SAA 2000 (Denver, Colorado)

Title of Session: Requirements for Long-Term Authenticity of Electronic Records: The Initial Findings of the InterPARES Project
Date: Thursday, August 31, 3:30 p.m. to 5:00 p.m.
Title of paper: Conceptualizing an Authentic Electronic Record
Presenter: Heather MacNeil

As Seamus has explained, the InterPARES research is divided into four complementary domains of inquiry. The first domain is the responsibility of the Authenticity Task Force, and its purpose is to identify what elements of electronic records need to be preserved to ensure their authenticity. Historically and in the InterPARES project, authenticity is connected to a record's identity and integrity. An authentic record, whether electronic or non-electronic, is one that can be proven to be what it claims to be, free of falsification or corruption.

The work of the Task Force is being accomplished in three steps. The first, already completed, was to develop a template to guide our analysis of electronic records. The second, currently in progress, is to carry out four rounds of case studies of electronic records and electronic systems in order to test the validity of the template. The third and final step will be to establish a typology of electronic records based on authenticity requirements. I will speak about the first and third steps.

In order to identify the elements of an electronic record that are relevant to a consideration of its authenticity, it is necessary first to decompose an electronic record into its constituent parts; define each part; explain its purpose; and indicate whether, and to what extent, that part is instrumental in verifying the record's authenticity. The *Template For Analysis* was created for this purpose.

The disciplinary perspectives that have shaped the identification of the elements of an electronic record are those of archival science and diplomatics. The template builds specifically on the findings of an earlier research project carried out at the University of British Columbia entitled "The Protection of the Integrity of Electronic Records." Of course, the record elements identified in the UBC project have been substantially expanded and refined by the International Team based on its knowledge and experience with various kinds of electronic records and electronic systems.

Viewed from the integrated perspectives of diplomatics and archival science an electronic record is a complex of elements and their relations, consisting of:

- EXTRINSIC ELEMENTS
- INTRINSIC ELEMENTS
- ANNOTATIONS
- MEDIUM
- CONTEXT

Extrinsic elements are those elements of a record that constitute or make-up its external appearance. They include:

- PRESENTATION FEATURES
- SPECIAL SIGNS
- SEALS
- ELECTRONIC SIGNATURES
- DIGITAL TIME STAMPS

PRESENTATION FEATURES are the set of perceivable features that present the record's content to our senses. Such features include the overall configuration or representation of the content, e.g., text, drawing, image, moving images, sound, or some combination thereof; as well as particular aspects of the record's formal presentation that are necessary for it to achieve the purpose for which it was created, e.g., deliberately employed typefonts or colours, special layouts, hyperlinks, sample rates of sound files, resolution of image files, scales of maps.

SPECIAL SIGNS. Symbols which identify one or more of the persons involved in the compilation, execution, or receipt of the record and which are distinct from a signature or seal, e.g., digital watermarks, or an agency crest or logo.

SEALS. Traditionally, a seal is a specific means of authenticating a record and/or ensuring that it is only opened by the intended addressee. In an electronic environment, a digital signature, i.e., an electronic signature based on public key cryptography, comes closest to fitting that definition. The digital signature allows the recipient to verify the origin of the record and to check that the record is complete and unchanged. A digital signature is usually accompanied by an Authentication Certificate of a Trusted Third Party, which is an attestation issued by a trusted third party for the purpose of authenticating the ownership and characteristics of a public key.

Of course a digital signature is not completely analogous to a traditional seal. For one thing, a traditional seal is associated exclusively with a juridical person and the same seal is used to authenticate any record issued by that

person. A digital signature is associated exclusively with a juridical person (through the private key) and with a specific record and so no two records will have the same digital signature even when issued by the same person.

For another thing, a traditional seal is an extrinsic element specifically because it is visible to the human eye. The appearance of the seal in itself communicates a meaning and significance that is immediately comprehensible in human terms. You can usually determine its ownership just by looking at it. With a digital signature what is visible to the human eye is simply an indication of the presence of a digital signature and the indication itself is (to humans) an incomprehensible sequence of numbers. Therefore, it does not visibly communicate its meaning and significance in the way a traditional seal does.

A digital signature is a specific type of electronic signature, which is the next element.

AN ELECTRONIC SIGNATURE is a digital mark having the function of a signature in, attached to, or logically associated with a record, and that is used by a signatory to indicate her approval of the content of that record.

A DIGITAL TIME-STAMP ISSUED BY A TRUSTED THIRD PARTY. This is an attestation by a T.T.P. that a record was sent at a particular point in time.

Whereas extrinsic elements refer to a record's external appearance, intrinsic elements refer to its internal composition or articulation. These are discursive elements within the record that communicate the action in which it participates and its immediate context. They fall roughly into 3 groups:

The first group consists of elements that convey aspects of the record's juridical and administrative context, i.e.,

- Name of Author
- Name of Originator
- Chronological Date
- Name of Place of Origin of Record
- Name of Addressee(s)
- Name of Receiver(s)

The second group consists of elements that communicate the action itself, i.e,

- Indication of Action (Matter) Subject line; caption
- Description of Action (Matter)

And the third group consists of elements that convey the record's documentary context and its means of validation, such as:

- Name of Writer
- Corroboration, i.e., explicit mention of the means used to validate the record
- Attestation, that is, the written validation of a record by those who took part in the issuing of it (author, writer, countersigner) and by witnesses to the action or to the signing of the record; and
- Qualification of Signature (the mention of the title and capacity of the persons signing a record)

Next we have annotations. These are additions made to a record after it has been created. Like the intrinsic elements, annotations fall roughly into 3 groups. The first includes additions made to the record after its creation as part of

the execution phase of an administrative procedure. Traditionally, this sort of annotation has been used only for the authentication and registration of records whose form is required by law. For example, the registration number added to a land deed by the land registry office, or the statement of the authenticity of the signatures in a will. For specific types of electronic records, namely, electronic mail records, the date, time, and place of transmission, and the indication of attachments also belong to this group.

The second group of annotations consists of additions to the record made in the course of handling the business matter in which the record participates, e.g., name of handling office, comments, notes and dates of transmission to other offices.

The third and final group are additions to the record made in the course of handling it for records management purposes, e.g., classification code, registration number, draft/version number, cross-references to other records.

In the InterPARES project, it is taken for granted that a record is a representation of a fact or act that is memorialised on a physical carrier, i.e., a medium and preserved by a physical or juridical person in the course of carrying out its activities.[1] It follows that a record does not exist until its elements have been inscribed on or affixed to a medium. On the basis of that reasoning, medium has been included in the template as a record element. Of course, with electronic records, storage on a hard, floppy, or optical disk, or on a magnetic tape, is necessary but not sufficient to re-present the content and form of a

[1] Maria Guercio, "Principi, metodi e strumenti per la formazione, conservazione e utilizzo dei documenti archivistici in ambiente digitale," *Archivi per la storia* XII, 1-2 (1999): 26.

record because what is inscribed on or affixed to the medium is not a record as such (or words, or pictures), but a bitstream. Re-presentation of an electronic record's content and form also requires the capacity to process the record through software.[2] Moreover, while affixing a bitstream to a medium is an essential pre-condition to the existence of an electronic record, this does not mean that the medium is an essential or even a relevant factor in verifying that record's authenticity. We assume that it is neutral with respect to the record's authenticity at least from the perspective of the records creator and the records preserver. The validity of that assumption will be confirmed or refuted in the course of analysing the case studies.

In the final set of elements, we move from the record itself to its context, meaning the broader structural, procedural, and documentary framework in which the record is created and managed. The identified elements correspond to specific frameworks and include:

- Juridical-Administrative Context, i.e., the legal and organizational system in which the records creator is situated. Indicators of juridical-administrative context are external laws, regulations, etc, that control how the creator conducts its business and manages its records.
- Provenancial Context refers to the mandate, structure, and functions of the records creator. Indicators are organizational charts, annual reports, etc.

[2] Ken Thibodeau, "Certifying Authenticity of Electronic Records: Interim Report of the Chair of the Preservation Task Force to the InterPARES International Team," (19 April 2000).

- Procedural Context refers to the business procedure in the course of which the record is created. Indicators are workflow rules, codes of administrative procedure, classification schemes, etc.
- Documentary Context refers to the broader aggregation (the fonds) to which the record belongs and its internal structure. Indicators are classification schemes, record inventories, indexes, registers, etc.
- Technological Context refers to the technological environment surrounding the record. This element, necessarily, has been broken down into numerous sub-elements, including
- Hardware (the storage, microprocessor, network, peripheral devices, and architecture)
- Software (the operating system, system software, network software, and application software)
- Data (the file structure, and file format)
- System models (i.e., abstract representations of the entities, activities and/or concepts in the system as well as their attributes, characteristics, and the functional relationship between them); and
- System administration (i.e., the set of procedures that ensure correct, secure, reliable, and persistent operation of the system).

The template for analysis, it should be stressed, is an idealized representation developed for the purpose of identifying all the known elements of an electronic record. No single record will, or should, include all the elements; the absence or presence of one or more of them in a specific instance will depend on

the record's purpose. For example, many types of records will not contain a signature because the procedure in which those records are created in itself validates them.

The purpose for developing the template and testing it through case studies is to define conceptual requirements for authenticity that can be translated into specific methods and procedures by the Appraisal and Preservation Task Forces. We anticipate two levels of requirements, the first level consisting of baseline or threshold requirements applicable to all electronic records and the second consisting of specific requirements associated with distinct types of electronic records. An electronic records typology is being developed as an aid to the identification of the requirements.

A typology is a system of groupings, usually called types, which are classes of things, persons, or events that have specific common attributes. The groupings are set up to aid inquiry by establishing a relationship among phenomena. The particular order elicited by a typology therefore will depend on the purpose of the investigation and on the phenomena so arranged. In anthropology and archaeology, for example, typologies have been based on variations in the style of artifacts, or variations in burial customs, social systems or ideologies. As Seamus Ross has pointed out, whatever the object under consideration, a typology must take into account the significant attributes of the object itself, its relationship to other objects, the processes of its production, and the meaning of the object to its maker. [3]

[3] Seamus Ross, "Dress-pins from Anglo-Saxon England: their production and typo-chronological development," D.Phil. dissertation, University of Oxford, 1992.

Traditional diplomatists have constructed typologies of medieval records based on variations in their documentary form and in chancery practice over time and across different juridical systems. The purpose of such typologies was to enable diplomatists to identify and characterize centuries-old records of uncertain provenance. They were, therefore, retrospective in orientation.

The purpose of the InterPARES typology of electronic records is substantially different. That purpose is to identify the specific elements of electronic records that must be preserved, over time and across technologies, in order to verify the records' authenticity. It follows that a typology based on the requirements for authenticity cannot be locked into either current record forms or current application classes because both of these, inevitably, will change.

We have determined, therefore, that the typological framework should be based on the relationship between a record and the action in which it participates. That relationship is shaped, in turn, by the juridical system in which the record is created. A juridical system is a collectivity, e.g., a corporate body, organised on the basis of a system of rules; the rules themselves may be imposed from outside or inside that collectivity.

Within this framework, which is based on contemporary diplomatics, there are four types of records.

DISPOSITIVE, PROBATIVE, SUPPORTING, NARRATIVE

A *dispositive* record is one whose written form is required by the juridical system as the essence and substance of an action, i.e., the action comes into existence with the creation of the record. A *probative* record is one whose written

form is required by the juridical system as proof that an action has taken place. The record is not contemporaneous with the action, as is the case with a dispositive record, but it procedurally follows it. A *supporting* record is similar to a probative record in that it is separate from, but procedurally linked to, an action. Unlike a probative record, however, its written form is not required by the juridical system and does not constitute proof of an action's occurrence. Finally, a *narrative* record is one that an individual creates for her own purposes in the course of carrying out an action. It is not procedurally linked to an action and its creation is entirely discretionary.

These four types represent main classes of the typology. Within each of these types, there will be subtypes and variants, corresponding to unique characteristics of specific records identified within each of the four types.

This typological framework will help us to determine how much of the active record's form, content and context must be preserved during its inactive life. It will also help us to determine how rigid we need to be in defining what constitutes a record within a given recordkeeping context.

The virtues of this particular framework are two-fold: first, it is broad enough to accommodate new electronic record forms; secondly, because it is tied to the consequences specific types of records are intended to generate, the framework can accommodate the identification of a range of requirements possessing varying degrees of rigor. As a corollary, it provides a useful basis on which records creators and preservers can assess the risk of non-compliance with specific requirements.

The typology is an essential first step in identifying the kinds of descriptive metadata and procedural documentation that must be carried forward with electronic records in order to preserve them as authentic memory and evidence for future generations. Once the typology is completed, it falls to the Appraisal and Preservation Task Forces to carry out the remaining steps in that work of identification.