrative procedure the function of a record that serves multiple activities and procedures, but of which only one original exist (see for example the series of the maps of the cadastre, which were and are used as records in several procedures having different purposes). Thus, this solution, per se, is not problematic as a principle or as a practice. What is problematic is that VanMap does not contain historic data, as it is

[30] This is where the uniqueness of the VanMap content comes into play because it is the coexistence and interaction of data of different, both internal and external, origin that makes VanMap a vital source for making decisions, defending and proving their factual basis.

constantly updated by overwriting superseded information. This means that, if a citizen who has been denied a building permit, for example, knowing the procedure and the part that VanMap has in it, requests access to the information contained in VanMap when his application was submitted, his request cannot be satisfied. Thus, this proposed solution would have to be accompanied by some additional procedure. One could recommend that each staff member using VanMap in the course of a transaction freeze the view related to each decision and save it as an attachment to the records of the business process in question. But this procedure is laborious without supporting the work of the staff members in any other way, therefore is not likely to be followed. Another, more effective procedure that would indeed make the work of some, if not all, departments more accurate, reliable and efficient, would be to configure the system in such a way that every day, at the closing of business, a complete image of VanMap be preserved live and fully functional within it, with of course the related attributes attached as metadata.[31] Of course, as these images would not be explicitly linked to any individual concrete business transaction, they would not be records. The record would be still the VanMap as a whole, but the presence in it of this historical stratification of data would make two things possible: first, accountability would be served, and second, the VanMap would become preservable as a record. The first outcome is quite obvious, while the second requires some explanation.

[31] The VanMap manager, having discussed this option with the case study team, found it not only technically feasible but also presumably exciting for several departments who would make good use of an historical stratification of data, such as the planning department. The idea of this option came to members of the team working on a science related case study, where the records creator maintains in a live system not only all the data of astronomic observations, and the images putting them into the needed relations, but also a stratification of the images taken at the end of each day representing all data accumulated in the previous twenty four hours.

VanMap, in its present configuration, constitutes a record that is constantly in a state of becoming and cannot be preserved. However, if the creator, in the usual and ordinary course of its business and for the purposes of its business, decided not only to have an historical stratification of daily images, but to remove them on a regular basis, yearly for example, from the active system into a live and fully functional, but inactive system, separate from the first by a firewall, thereby setting them aside as an indivisible whole related by procedure to all the city business which needs to use it either for reference or for accountability purposes, then each year of the VanMap could be regarded as a record with stable content and fixed form, that is a finished record, fulfilling all the requisites implied in the archival definition of record.[32] I

In conclusion, all InterPARES case studies have pointed to the fact that dynamic, interactive and experiential digital objects can only be records in becoming, potential records, that is. If the creator treats them as records, associates them with entities that are undeniably records, and do so in the course of activity and for its purposes, these objects only need a stable content and a fixed form to materialize themselves as fully finished records. If the acquisition of these two characteristics occurs as a matter of course at the hand of the creator, because the creator will need either to have records of the actions for which the written form is required by the legal system, to provide an account to itself and

[32] Some participants in the professional discourse on the concept of record have reduced it to a polarization between those who think that data or recorded information are records and those who think that data files can never be regarded as records. This is indeed a false dichotomy, because every entity that becomes associated in the usual and ordinary course of affairs with an archival aggregation (i.e., dossier, series, fonds), be it digital or not, regardless of its original nature, is subject to the law that governs every universality, according to which every individual entity that becomes a member of a collectivity, subject its individuality to the nature of the collectivity and shares the attributes of all other members of it (Archives as *universitas* is a very old concept. *Universitas* is the term from which the word university derives, a word referring to an institution whose members share the same nature, rights and duties with respect to the raison d'etre and the purposes of the institution). This is the reason why the VanMap used by the city staff is a record of the City, regardless of the fact that it is a collection of data, while the VanMap used by the public it is not.

to its stakeholders of the activities carried out, or to use the records itself for further action, future activities or reference, then the results will be records in all respects, by any analysis and standard (certainly from an archival, diplomatic, administrative, and legal point of view). A proviso is however required in the context of e-government and of business organizations: the stabilization of the content and the fixing of the form must be carried out by a neutral third party having no stake in the content of the records of any type of business transaction carried out by the creator, and who is competent in the science of the records. This is of course the definition of a record officer. On the contrary, as it regards the sciences and the arts, the author (be it an individual or a collective person), that is, the person who has the highest stake in the content of the record, would also be the most reliable person to make the record because of the nature and purposes of the scientific and artistic activities.

Meeting the challenge of identifying the record in the digital world is much more than establishing policies and procedures or developing metadata schemata. Every time a solution is proposed, more questions come up. Thus, I do not believe that a definitive answer that is valid in the context of all future technological environments can be provided, but what InterPARES 2 can do is to establish the conceptual underpinning, the parameters, and the method of analysis that will determine the answer to the question of what entity corresponds in a known given environment to the archival definition of the record, a definition that has survived the administrative and technological changes brought about by centuries of human activities and is likely to remain valid for the foreseeable future.