Providing Grounds for Trust: Developing Conceptual Requirements for the Long-Term Preservation of Authentic Electronic Records

HEATHER MACNEIL

RÉSUMÉ Depuis 1999, le projet InterPARES (International Research on Permanent Authentic Records in Electronic Systems) se penche sur les problèmes reliés à la conservation à long terme de documents électroniques authentiques. La formulation des exigences conceptuelles nécessaires à la vérification de l'authenticité des documents électroniques est sous la responsabilité du groupe de travail sur l'authenticité. Celui-ci a divisé ses tâches en trois étapes: (1) identifier et définir, en utilisant la diplomatique archivistique contemporaine, les éléments d'un document électronique qui sont liés à son authenticité; (2) vérifier la validité de ces éléments au moyen d'études de cas de systèmes électroniques; et, (3) développer des exigences générales et spécifiques pour la conservation à long terme de documents électroniques authentiques. Cet article présente le travail qui a été accompli jusqu'à présent par le groupe de travail dans chacune de ces trois étapes.

ABSTRACT Since 1999, the International Research in Permanent Authentic Records in Electronic Systems (InterPARES) Project has been investigating the issues associated with the long-term preservation of authentic electronic records. The identification of conceptual requirements for the verification of authentic electronic records is the responsibility of the InterPARES Authenticity Task Force. The work of the task force is being carried out in three stages: (1) identifying and defining, using contemporary archival diplomatics, the elements of an electronic record that are relevant to a consideration of its authenticity; (2) testing the validity of the elements through case studies of electronic systems; and (3) developing general and specific requirements for the preservation of authentic electronic records over the long term. This article reports on the work accomplished by the task force to date in each of the three stages.

According to Webster's dictionary, *authentic* means "worthy of acceptance or belief as conforming to or based on fact ...; conforming to an original so as to reproduce essential features ...; made or done the same way as an original." Authentic is synonymous with the terms *genuine* and *bona fide*. *Genuine* "implies actual character not counterfeited, imitated, or adulterated [and] connotes definite origin from a source." *Bona fide* "implies good faith and sincerity of intention."[1]

1 *Merriam-Webster Online Dictionary*, <http://www.m-w.com/cgi-bin/dictionary>, s.v. "authentic."

It follows that an authentic record is one that can be proven to be (i) what it claims to be and (ii) free of falsification or inappropriate modification. The authenticity of a record is assessed in relation to its identity (i.e., was it written by the person who purports to have written it?) and its integrity (i.e., has it been altered in any way since it was first created and, if so, has such alteration changed its essential character?). Proving the authenticity of a record thus implies the need to preserve its identity and integrity over time.

Preserving a record's identity and integrity over time is predicated on its endurance and stability over time. According to David Levy, a computer scientist who has studied the nature of documents in the digital age:

> Assessments of authenticity in the world of paper and other stable, physical media rely heavily on the existence of enduring physical objects. If you want to determine whether the document in front of you is the unique individual it purports to be (someone's last will and testament, for example), you can try to determine its history. But you can do this only because it *has* a history, an extended existence in time.[2]

Preserving the identity and integrity of a record in the digital world is complicated by the fact that, in such a world, there are no stable and enduring physical objects. As Ken Thibodeau observes, "strictly speaking, it is not possible to preserve an electronic record. It is only possible to preserve the ability to reproduce an electronic record. It is always necessary to retrieve from storage the binary digits that make up the record and process them through some software for delivery or presentation."[3] Given that exact replication of digital objects is unfeasible and that loss and change are inevitable and unavoidable in the digital world, on what grounds should we base our trust in the authenticity of digital objects that will be preserved over the long term?

The need to establish specific and defensible grounds for such trust is the driving force behind a number of current research initiatives, including the InterPARES[4] project. InterPARES focusses its attention on a specific class

2 David Levy, "Where's Waldo? Reflections on Copies and Authenticity in a Digital Environment," *Authenticity in a Digital Environment* (Washington, D.C., 2000), p. 30.

3 Ken Thibodeau, "Certifying Authenticity of Electronic Records: Interim Report of the Chair of the Preservation Task Force to the InterPARES International Team," unpublished report (19 April 2000), p. 1.

4 The InterPARES project (the acronym stands for "International Research on Permanent Authentic Records in Electronic Systems") began in January 1999 and will conclude in January 2002. The researchers in InterPARES are an international and multi-disciplinary group consisting of archival scholars and practitioners as well as scholars and other specialists drawn from the humanities and social sciences, and from the computer, mathematical, and chemical sciences. A number of national archival institutions are also participants in the project. A detailed description of the project, including its origins, goals, objectives, and methodology, may be found on the project's Web site at <http://www.interpares.org>.

of digital objects, i.e., electronic records.[5] Its overarching goal is to "develop the theoretical and methodological knowledge essential to the permanent preservation of authentic records generated and/or maintained electronically, and, on the basis of this knowledge, to formulate model policies, strategies, and standards capable of ensuring that preservation."[6] To accomplish that goal, the project is divided into four complementary domains of inquiry: (1) conceptual requirements for preserving authentic electronic records; (2) appraisal criteria and methods for selection of authentic electronic records; (3) methods and responsibilities for preserving authentic electronic records; and (4) framework for the formulation of policies, strategies, and standards. This article explores the work that has been accomplished to date in the first domain of inquiry.

Research in the first domain, which provides the foundation for the three subsequent domains, is the responsibility of the Authenticity Task Force of InterPARES.[7] The work of the task force is being carried out in three steps. The first step is to identify and define, in the abstract, the elements of an electronic record that are relevant to a consideration of its authenticity. The second step is to test the validity of the elements through case studies of electronic systems. The third step is to develop, on the basis of the findings in the first two steps, conceptual requirements for the preservation of authentic electronic records over the long term.

The disciplinary perspective that has shaped the identification of the elements is contemporary archival diplomatics.[8] Viewed from this perspec-

5 For the purpose of the project an electronic record is defined as a record created in electronic form. A record is defined as any document created – meaning made or received and set aside either for action or reference – by a physical or juridical person in the course of practical activity as an instrument and by-product of it.

6 InterPARES Project, "Project Background," available on the project Web site.

7 The members of the Authenticity Task Force are: Heather MacNeil (Chair), Luciana Duranti, Anne Gilliland-Swetland, Maria Guercio, Babak Hamidzadeh, Sue McKemmish, John Roeder, Seamus Ross, and Wai-kwok Wan.

8 Contemporary archival diplomatics is an adaptation of traditional diplomatic concepts and methods to contemporary record-keeping environments and an integration of these concepts and methods with those of archival science. It provided the conceptual foundation for a three-year project carried out between 1994 and 1997 at the University of British Columbia entitled "The Preservation of the Integrity of Electronic Records." The goal of the UBC project was to identify and define conceptually the nature of an electronic record and the conditions necessary to ensure its reliability and authenticity based on the concepts and methods of diplomatics and archival science. This work resulted in the identification of the elements of a *record*, a *reliable record*, and an *authentic record* in both paper and electronic record-keeping environments. For an overview of the findings of the UBC project see Luciana Duranti and Heather MacNeil, "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project," *Archivaria 42* (Fall 1996), pp. 46–67. The elements of an electronic record included in the template for analysis draw specifically

tive, an electronic record, like its traditional counterpart, is a complex of elements and their relationships. It possesses a number of identifiable characteristics,[9] among them: a fixed documentary form,[10] a stable content, an archival bond with other records either inside or outside the system, and an identifiable context. It participates in or supports an action, either procedurally or as part of the decision-making process (meaning its creation may be mandatory or discretionary), and at least three persons (author, writer, and addressee) are involved in its creation (these three conceptual persons may in fact be only one physical or juridical person).

In a traditional record-keeping environment, these characteristics manifest themselves in explicit and implicit ways. For example, the archival bond may be expressed in a classification code or some other unique identifier that appears on the face of a record. The names of the author and addressee typically appear in the "to" and "from" fields in a memorandum. The name of the author may appear in the letterhead in other types of records. The action or matter to which the record relates is typically expressed in a subject line in a textual record or in a caption in a visual record. The purpose served by these individual elements also depends on their specific form of expression. For example, the identification of the name of the author that appears in the letterhead serves the purpose of identifying the record's immediate juridical-administrative context. When that same name appears as a signature at the bottom of the record, it serves the purpose of attesting the validity of the record or its content, or both.

The working hypothesis of the task force is that, while they may manifest themselves in different ways, these same or similar elements are present, either explicitly or implicitly, in electronic records. To test that hypothesis, the task force has created a template for analysis.[11] The template is a decom-

---

on those identified in the UBC project. At the same time, the elements have been substantially revised and extended by the InterPARES researchers based on their combined knowledge and experience with various kinds of electronic records and electronic systems.

9 These characteristics are identified as selection criteria in the Authenticity Task Force, "[Draft] Research Methodology Statement," 7 November 2000. The statement is available on the project Web site.

10 According to the research methodology statement, a fixed form "means that (1) the binary content of the record, including indicators of its documentary form, are stored in a manner that ensures it remains complete and unaltered; and (2) technology has been maintained and procedures defined and enforced to ensure that the content is presented or rendered with the same documentary form it had when it was set aside."

11 The Template for Analysis is available for viewing on the InterPARES Web site. See Authenticity Task Force, "[Draft] Template for Analysis," 7 November 2000. Unless otherwise indicated, definitions of the elements of an electronic record included in the template are drawn from the "Template for Analysis." The following student researchers at the University of British Columbia have assisted the Authenticity Task Force in the development and interpretation of the template: Marta Maftei, Ian McAndrew, Shauna McRanor, April Miller,

position of an electronic record into its constituent elements which defines each element, explains its purpose, and indicates whether, and to what extent, that element is instrumental in verifying the record's authenticity. The validity of the template is being tested through four rounds of case studies of electronic systems that either contain, generate, or have the potential to create electronic records. Two rounds of case studies have been carried out in Canada, the United States, the United Kingdom, the Netherlands, and Italy and cover both public and private sector agencies. The studies completed thus far include large and small scale databases (used to manage, for example, student records, financial aid, securities transactions, granting of patents, and the registration of last wills), document management systems (used to support agency-wide administrative functions, such as the drafting and management of procedures, as well as specific operational functions, such as the issuing of permits for the transportation of hazardous waste), a geographic information system (used to manage mappable thematic data related to land inventory and land use), and a Web-based application system (used to support on-line trademark applications).

The purpose for conducting the case studies is to assess whether and to what extent the elements identified in the template are present in the systems being examined, as well as to identify any relevant elements present in these systems that are not taken into account in the template. The case studies will assist the researchers in determining whether and how the elements are brought together as a record, e.g., are the elements embedded in the record, or are they linked to it? If they are linked to the record, how determined and enforced is that link? Do the elements manifest themselves in ways that are similar to the way they manifest themselves in traditional records or is their manifestation different? Finally, the case studies will assist the researchers in ascertaining which specific elements the creator considers essential for verifying the record's authenticity and the kinds of procedural controls exercised over the systems and the records contained within them which, in the creator's view, support a presumption of authenticity.

The elements of an electronic record included in the template for analysis fall into four main categories: *documentary form* (which includes *intrinsic elements* and *extrinsic elements*), *annotations*, *context*, and *medium*. The elements examined in the categories of documentary form and annotations are those that are (conceptually at least) inside the record, i.e., they are visible on the face of the record, or embedded in it, or linked to it. The elements examined in the category of context are those that are outside the record, i.e., they are part of the larger documentary and administrative framework in which the records are created, maintained, and used. Medium is considered

Lara Wilson, and Joleen Wright.

to reside both inside and outside the record.

Documentary form is defined as the rules of representation according to which the content of a record, its immediate administrative and documentary context, and its authority are communicated. Documentary form possesses both extrinsic and intrinsic elements. Intrinsic elements refer to a record's internal composition or articulation. These are discursive elements within the record that communicate the action in which it participates and its immediate context. Intrinsic elements fall into three groups: *elements that convey aspects of the record's juridical and administrative context* (e.g., the name of the author and addressee); *elements that communicate the action itself* (e.g., the indication of the subject or matter); and *elements that convey aspects of the record's documentary context and its means of validation* (e.g., the name of the writer, attestations). With traditional records, the three groups of elements typically corresponded to three physical subsections of a record: indications of the record's juridical and administrative context were found in the protocol (i.e., the top part of the record), indications of the action of which the record formed a part were located in the text (i.e., the main body of the record), and indications of the record's documentary context and means of validation appeared in the eschatocol (i.e., the bottom part of the record). While this correspondence continues to exist in some types of electronic records, it does not by any means prevail in all types.

Extrinsic elements refer to specific features of the record's external appearance that are instrumental in communicating and achieving the purpose for which the record was created. For traditional diplomatists examining medieval acts, extrinsic elements, which could only be examined on the original document, constituted the first and most obvious proof of authenticity. Such elements included the layout, paragraphing, colour of ink, type and size of letters, and so on, as well as the seals moulded into or appended to the record. For electronic records, *presentation features*, *electronic signatures*, *electronic seals*, *digital time stamps*, and other *special signs* are treated as extrinsic elements. Although, in an electronic environment, these elements manifest themselves somewhat differently than their traditional counterparts, their purpose is analogous.

The intrinsic elements of form that convey aspects of the record's juridical and administrative context include the *name of the author*,[12] the *name of the originator*,[13] the *chronological date*,[14] the *name of the place of origin of the*

12 The author is the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.

13 The originator is the physical or juridical person assigned the electronic address in which the record has been generated and/or sent.

14 The chronological date is the day, month, year, and, possibly, the time of the record included in the record by the author or the electronic system on the author's behalf in the course of its compilation.

record,[15] the *name(s) of the addresee(s)*,[16] and the *name(s) of the receiver(s)*[17] (i.e., recipients) of the record. In an electronic record-keeping environment, the type of system in which the records are created, maintained, and used will determine whether the inclusion of all or some of these elements is mandatory or discretionary and whether they are added by the author or by the electronic system on the author's behalf. For electronic records maintained in document management systems, for example, many of these elements are included in the profile associated with the record. With electronic mail records, the names of the author and originator, addressee, and receiver all appear in the top portion of the record (i.e., in a header). The author's name may only appear in the form of an attestation (in which case it is considered below under the elements of validation and documentary context). In certain kinds of electronic records (like with certain kinds of traditional records) the name of the author will not be mentioned explicitly but it may be inferred from the record's context. Similarly, the name of place of origin of a record may not be explicitly identified but it may be inferred from a filing prefix (in which case it is considered below under annotations).

The elements that communicate the action itself include the indication and description of the action or matter. For textual records, the *indication of the action or matter* typically appears as a subject line(s) or a title at the top of the record; in other types of records, such as images, it may take the form of a caption. The subject may only be identifiable through a classification code (in which case it is considered below as an annotation). The *description of the action or matter* (i.e., the record's content) typically occupies the body of the record and refers to the message the record is intended to convey. Depending on the type of record, the content may be entered directly by an individual or extracted, in whole or in part, from the electronic system. It may be standardized or free form.

Since a stable content is considered one of the identifying characteristics of a record, the case studies will seek to determine at what point in time the content is considered complete, stable, and unchangeable. If there is no such point in time, the question then becomes: in what specific ways can the content be changed – by addition of new content, by deletion or substitution of existing content? If the content can be changed, who has the authority to make that change, and how and to what extent are such changes tracked by the system?

15 The name of the place of origin of the record is the name of the geographic place where the record was generated, included in the content of the record by the author or the electronic system on the author's behalf.

16 The addressee is the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

17 The receiver is the person to whom the record is copied for information purposes.

The visible means by which the content of an electronic record is communicated is governed by *presentation features*, which are included among the extrinsic elements of form. Presentation features are the set of perceivable features generated by means of encoding and programme instructions, which are capable, when used individually or in combination, of presenting a message to our senses. Such features include the overall configuration or representation of the content, e.g., text,[18] graphic,[19] image,[20] moving images,[21] sound,[22] or some combination thereof. They also include particular aspects of the record's formal presentation that are necessary for it to achieve the purpose for which it was created, e.g., standardized spacing and fonts, deliberately employed colours, special layouts (e.g., spreadsheets), hyperlinks, sample rates of sound files, resolution of image files, scales of maps. Understanding the role such elements play in communicating a record's content is essential to determining whether, and to what extent, these presentation features will be preserved in certain records over time.

The intrinsic elements that convey the record's documentary context and its means of validation include the *name of the writer*[23] (which may be explicitly identified or simply implied from the name of the author or the record's context), the *attestation*, *corroboration*, and the *qualification of signatures*. In traditional records, the attestation is the commonest means of validation and it consists of 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. In traditional records, attestations usually appear as signatures at the bottom of the record. However, some records carry the attestation in the protocol, e.g., in a memorandum signed or initialled beside the superscription. In some records, the qualification of signature, i.e., the mention of the title and capacity of the persons signing a record, may appear in conjunction with an attestation.

In an electronic record-keeping environment, the attestation may assume a number of forms, for example, a scanned image of a handwritten signature in

18 Text is defined as words, numbers, or symbols.

19 Graphic is defined as a representation of an object or outline of a figure, plan, or sketch by means of lines; a representation of an object formed by drawing.

20 Image is defined as an artificial imitation or representation of the external form of any object, or an optical appearance or counterpart of an object, such as is produced by rays of light, refracted as through a lens, or falling on a surface after passing through a small aperture.

21 Moving images, which are a subset of image, are defined as visual images, with or without sound that, when viewed, present the illusion of motion.

22 Sound is defined as an aural representation of words, music, or any other manifestation of sound.

23 The writer is the person having the authority and capacity to articulate the content of the record. It may be the same name as the author and/or originator of the record.

a word processing document (its weight as an attestation will depend on whether the scanned image is subject to procedural controls that prevent its misuse, such as maintaining it in a restricted part of a database), the name of the author as it appears in the header of an electronic mail message, or the name that is included in a document profile (assuming that the assignment of the name included in the header or profile is subject to strict technical and procedural controls). The qualification of signature that accompanies an attestation may be added by the writer or automatically assigned by the electronic system.

The extrinsic elements of form that are closely associated with the attestation function in an electronic record-keeping environment are *electronic signatures* and *electronic seals.* In the area of electronic commerce and contracting law, electronic signatures are becoming the standard method of authentication for electronic records. In the template for analysis, an electronic signature is defined as a digital mark having the function of a signature in, attached to, or logically associated with a record, and which is used by a signatory to indicate her approval of the content of that record. A number of electronic signature techniques, such as electronic pens and digital signatures, are currently being used or are under development. Pen-based electronic signatures rely on authentication through a biometrical device based on handwritten signatures:

> In such a device, the signatory would sign manually, using a special pen, either on a computer screen or on a digital pad. The hand-written signature would then be analysed by the computer and stored as a set of numerical values, which could be appended to a data message and displayed by the recipient for authentication purposes. Such an authentication system would presuppose that samples of the hand-written signature have been previously analysed and stored by the biometrical device.[24]

Digital signatures, on the other hand, rely on public key cryptography. Public key cryptography is based:

> ... on the use of algorithmic functions to generate two different but mathematically-related "keys" (i.e., large numbers produced using a series of mathematical formulae applied to prime numbers). One such key is used for creating a digital signature or transforming data into a seemingly unintelligible form, and the other one for verifying a digital signature or returning the message to its original form. ... The complementary keys used for digital signatures are named the "private key" which is used only by the

24 United Nations Commission on International Trade Law, *Draft Guide to Enactment of the UNCITRAL Uniform Rules on Electronic Signatures* A/CN.9/WG.IV/WP.86 (New York, 2000), p. 16, para. 31.

signatory to create the digital signature, and the "public key" which is ordinarily more widely known and is used by the relying party to verify the digital signature. ... In addition to the generation of key pairs, a "hash function" is used in both creating and verifying a digital signature. A hash function is a mathematical process, based on an algorithm which creates a digital representation or compressed form of the message, often referred to as a "message digest" or "fingerprint" of the message, in the form of a "hash value" or "hash result". ... Any change to the message invariably produces a different hash result when the same hash function is used. To sign a document ... the signatory first delimits ... what is to be signed. Then a hash function in the signatory's software computes a hash result unique ... to the [document] to be signed. The signatory's software then transforms the hash result into a digital signature using the signatory's private key. The resulting digital signature is thus unique to both the [document] being signed and the private key used to create the digital signature. ... Verification of a digital signature is accomplished by computing a new hash result of the original message by means of the same hash function used to create the digital signature. Then, using the public key and the new hash result, the verifier checks whether the digital signature was created using the corresponding private key, and whether the newly computed hash result matches the original hash result that was transformed into the digital signature during the signing process.[25]

Though they assume different forms, both these techniques share a common purpose, i.e., "to provide functional [and legally binding] equivalents to (1) hand-written signatures; and (2) other kinds of authentication mechanisms used in a paper-based environment (e.g., seals or stamps)."[26]

In the template, digital signatures are considered an example of *electronic seals*. This is because digital signatures are functionally analogous (though not equivalent) to medieval seals in general and the sovereign's seal in particular. Medieval seals performed three functions: "closure and guarantee of the integrity of ... texts; claim and proof of ownership; and authentication of documents, converting them into executory instruments by affirming that the text represents the sealer's will."[27] The affixing of a seal did not simply furnish a medieval document with a means of proving its genuineness. It also rendered that document indisputable as to the terms of the transaction it recorded. The non-repudiation function of the medieval seal stemmed from the Germanic principle concerning the indisputability of the king's word according to which "Who gives him the lie forfeits life."[28] The king's seal

25 Ibid., pp. 17–18.

26 Ibid., p. 15, para. 20.

27 Brigitte Bedos Rezak, "Seals and Sigillography, Western European," in Joseph R. Strayer, ed., *Dictionary of the Middle Ages,* Vol. 11 (New York, 1989), p. 124.

28 John Henry Wigmore, *Evidence in Trials at Common Law*, Vol. 9, ed. and rev. by James H. Chadbourn (Boston, 1978), para. 2426. Hereafter cited as *Wigmore on Evidence*.

to a document therefore rendered its truth incontestable. As the use of the seal extended downward from the king to the people at large it carried this non-repudiation function along with it.[29] The authority of the medieval seal also derived from the controls exercised over the matrix used to make the seal's impression. According to Brigitte Bedos Rezak, "the matrix might not be lost, stolen, or misused without serious consequences for its owner and, in these circumstances, would be publicly disclaimed. Matrices were routinely changed upon modification of the owner's social status, title, or function; and at the owner's death the matrix was defaced, destroyed, or buried with him. ... By the fourteenth century, custom called for the destruction of royal, imperial, and papal matrices at the death of its owner."[30]

The digital signature is characterized as an electronic seal because, like the traditional seal, it allows the recipient to verify the origin of the record and check that it has not been altered during its transmission. The authority and indisputability of a digital signature depends on the verifier having access to the signatory's public key and obtaining some assurance that it corresponds to the signatory's private key. One means of providing that assurance is to use one or more trusted third parties to associate an identified signatory or the signatory's name with a specific public key. The trusted third party is generally referred to as a certification authority. The certificate issued by a certification authority accompanies a digitally signed record and serves to authenticate the ownership and characteristics of a public key. Certification authorities, in turn, may be organized hierarchically into what is commonly referred to as a public key infrastructure (PKI). According to Clifford Lynch, a computer scientist and executive director of the Coalition for Networked Information, the procedures of a PKI may be trusted to accomplish the following:

- To verify, according to published policies, a user's right to an "identity" and to subsequently document the binding between the identity and a public/private key pair. ...
- To provide a means for determining when a key pair/identity binding has been compromised, expired, or revoked and should no longer be considered valid.[31]

Of course a digital signature is not completely analogous to the medieval seal. For example, a traditional seal is associated exclusively with a physical

29 Ibid.

30 Brigitte Bedos Rezak, "Seals and Sigillography," p. 127.

31 Clifford Lynch, "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust," *Authenticity in a Digital Environment*, pp. 44–45. For a detailed discussion of PKI, see UNCITRAL, *Draft Guide to Enactment of the UNCITRAL Uniform Rules on Electronic Signatures*, pp. 19–22.

or juridical person and the same seal is used to authenticate any record issued by that person. A digital signature is associated with a specific physical or juridical person *and* a specific record. Consequently, no two records will have the same digital signature even when issued by the same person. Moreover, a digital signature, in itself, does not communicate its meaning and significance with the same immediacy as a traditional seal. A seal appended to or moulded into a record is a tangible visual symbol of the owner's authority and identity. In contrast, the digital signature attached to and transmitted with an electronic record is simply a hash result that manifests itself as an incomprehensible sequence of numbers.

Other extrinsic elements of form associated with attestation and identification are *digital time stamps issued by a trusted third party* and *special signs*. Digital time stamps are typically used in situations involving legal relationships where proof of the exact time that a record was transmitted or received is critical to establish rights (e.g., intellectual property rights) or avoid liability (e.g., in contracts). In these situations, the digital time stamp provided by a certification authority or other trusted third party serves as an attestation that a record was transmitted or received at a particular point in time.

Finally, *special signs* are symbols that 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. In medieval documents, such signs typically included the chrismon, the signum manus, or the monogram. Special signs that may be found in or on electronic records include identifiers that use symbols or images rather than words to identify the author, originator, or writer of a record (e.g., an agency crest, a personal logo). Digital watermarks used to protect intellectual property are another type of special sign related to identification and attestation.[32]

In addition to an attestation, certain kinds of records may also include a *corroboration*, which is the explicit mention of the means used to validate the record and guarantee its authenticity. For example, an official student transcript issued by the University of British Columbia includes the phrase, "Issued under the seal of the University of British Columbia." An example of a corroboration specifically associated with digital signatures is the certificate issued by a certification authority, which accompanies a digitally signed record. The information provided in the certificate will depend on the level of trust that is required between the parties in a particular transaction but, typically, it will include the name or pseudonym of the signatory, the name

32 A digital watermark is a copyright claim that is attached to a digital object. Digital watermarks raise a number of authenticity-related issues. For a discussion of some of these issues see Ibid., pp. 42–44.

of the certification authority, the public key of the signatory, the algorithm, and the type of key.

Annotations, i.e., additions made to a record after it has been created, constitute the next category of elements included in the template for analysis. Annotations are an important means by which a record's archival bond as well as its documentary and administrative context are expressed. In medieval documents, annotations typically took the form of chancery or notarial notes, which were added on the bottom of the document or on its verso. In contemporary bureaucratic record-keeping environments, the annotations that either appear on the face of a record, or are linked inextricably to it, assume a wide variety of forms.

Annotations fall into three basic groups. The first group 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. Digital signatures, which function as attestations, are considered to belong also to this group of annotations.

The second group consists of *additions made to the record in the course of handling the business matter in which the record participates*. Examples of this type of annotation include, but are not limited to, the identification of the name of the office handling the matter, comments noted on the face of the record or embedded in it, and dates of transmission to other offices. The manner in which such annotations manifest themselves in an electronic record-keeping environment depends on the application being used. For example, word-processing applications typically provide for the insertion of comments into a record, along with the identification of the individual making the comment and the date. These comments are embedded in the record and may be viewed by clicking on highlighted text. In other types of applications, annotations made in the course of handling the matter are included in the profile associated with the record or its functional equivalent.

The third group of annotations consists of *additions made to the record in the course of handling it for records management purposes*. Such additions typically include the classification code or file number assigned to the record, its draft and/or version number, cross-references to other records, the identification of the records creator (i.e., the person in whose fonds the record belongs), an indication of scheduling actions, and so on. As with the previous category, how these annotations manifest themselves in an electronic environment depends on the application. In document management applications, for example, annotations of this type are typically found in the

profile. The profile itself is also considered an annotation (as well as a repository of annotations) because it is inextricably linked to a record and exists for as long as the record does.

The final two categories of elements included in the template are *context* and *medium*. The examination of a record's context shifts the analysis away from the record itself to the broader structural, procedural, and documentary framework in which the record is created and managed. The identified elements of context correspond to a hierarchy of frameworks ranging from the general to the specific. They include the record's *juridical-administrative context*, its *provenancial context*, its *procedural context*, its *documentary context*, and its *technological context*. Although the record itself may contain indications of one or more of these contexts (e.g., the classification code or file number that appears on the record or in its profile is a kind of shorthand indication of the record's documentary, procedural, and provenancial contexts), the greater part of our understanding derives from an examination of sources outside the record (although all or some of these sources may be incorporated into the electronic system in which the records reside). Indicators of the juridical-administrative context are laws and regulations external to the creator that control how the creator conducts business and manages records. Indicators of provenancial context include organizational charts, annual reports, and so on that identify the creator's structure, mandate, and functions. Indicators of procedural context include workflow rules, codes of administrative procedure, task lists, classification schemes, and so on that explain the business procedure in the course of which the record is created, maintained, and used. Indicators of documentary context include classification schemes, record inventories, indexes, registers, and so on that situate the record within the broader aggregation to which it belongs (i.e., the fonds). Specific indicators of the record's technological context include workflow models, data models, and so on that explain the technological environment surrounding the record, including the hardware, software, data, system models, and system administration.[33]

An examination of these contexts is important to understand, among other things, the business processes in the course of which electronic records are created, maintained, and used, the types of records generated from these processes, and the connection between those processes and the creator's

33 Hardware refers to the storage, microprocessor, network, peripheral devices, and architecture. Software refers to the operating system, system software, network software, and application software. Data refer to the file structure and file format. System models refer to the abstract representations of the entities, activities, and/or concepts in the system as well as their attributes, characteristics, and the functional relationship between them. System administration refers to the set of procedures that ensure correct, secure, reliable, and persistent operation of the system.

broader functions and mandate. That understanding in turn provides a foundation on which to identify more precisely the kinds of documentation and information that are essential to support the verification of a record's authenticity over time and which, therefore, must be preserved and transferred along with the records when they become inactive and are transferred to the record preserver.

In identifying and positioning the elements included in the template for analysis, the Authenticity Task Force has struggled with the question of whether to treat the *medium*, i.e., the physical carrier on which a record is stored, as a part of the record itself or as part of its technological context. For diplomatists examining medieval documents, the medium is an essential component of a record because the examination of the physical carrier on which the document is inscribed is one of the most obvious proofs of its authenticity.[34] In the translation of traditional diplomatic concepts into modern paper-based record-keeping environments, the medium has continued to be treated as a part of the record itself, mainly because the medium and the message are inextricably linked. The question is whether, in an electronic record-keeping environment, the medium should continue to be treated as an essential part of the record itself given that: (1) the medium and the message are no longer inextricably linked; and (2) what is inscribed on or affixed to the medium is not a record as such (or words, or pictures), but a bitstream.

It is taken for granted that a record is a representation of a fact or act that is memorialized on a physical carrier, i.e., a medium, and preserved by a physical or juridical person in the course of carrying out its activities.[35] It follows that a record cannot exist before its elements have been inscribed on or affixed to a medium. Similarly, in an electronic environment, the bitstream, i.e., the source of the record, cannot endure for any length of time unless it is affixed to a medium.

Of course, with electronic records, storage of a bitstream on a hard, floppy, or optical disk, or on a magnetic tape, is necessary for the bitstream to endure but it is not sufficient to re-present the content and form of a record. Re-presentation of an electronic record's content and form also requires the capacity to process the record through software.[36] Moreover, although affixing a bitstream to a medium is considered an essential pre-condition to the

34 For example, a royal diploma of Childebert I (King of Francs, sixth century) that is written on parchment instead of papyrus is considered false. The medium also provides evidence of the manner in which medieval documents were prepared. The documents from the German chancery have many erasures and corrections in comparison to the documents of the papal chancery, indicating a lesser degree of care and accuracy in the preparation of the final documents.

35 Maria Guercio, "Principi, metodi e strumenti per la formazione, conservazione e utilizzo dei documentiarchivistici in ambiente digitale," *Archivi per la storia* XII, nos. 1–2 (1999), p. 26.

36 Thibodeau, "Certifying Authenticity of Electronic Records," p. 1.

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. It is assumed that it is neutral with respect to the record's authenticity at least from the perspective of the records creator and the records preserver.

For the moment, the problem of medium has been resolved by treating it in the template as both an element of the record itself and as part of its technological context (i.e., it is treated as something that is both inside and outside the record). Whether this solution – which acknowledges that medium is part of a record's technological context, yet continues to accord it a privileged role in determining the existence of an electronic record (a role not accorded to any other aspect of the technological context) – will prove to be supportable by the end of the project remains to be seen.

It is important to emphasize that the template for analysis is a generalized representation of an electronic record developed for the purpose of identifying all its known elements. It is not expected that any single electronic record will, or should, include all the elements identified in the template. The absence or presence of one or more of them in a specific instance will depend on the record's purpose. For example, although the attestation is probably the commonest means of validating a traditional record, it is by no means present in every record because, in many cases, the procedural controls exercised over the records' creation validates them, obviating the need for an explicit attestation.

The case studies currently underway will test the effectiveness of the template as a tool for identifying and analysing the elements of electronic records across a range of record-keeping environments and technologies. Some of the Canadian case studies that will be carried out in rounds three and four specifically target the field of digital music. Digital music records raise a host of authenticity-related issues, foregrounding subtle and complex questions which typically do not present themselves in traditional administrative record-keeping environments but which are essential considerations for records generated in music and other creative and performing arts.[37] The case studies of electronic systems containing digital music will assist the researchers in determining whether the archival diplomatic concept of a record, a concept based primarily on the nature of records created in the course of carrying out administrative and bureaucratic activities, is sufficiently robust to accommodate records created in the course of carrying out cultural and creative activities.

The purpose for identifying the elements of an electronic record and testing their validity through case studies is to define conceptual requirements for

37 The authenticity-related issues raised by digital music records are explored by Brent Lee in an article that appears in this issue of *Archivaria*.

verifying the authenticity of electronic records over the long term. On the basis of the work completed thus far, the Authenticity Task Force has prepared a discussion document, entitled "Draft Requirements for Authenticity," which outlines, in a preliminary way, the nature of those requirements.[38] Two levels of requirements have been identified, the first consisting of foundation or threshold requirements applicable to all electronic records and the second consisting of specific requirements associated with distinct types of electronic records.

It is generally acknowledged that verification of the authenticity of electronic records over the long term depends on the development and implementation of trust management systems.[39] The draft requirements for authenticity are built on this notion of trust management and are intended to establish a foundation on which to establish a presumption of authenticity for records that will be preserved over the long term. As stated in the "Draft Requirements," the requirements are based on the following premises:

> Establishing requirements for the authenticity of electronic records over the long term amounts to establishing requirements for the production of *authentic electronic copies* of *authentic electronic records*. The authenticity of *electronic records* must be verifiable from elements of the records (i.e., either on their face or linked to them) and contextual to the records (i.e., belonging to their documentary, administrative or technological context), while the authenticity of *electronic copies* of authentic electronic records is attested by the preserver, who has taken responsibility for the process of reproduction. ... In other words, *any electronic copy of an authentic electronic record is authentic if declared to be so by an officer entrusted with such function*, namely the official preserver.[40]

From these premises it follows that foundation requirements for the authenticity of the electronic records kept by the *creator* (either in live

38 [Authenticity Task Force], InterPARES Project, "Draft Requirements for Authenticity," version 1.1 (21 November 2000). The requirements will be issued in final form once all four rounds of case studies are completed.

39 Clifford Lynch explores the issues associated with the development and management of what he calls "identity and trust management systems" in the general context of digital objects and the specific context of digital signatures. See Lynch, "Authenticity and Integrity," pp. 32–50. See also Margaret Hedstrom, "Building Record-Keeping Systems: Archivists Are Not Alone on the Wild Frontier," *Archivaria* 44 (Fall 1997), pp. 44–71. In that article, Hedstrom examines trusted systems that are associated with electronic record-keeping. She characterises a trusted record-keeping system as "a type of trusted system where rules govern which documents are eligible for inclusion in the record-keeping system, who may place records in the system and retrieve records from it, what may be done to and with a record, how long records remain in the system, and how records are removed from it." Ibid., p. 57.

40 [Authenticity Task Force], "Draft Requirements for Authenticity," p. 3.

systems or outside the systems in which they were created) are essential to enable the *preserver* to verify such authenticity before the records selected for preservation are acquired and reproduced.[41]

It is accepted, both as a matter of law and of general principle, that records (at least those generated in a business context) relied upon by a creator for carrying out its business are presumptively authentic. The authentic records of the creator include (1) records that exist as created, i.e., they have not undergone processing that has altered their documentary form or any part of their technological context; and (2) any copies of those records that result from a migration process either to another electronic system or to another medium. Both types of records are considered authentic with respect to the creator, because the creator treats them as such by relying on them for action or reference in the usual and ordinary course of business.

The inference of trustworthiness that derives from the creator's need for accurate and authentic records does not, however, obviate the need for foundation requirements. Once records are no longer being used actively by the creator in the usual and ordinary course of business, the inference of trustworthiness is less supportable because the motivation to maintain accurate and authentic records ceases to be compelling. Moreover, while they should not be held to a higher standard of authenticity than that required for paper records, electronic records may carry fewer visible indicators of their identity and may be more vulnerable than paper records to undetectable modification. For these reasons, it is important to verify that the electronic records the creator relies on are clearly identifiable and of demonstrable integrity and that accidental corruption or purposeful tampering have not occurred after the records are no longer in active use by the creator.

The authenticity of electronic records is assessed in relation to their identity and integrity. The identity of a record refers to its provenance, author, addressee, writer, date, action or matter, and archival bond. The integrity of a record refers to its soundness (i.e., its condition is unimpaired) and completeness (i.e., it possesses all the necessary parts). Assessments of the integrity of a record (i.e., determining if it is sound and complete) are intimately connected to the question of what constitutes the essence of a record[42] and the status of copies relative to an original. As David Levy explains:

to be a copy ... is to stand in a certain relation to an original, that is, to its origin. To

41 Ibid. Although the creator and preserver of electronic records are treated as two conceptually distinct juridical persons, it is understood that the context in which they fulfil their separate roles will differ depending on whether the creator maintains its own historical records, as is usually the case with private corporate bodies or whether the creator's records are routinely transferred to a central archival depository, as is usually the case with public bodies.

42 The question of what constitutes the essence of a digital document is explored by Clifford

be a copy in this sense is to be faithful to the original. The definition of "faithful," however, depends on the circumstances in which the copy is being made and on the uses to which it will be put. The context of use, in other words, determines which properties of the original must be preserved in the copy. ... The point is, a document can be *identical* only with itself, if "identical" is taken to mean "the same in every respect." When we say that something is "the same," we generally mean one of two things. We either mean that it is "the very same" thing (as in "This is the same car I drove yesterday") or that it is "of the same type" as something else ("I read that same book last year"). It is this second notion of sameness – sameness of type, sameness in virtue of sharing certain properties – that is at issue in copying.[43]

In light of this reality, assessments of the integrity of a record cannot be made in any absolute sense but, rather, in relation to the purpose the record serves in the environment in which it has been created, maintained, and used. Thus, in the draft requirements, integrity refers to the fact that the elements conveying its identity are intelligible and the message that it is meant to communicate in order to achieve its purpose is unaltered. This implies that the precise number of bits in an electronic record need not be replicated in a copy, provided that the articulation of the content and its required formal elements remain the same.

The foundation requirements for authentic electronic records identify the kinds of procedural controls that will support the preserver's verification of authenticity. The case studies completed thus far suggest that, before the records selected for preservation are acquired, the preserver should verify whether the creator has, for example:

- implemented and monitored access privileges in the electronic system;
- designed a profile (or the functional equivalent of a profile) that is linked to each record as an annotation and that includes fields that allow the verification of the record's identity – including the name of the persons (author, writer, addressee, etc.), the action or matter, the chronological and archival dates and the expression of the archival bond (classification code,

---

Lynch in "Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information," *D-Lib Magazine* 5 (September 1999), at <http://www.dlib.org/dlib/september99/09lynch.html>. In that article, Lynch examines the problem of determining the effect of reformatting on the integrity of digital objects and the need for a more precise articulation of what constitutes the essence of a digital object in a given situation. He proposes canonicalization as a means of making precise what is important about a class of digital objects and for verifying that the integrity of these objects has been preserved in the reformatting process.

43 David Levy, "Where's Waldo? Reflections on Copies and Authenticity in a Digital Environment," *Authenticity in a Digital Environment*," p. 26.

dossier identifier, etc.) – and its integrity – including an indication of any additions, deletions, and migrations;
- established audit procedures by maintaining an audit trail of access to the records system to control the administration and use of access privileges; and maintaining an audit trail of every transmission (date, time, persons, action, or matter) within the record system;
- established procedures to prevent loss or corruption of records because of intentional or inadvertent unauthorized additions, deletions, or alterations; established procedures to prevent the loss of records due to technological obsolescence;
- established a procedure for taking records out of the live system for preservation purposes by: identifying the officers authorized to remove records from the system, determining storage medium and location for records removed from the system, and determining what has to be removed along with the records (e.g., indexes, data directories, data dictionaries, profiles, etc.);
- determined methods of transfer of inactive records to the entity competent for their preservation and the form in which the records will be transferred.[44]

Once the final two rounds of case studies are completed, these procedural controls will be reassessed to determine whether they are appropriate and relevant to electronic record-keeping environments that are different from the ones that have been examined thus far. The requirements will then be revised, qualified, and augmented in light of that determination.

While the verification of authentic electronic records is predicated on the existence of a trusted record-keeping system, the verification and attestation of the authenticity of copies of electronic records by the preserver (who assumes responsibility for the process of reproduction) is predicated on the more general notion of trust management and on the role of the preserver as a trusted custodian. In archival history, the role of trusted custodian dates back to Roman antiquity when citizens would deposit private records in the Tabularium for the express purpose of rendering them authentic. As a trusted custodian of records, ancient archival institutions sustained and lent credibility to contractual relationships between citizens. They also lent credibility to the implicit social contract between citizens and the state by preserving the records of the state's past actions on the basis of which the state could be held to account.

In the modern world, the role of trusted custodian has a cultural as well as

44 For a full list of these requirements see [Authenticity Task Force], "Draft Requirements for Authenticity," pp. 5–8.

a juridical dimension. The cultural dimension is highlighted by Charles Cullen in his discussion of the trust role played by librarians (and archivists) working in the realm of rare books and manuscripts. As Cullen elaborates,

> ... trusted librarians help authenticate their print holdings through recognized acquisition processes, accepted cataloging procedures, and careful stewardship of their collections, especially those in manuscript form. If a special collection librarian tells us, either directly or by means of a catalog card, that the book in hand is one of two extant copies of Ariosto's *Orlando Furioso* printed on vellum in Venice in 1542, and that it was prepared for the dauphin of France, the library's and the librarian's reputation go a long way toward instilling some degree of confidence that the document is indeed authentic. Moreover, all of this information can be checked. If another librarian delivers to a reader a box of letters cataloged as Ernest Hemingway's, authentication is assumed until internal or physical evidence suggests someone has made a mistake. Knowing that the materials – hard copy objects – have gone through a process of description and identification, if not authentication, conveys a sense of trust that they are authentic, at least until proved otherwise.[45]

The juridical and cultural dimensions of trusted custodianship are intimately connected and mutually reinforcing. Both dimensions are relevant and transportable to the electronic record-keeping environment. Clifford Lynch maintains that "... provenance and chain of custody in the digital world begin to reflect our evaluation of archives and custodians as implementers and operators of 'trusted systems' that enforce the integrity and provenance records of objects entrusted to them."[46]

For the preservers of electronic records to function effectively as trusted custodians, however, it is not sufficient that they simply declare that the records in their custody are presumptively authentic; they also provide grounds for such declaration. Verification of the authenticity of electronic copies of authentic electronic records depends on the accuracy of the documentation of the reproduction process, and on the preservation of the documentary and administrative context of the records themselves. Accordingly, the draft requirements stipulate that the preserver must take responsibility for:

- fully documenting the activity of reproduction (demonstrating the relationship between the records acquired from the creator and those reproduced, and the impact of the technology chosen for the preserved

45 Charles T. Cullen, "Authentication of Digital Objects: Lessons from a Historian's Research," *Authenticity in a Digital Environment*, pp. 3–4.

46 Clifford Lynch, "Authenticity and Integrity," p. 35.

copies on the form, content, accessibility, and use of the records), including the date of each reproduction and the name of the responsible person;
- ensuring that the identity of the record is clearly expressed by preserving where appropriate (e.g., in the record profile or its functional equivalent, on the face of the record, in a register) the elements that are necessary to determine it. The minimum elements necessary to express identity are the names of the persons involved in the creation of the record, the action or matter, the date of the record, and the expression of the archival bond;
- ensuring that the documentary and administrative (juridical, provenancial, and procedural) context of the records is accessible and clearly understandable both through their means of preservation and their archival description;
- maintaining and demonstrating unbroken custody of the record; and
- implementing and monitoring security and control procedures.[47]

As with the requirements for the verification of authentic electronic records, the requirements for the verification of authentic electronic reproductions will be revised, qualified, and augmented in light of the findings in the next two rounds of case studies.

Before turning to the specific requirements that are being developed for distinct types of electronic records, it is worth examining the broader epistemological framework in which the foundation requirements established for the production of authentic electronic copies of authentic electronic records are situated. The assessment of authenticity underpinning the requirements operates within a framework of probabilities, rather than certainties. Such assessment is similar in many respects to the common law's assessment of documentary evidence in general, especially as it concerns the relationship between admissibility and weight, the rules of relevancy, and the rules of auxiliary probative policy. In common law jurisdictions, the specific purpose of evidence law is to ensure the integrity of decisions reached in adjudication. The legal rules governing the admissibility of documentary evidence further that end by requiring that records meet a certain standard of trustworthiness before they are admitted as evidence in court.

Admissibility means that a particular fact is relevant, and that it has also met the requirements of specific auxiliary tests and extrinsic policies. As John Henry Wigmore makes clear in his *Treatise on the Anglo-American System of Evidence in Trials at Common Law*, it does not mean "that the particular fact has demonstrated or proved the proposition to be proved, but merely that it is received by the tribunal for the purpose of being weighed with other

47 For a full list of these requirements see [Authenticity Task Force], "Draft Requirements for Authenticity," pp. 11–12.

evidence."[48] The admissibility of evidence is determined by the judge, while the weight of evidence is determined by the trier of fact, usually the jury. The role of the preserver of records is analogous to that of a judge, while the role of users is analogous to that of the jury. This analogy suggests that, while the records preserver has a role to play in establishing threshold standards for the determination of authenticity, the users of records play an equally important role in assessing the degree of trustworthiness records ought to be accorded in specific circumstances. Moreover, the users' assessment is based on a wider range of considerations than are typically taken into account by the preserver. As Clifford Lynch observes,

> At some level, authenticity and integrity are mechanical characteristics of digital objects; they do not speak to deeper questions of whether the contents of a digital document are accurate or truthful when judged objectively. An authentic document may faithfully transmit complete falsehoods. There is a hierarchy of assessment in operation: forensics, diplomatics, intellectual analyses of consistency and plausibility, and evaluations of truthfulness and accuracy. Our concern here is with the lower levels of this hierarchy (i.e., forensics and diplomatics as they are reconceived in the digital environment) but we must recognise that conclusive evaluations at the higher levels may also provide evidence that is relevant to lower level assessment.[49]

Foundation requirements, in other words, are not the final word on authenticity-related questions nor are they immune to challenge. They simply establish grounds for a presumption of authenticity which means that, until proof to the contrary is shown, records that meet the requirements are considered authentic.

The rules of admissibility governing relevancy deal with the probative value of specific facts. The rules of auxiliary probative policy aim at increasing or safeguarding their probative value. The rules of relevancy derive from principles of logical relevancy, which are expressed in terms of the relationship between evidence and probability. As legal evidence scholar Peter Tillers explains, "[k]nowledge of facts is always a matter of probabilities. We may acquire knowledge of matters of fact by drawing inferences from evidence, but these inferences can only alter the probability that some fact does or does not exist and can never establish with certainty that some fact does or does not exist."[50] Inferences, in turn, rest on generalizations based on common sense experience and logic:

48 *Wigmore on Evidence* vol. 1, para. 12.
49 Lynch, "Authenticity and Integrity in the Digital Environment," pp. 35–36.
50 *Wigmore on Evidence* vol. 1A, para. 37.4.

> We draw an inference when the existence of one fact, the factum probans, alters our estimate of the existence of another fact, the factum probandum, but we do not draw that inference because of any intrinsic relationship between the factum probans [the existent fact, i.e., the evidence] and the factum probandum [the hypothetical fact, i.e., the proposition]; we draw that inference because we hold some principle that leads us to believe that the existence of the factum probans makes the existence of the factum probandum more or less probable. These connective principles are called "generalizations" or "evidential hypotheses," and they are furnished by experience or logic. They take the form of relative frequency statements that assert that when events of type A occur, events of type B occur with a certain frequency (e.g., "very often," or "almost always").[51]

Inferences from evidence usually involve a series or chain of inferences and a chain of inferences is only as strong as its weakest link. "The greater the number of links in the chain – the greater the number of intermediate inferences – the weaker the final inference produced by the chain of inferences."[52]

Similarly, the strength of the preserver's declaration of authenticity is only as strong as the evidence on which that declaration rests. An archives is not a rehabilitation centre for records whose identity and integrity have been lost or compromised while they were in the hands of the creator, and the preserver cannot declare records to be authentic in the absence of evidence to support such a claim. In such cases, the best the preserver can offer is a commitment to maintain the records as authentic as they were when they were transferred to archival custody and to try to avoid further slippage.

The rules of auxiliary probative policy operate within this larger framework of logical relevancy and are "designed to strengthen here and there the evidential fabric and *to secure it against dangers and weaknesses pointed out by experience*."[53] The best evidence rule, the business records exception to the hearsay rule, and the rules governing authentication of documents are all rules of auxiliary probative policy that are used to assess the trustworthiness of documentary evidence specifically. The foundation requirements for authentic copies of authentic electronic records rely on common sense inferences and generalizations about what constitutes a reliable and authentic record that are similar to those that underlie the legal rules governing probative policy. And, like those legal rules, the requirements, and the inferences and generalizations on which they rest, must be tested and regularly reassessed to determine their continuing validity.

51 Ibid.
52 Ibid.
53 *Wigmore on Evidence* (Chadbourn rev., 1972), vol. 4, para. 1171.

In addition to identifying foundation requirements, the Authenticity Task Force is also responsible for identifying specific requirements associated with distinct types of electronic records. An electronic records typology is being developed as an aid to the identification of these requirements. The word typology comes from the Greek word *typos*, which means an impression or a pattern. A typology is a system of groupings, usually called types, which are classes of things, persons, or events that have specific common attributes. The primary purpose of a typology is "to produce ordered and reproducible sets that can support the rapid identification of members of groups of sets in general and members of individual sets or subsets in particular."[54] As Seamus Ross points 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.[55]

There are two approaches to the design and implementation of a typology, the first is top-down and the second is bottom-up. As Ross explains,

> In the former approach a researcher begins within the premise that a "group of entities" ... forms a bounded set. Then the researcher attempts to select and define characteristics shared by the material and to determine whether objects/entities proposed as members of the group have the required attributes. In this approach the set becomes equivalent with the type. In the second approach the investigator starts with the objects and proceeds to describe the component elements. The elements are then grouped into attributes and the attributes subsequently grouped into restricted sets. These are shared component types that carry meaning.[56]

The purpose for developing a typology of electronic records is to define the authenticity requirements specific to different types of electronic records. The criterion for developing the typology is the significance of the extrinsic and intrinsic elements of the records and their annotations for carrying out or attesting to the action or matter in which the records participate. Following the completion of the first two rounds of case studies, the task force has

54 Seamus Ross, "Dress-pins from Anglo-Saxon England: their production and typo-chronological development," (D.Phil. dissertation, University of Oxford, 1992), p. 68. For the work accomplished to date in establishing the conceptual and methodological basis for typological analysis the task force is indebted to Ross's exploration of typological analysis as it is used in archaeological research and to Ian McAndrew's summary of Ross's work, "Typologies and Typological Analysis: Definitions and Characteristics," unpublished report to the Authenticity Task Force, October 2000. The discussion of typologies and typological analysis that follows is based on Chapter 3 of Ross's dissertation, "Re-thinking Typology: Designing Material Culture Models."

55 Ibid., p. 9.

56 Ibid., p. 86.

adopted a top-down approach for the initial basic typology (i.e., for the highest level of categorization). Once the next two rounds of case studies have been completed, it is anticipated that this top-down approach will be supplemented by a bottom-up approach for the development of sub-types and, possibly, the creation of additional primary types. This is in keeping with the iterative nature of typological analysis. As the work proceeds, it is expected that concepts may be redefined, premises re-examined, and initial types reconsidered in light of the new findings.[57]

The initial basic typology reflects the four categories of records identified by contemporary archival diplomatics.[58] The categories are based on the relationship between a record and the action in which it participates. The choice of this categorization is based on the premise that groups of records sharing the same function with respect to an action or matter form a bounded set. The categories are *dispositive* records (records whose written form is required by the juridical system as the essence and substance of an action), *probative* records (records whose written form is required by the juridical system as proof that an action has taken place prior to its documentation), *supporting* records (records whose written form is discretionary; they are created to provide support for, and are procedurally linked to, an action), and *narrative* records (records whose written form is also discretionary; they do not participate procedurally in the action but are created as part of the process of setting oneself to work).

An extended definition of these four categories of records is currently being tested. According to this definition, the terms dispositive, probative, supporting, and narrative refer to the smallest indivisible aggregation of records (e.g., the file unit) in each system rather than to individual records. Dispositive, probative, supporting, or narrative aggregations of records may contain one or more types of records. This definitional extension of the record categories implies an extension of the authenticity requirements because the requirements for a given category of records will apply to all the records within the aggregation, regardless of the different types of individual records contained within it.

On the basis of this preliminary categorization, the researchers have drawn a number of inferences about the specific requirements for authentic electronic records and authentic reproductions of authentic electronic records. For example, for dispositive and probative aggregations of records, i.e., records whose written form is required and therefore mostly prescribed as to elements of extrinsic and intrinsic form and to annotations, the specific requirements for verifying their authenticity before they are acquired by the preserver might

57 Ibid., pp. 72, 88.

58 The discussion of the preliminary categorization of records and its implications that follows is based on "Draft Requirements for Authenticity," pp. 8–12.

be (1) the presence, on the face of the record, of all the elements prescribed by the juridical system; and (2) the inclusion, on the record profile, of all the data related to responsibility for, and any changes to, the record. As it concerns the preservation of authentic reproductions of authentic electronic records, this categorization implies that dispositive and probative aggregations of records should be preserved as authentic copies in the form of an original. A copy in the form of an original is a record that, on its face, looks in all essential ways like the original, i.e., it presents the same extrinsic and intrinsic elements of form, identical content, and has all the annotations that are linked to the original. Such copy is considered to be as complete and effective as the original record. Reproductions of supporting and narrative aggregations of records, on the other hand, only require their reproduction to be as accurate as needed for the purposes for which they were used. In certain cases, a simple copy, i.e., a copy that only reproduces the content of the original, would be sufficient. In other cases, formal elements would need to be carried forward for the record to be either intelligible or capable of being used as it was when it was current. In such cases, an imitative copy, i.e., a copy that reproduces, completely or partially, the content and form of the original, would be necessary. The validity of these preliminary categorizations, and the inferences drawn from them, will be tested in the next two rounds of case studies.

This article has chronicled the efforts of the Authenticity Task Force to identify and elaborate the grounds on which we might base our trust in the authenticity of electronic records that will be preserved over the long term. The development of conceptual requirements for authenticity is an essential first step towards identifying the kinds of descriptive metadata and procedural documentation that should be carried forward with electronic records to help preserve them as authentic memory and evidence for future generations. The authenticity requirements also provide the framework in which research in the other domains of inquiry in the InterPARES project is currently being carried out. The Appraisal Task Force (responsible for domain two) is developing a set of appraisal criteria and specific appraisal procedures for electronic records that are consistent with the requirements for authenticity. The Preservation Task Force (responsible for domain three), for its part, is formulating procedures and rules for implementing the requirements. But that, as they say, is another story, and one best told by the Appraisal and Preservation Task Forces.[59]

59 As the findings of the Appraisal and Preservation Task Forces develop, they will be posted on the InterPARES Web site.