From Digital Diplomatics to Digital Records Forensics*

LUCIANA DURANTI

RÉSUMÉ Il y a quinze ans, Elizabeth Diamond décrivait l'archiviste comme un scientifique médicolégal. Depuis quelques années, plusieurs auteurs dans le domaine de l'archivistique ont qualifié les professionnels responsables de la préservation des documents numériques de conservateurs de confiance (« *trusted keepers* »), ou de gardiens (« *custodians* »). Sans doute, dans l'environnement numérique, on fait de plus en plus appel aux professionnels de l'information pour évaluer et préserver l'authenticité des documents dont ils sont responsables, et pour agir en tant que tierce parties neutres. Mais sont-ils qualifiés pour remplir ce rôle? Cet article tente d'identifier les connaissances que doit avoir le professionnel d'information de confiance pour être capable d'évaluer la véracité (« *trustworthiness* ») des documents numériques et pour assurer que leur authenticité puisse être démontrée, au besoin, à n'importe quel point dans leur cycle de vie. Pour ce faire, l'article présente des concepts développés par le projet InterPARES dans le domaine de la diplomatique des documents numériques; il compare ceux-ci aux concepts pertinents dérivés d'une discipline relativement nouvelle, le numérique médicolégal (« *digital forensics* »); il discute des méthodologies dont se servent les deux disciplines; et il propose des domaines qui pourraient être explorés conjointement par les experts en diplomatique et en numérique médicolégal afin de développer un corpus de savoir intégré que l'on pourrait nommer la science médicolégale des documents numériques (« *Digital Records Forensics* »).

ABSTRACT Fifteen years ago, Elizabeth Diamond described the archivist as a forensic scientist. In the past few years, several archival writers have referred to professionals responsible for keeping digital records as trusted keepers or custodians. Undoubtedly, in the digital environment, record professionals are increasingly called to assess and preserve the authenticity of the records they are responsible for, and to act as neutral third parties. But, are they qualified to fulfill this role? This article aims to begin identifying the body of knowledge that a trusted record professional needs in order to assess the trustworthiness of digital records and ensure that their continuing

* I dedicate this article to the memory of Elizabeth Diamond who encouraged and inspired me when I was learning to be a Canadian archivist. The fact that it took me fifteen years to truly understand her call and bring it to fruition shows her foresight and imagination.

ARCHIVARIA 68 (Fall 2009): 39–66

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

authenticity can be demonstrated, if required, at any point during their life cycle. To do so, it presents some of the concepts developed by the InterPARES Project in the area of diplomatics of digital records; compares them with the relevant concepts of a relatively new discipline called digital forensics; discusses the methodologies used by the two disciplines; and proposes areas that can be jointly investigated by diplomatics and forensics experts to develop an integrated body of knowledge that might be called Digital Records Forensics.

Introduction

In 1994, Elizabeth Diamond wrote an article that described the archivist as a "forensic scientist"[1] and went on to explain that "[i]t is his or her major professional function to clarify the meaning" of the records, "what Sir Hilary Jenkinson called the 'material evidences'[2] of the historical case." The archivist, just like a forensic scientist, asks questions about the characteristics of the records and their use, questions that are increasingly important in the digital environment. "The archivist – Diamond continues – must 'translate' the records and be able to testify that they have not been tampered with or falsified. This expert testimony is essential; without it, the electronic record cannot be used as evidence. It is clearly a modern example of Jenkinson's moral defense of archives." "Archivists, like forensic scientists, become expert witnesses, testifying to the nature of the documents."[3]

Fifteen years later, Diamond's words could not sound truer. Indeed, the enduring trustworthiness of our documentary heritage is becoming a central responsibility of its designated custodian, of this neutral third party whose key function Diamond had stressed in her article: "The services of the forensic scientist are available impartially to all sides in the case. A forensic scientist may have an opinion about the rights or wrongs in any particular argument, but does not express that opinion in an official capacity; she or he must avoid advocacy."[4] The question arises, however, of what is required for a professional responsible for keeping records to qualify as an impartial third party.

The concept of "trusted third-party recordkeeper" was developed in the context of electronic contracting, and refers to a physical or juridical person[5] who is entrusted with the maintenance of the records of electronic

1 Elizabeth Diamond, "The Archivist as a Forensic Scientist. Seeing Ourselves in a Different Way," *Archivaria* 38 (Fall 1994), pp. 139–54.

2 Sir Hilary Jenkinson, "The English Archivist: A New Profession," in *The Selected Writings of Sir Hilary Jenkinson* (Gloucester, 1980), pp. 246–47.

3 Diamond, p. 142.

4 Ibid., p. 140.

5 A juridical person is a collection or a succession of physical persons. Examples are organizations of any kind, committees, and positions.

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

data interchange (EDI) partners. To qualify as such, a trusted recordkeeper must demonstrate that it has no reason to alter the records and no interest in allowing others to do so, and must have the knowledge necessary to implement procedures that ensure the integrity and accuracy of the records.[6] The InterPARES Project – an international, multidisciplinary research collaboration aimed at developing the theory and methods necessary for the permanent preservation of authentic electronic records[7] – similarly defined a trusted custodian as a neutral third party who must demonstrate that it has no reason to alter or to allow others to alter the records in its care, and that it has the knowledge required for attesting to, and ensuring the continuing authenticity of, the records.[8] Ensuring authenticity is the key to the identity of the trusted recordkeeper or custodian, or, more generally, of the records professionals responsible for assessing and guaranteeing over time the trustworthiness of our documentary heritage. Diamond's article argues that any records manager/archivist who has received a dedicated graduate archival education should be endowed with the intellectual armour required to fulfill the role of trusted keeper or custodian of traditional records. The question is whether this is also true with respect to digital records, which present several challenges related to both their identity and their integrity.

The purpose of this article is to begin identifying the body of knowledge that a trusted record professional needs in order to assess the trustworthiness of digital records and ensure that their continuing authenticity can be demonstrated, if required, at any point during their life cycle. To do so, it will present some of the concepts developed by the InterPARES Project in the area of diplomatics of digital records; compare them with the relevant concepts of a relatively new discipline called digital forensics; discuss the methodologies used by the two disciplines; and propose areas that can be jointly investigated by diplomatics and forensics experts to develop an integrated body of knowledge that might be called Digital Records Forensics.

6 Bernard D. Reams Jr., L.J. Kutten, and Allen E. Strehler, *Electronic Contracting Law: EDI and Business Transactions, 1996–97 Edition* (New York, 1997), p. 37.

7 For a succinct description of the InterPARES Project see the home page of its website at http://www.interpares.org.

8 "Authenticity Task Force Report," in Luciana Duranti, ed., *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (Rome, 2002), p. 21, http://www.interpares.org/book/interpares_book_d_part1.pdf (accessed on 19 August 2009). This definition of a trusted custodian and the need for such a role were confirmed in the course of the second phase of the InterPARES Project and identified as a required component of any record policy. See Appendix 19: "A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records," in Luciana Duranti and Randy Preston, eds., *InterPARES 2: Interactive, Dynamic and Experiential Records* (Padova, 2008), pp. 10, 14, http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_19.pdf (accessed on 19 August 2009).

Diplomatics of Digital Records and Digital Forensics

In order to enable records professionals to understand digital records and be responsible for their trustworthiness over time, the InterPARES Project took traditional diplomatic and archival knowledge, applied it to all manner of entities existing in a variety of digital environments, and developed from it a new body of knowledge aimed at serving current and future needs.[9] This new body of knowledge could be named "diplomatics of digital records," and considered a product of special diplomatics.[10] It may not be sufficient, however, for dealing with the challenges presented by increasingly complex digital environments, which might require that concepts, principles, and methods developed in the context of other disciplines be brought to bear on digital diplomatics. One such discipline is "digital forensics science," defined by Ken Zatyko as "the application of computer science and investigative procedures for a legal purpose involving the analysis of digital evidence after proper search authority, chain of custody, validation with mathematics, use of validated tools, repeatability, reporting, and possible expert presentation."[11] More specifically, the Digital Forensics Research Workshop, in 2001, had defined "digital forensics" as "the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations."[12]

Digital forensics has developed methodologies and a very large body of practices carried out in an investigative context[13]; the theory supporting digital forensics processes, however, has not been fully articulated yet, although some

9 The methodologies by means of which new knowledge was developed in the course of the InterPARES Project are described in detail in the two InterPARES books in the context of each chapter discussing a specific part of the research and its findings, products, and outcomes. See Duranti, *The Long-term Preservation of Authentic Electronic Records*; and Duranti and Preston.

10 Special diplomatics is a branch of diplomatics, a discipline in which "the theoretical principles formulated and analyzed by diplomatics individualize, develop and clarify themselves being applied to single, concrete, real, existent and easily exemplifiable documents, rather than to an abstract and atypical general documentation." Luciana Duranti, "Diplomatics: New Uses for an Old Science. Part I" *Archivaria* 28 (Summer 1989), p. 9.

11 Ken Zatyko, "Commentary: Defining Digital Forensics," *Forensic Magazine* (Feb/March 2007), pp. 1–5.

12 *Digital Forensics Research Workshop*, 2001, p. 15, http://www.dfrws.org/2001/dfrws-rm-final.pdf (accessed on 25 August 2009).

13 Sarah Mocas, "Building Theoretical Underpinnings for Digital Forensics Research," *Digital Investigation*, vol. 1, no. 1 (2004), p. 62, www.elsevier.com/locate/diin (accessed on 25 August 2009).

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

very clear concepts and principles have been enunciated and agreed upon. Digital forensics experts are still mostly practitioners and the discipline has only recently entered academia; thus, it lacks the kind of *fora* offered to theory development by a multiplicity of scholarly journals and a well-established community of academics writing for them. Most literature on the subject is confined to magazines, newsletters, and conference proceedings, while substantive articles in terms of theoretical and methodological content can only be found in legal journals or in the one international journal for the field, the *International Journal of Digital Evidence*.[14] In 2006, the Association of Digital Forensics, Security and Law began publishing a double-blind refereed journal entitled *The Journal of Digital Forensics, Security and Law (JDFSL)* with the intent of providing a forum for high quality research, communication, and debate on the subject of digital forensics and related fields. The mission of the *JDFSL* is "to significantly expand the domain of digital forensics research to a wide and eclectic audience of academics, consultants, and executives who are involved in the curriculum, research, and use of digital forensics. *JDFSL* publishes original research and comments about digital forensics and its relationship to security and law. Contributions are particularly welcome which analyze the results of interdisciplinary research and relate to the intersection of theory, method, and empirical findings."[15]

Presently, the goals of digital forensics professionals are very different from those of a trusted recordkeeper or custodian, but are similar to those that gave origin to diplomatics in the seventeenth century and resulted in its study in the European faculties of law in the eighteenth century. Diplomatists were the forensic scientists of their day; they were called upon to authenticate records in a court of law when the rights they attested to were challenged and their trustworthiness as records questioned.[16] Comparing the concepts, principles, and methods of digital forensics and diplomatics of digital records might consequently strengthen both disciplines and lead to the development of a new body of knowledge that might be called Digital Records Forensics. But, before elaborating on the idea of a new science for fundamentally old uses,[17]

14 Available at http://www.utica.edu/academic/institutes/ecii/ijde/about.cfm (accessed on 25 August 2009).

15 This text is taken from the mission statement of the *JDFSL*, http://www.jdfsl.org/mission.htm (accessed on 25 August 2009).

16 Duranti, "Diplomatics. Part I," pp. 13–14.

17 When I first introduced general diplomatics to North American audiences, I entitled my writings "Diplomatics: New Uses for an Old Science." I find it ironic now to be developing from that initial effort a new science for the same purpose that general diplomatics was created to fulfill, establishing the trustworthiness of records, but in a context that is both technologically and juridically different. The six articles by that title, first published in *Archivaria* 28–32 (1989–1992), were later collected in Luciana Duranti, *Diplomatics: New Uses for An Old Science* (Chicago, 1998).

it is first necessary to present the central concepts of diplomatics of digital records – hereinafter called "digital diplomatics" for reason of brevity – and contrast them with the concepts used by digital forensics. It is simply natural to begin with the concept of digital record.

The Concept of Digital Record

Digital diplomatics defines a digital record as a digital component, or group of digital components, that is saved, and treated and managed as a record, or, more specifically, "a record whose content and form are encoded using discrete numeric values (such as the binary values 0 and 1) rather than a continuous spectrum of values (such as those generated by an analogue system)." A digital record is distinguished from an "analogue record" and an "electronic record." InterPARES considers analogue the representation of an object or physical process through the use of continuously variable electronic signals or mechanical patterns. In contrast to a digitally encoded representation of an object or physical process, an analogue representation resembles the original. InterPARES defines an electronic record as any analogue or digital record that is carried by an electrical conductor and requires the use of electronic equipment to be intelligible by a person (e.g., a fax). InterPARES defines a record, using the traditional archival concept, as "a document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference."[18]

Digital forensics – when endeavouring to provide a general definition for records – adopts the definition provided by law or statute in each of the countries in which the discipline is applied, and rightly so, because its purpose is of a legal nature, different from the purpose of digital diplomatics, which is of an archival nature: definitions must be consistent with the purpose for which they are used. Thus, in Canada, the relevant definition of record for digital forensics experts is: "the whole or any part of any book, document, paper, card, tape or other thing on or in which information is written, recorded, stored or reproduced, and ... any copy or transcript admitted in evidence."[19] Another definition is:

> … any correspondence, memorandum, book, plan, map, drawing, diagram, pictorial or graphic work, photograph, film, microform, sound recording, videotape, machine readable record, and any other documentary material regardless of physical form or characteristics, and any copy thereof. When used in reference to a person, [the term] means all recorded information, regardless of physical form or characteristics, that: (i)

18 InterPARES 2, *Terminology Database*, http://www.interpares.org/ip2/ip2_terminology_db.cfm (accessed on 25 August 2009).

19 *Canada Evidence Act*, R.S.C. 1985, c. C-5, s. 30(12).

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

relates to the person; (ii) is recorded in connection with the provision of an approved service, or a service purchased by an approved agency, to the person or a member of the person's family; and (iii) is under the control of a service provider.[20]

Clearly, these legal definitions are outdated, and laws and statutes will soon have to substitute them with more comprehensive ones based on the concept of record rather than on a list of media, forms, characteristics, or degree of perfection (e.g., copy), and capable of dealing with the growing number of new record types generated by digital technology.

However, if general definitions serve the very useful purpose of circumscribing the object examined, they are not very helpful in determining the identity and integrity of each instance of such object. Although digital forensics experts do not define "digital" records, they demonstrate an awareness of the challenges such records might present in the absence of a more detailed description of their necessary attributes, components, or even behaviour. This is particularly evident when they discuss whether, for example, "digital evidence falls under the Daubert guidelines as scientific evidence or the Federal Rules of Evidence as non-scientific technical testimony,"[21] like business records do. The significance of this question will be shown later on, after the following analysis of the concept of digital record according to digital diplomatics.

A digital record must have 1) an identifiable context; 2) an originator,[22] an author,[23] a writer,[24] an addressee,[25] and a creator;[26] 3) an action, in which the record participates or which the record supports either procedurally or as part of the decision-making process; 4) explicit linkages to other records within or outside the digital system, through a classification code or other unique identifier; 5) a fixed form; and 6) a stable content.[27] Most of these requirements

20 Daphne A. Dukelow, ed., *The Dictionary of Canadian Law*, 3rd ed. (Toronto, 2004).

21 Brian Carrier, "Open Source Digital Forensics Tool. The Legal Argument," p. 7, http://www.digital-evidence.org/papers/opensrc_legal.pdf (accessed on 25 August 2009). The Daubert Test is used in the United States to determine the admissibility of scientific evidence. See *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 572 (1993).

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

23 The physical or juridical person(s) having the authority and capacity to issue the record, or in whose name or by whose command the record has been issued.

24 The physical or juridical person(s) 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.

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

26 The physical or juridical person in whose *fonds* the record exists.

27 These necessary characteristics are discussed in detail in the article by Heather MacNeil, "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-term Preservation of Electronic Records," *Archivaria* 50 (Fall 2000), pp. 52–78. It must be

are self-explanatory; the concepts of fixed form and stable content, however, require elaboration, as these two characteristics of a digital record are the most problematic not only for the purposes of trusted recordkeepers, but also for those of digital forensics experts.[28]

A digital record has a fixed form if its binary content is stored so that the message it conveys can be rendered with the same documentary presentation it had on the screen when first saved, even if its digital presentation has been changed, for example, from Word to .pdf.[29] A digital record has a fixed form as well if the same content can be presented on the screen in several different ways but in a limited series of pre-determined possibilities; in such a case we would have different documentary presentations of the same record (e.g., statistical data viewed as a pie chart, a bar chart, or a table). This situation raises the issue of the difference between a stored record and a manifested record.

A "stored record" is constituted of the linked digital component(s)[30] that are used in re-producing the record, which comprise the data to be processed in order to manifest the record (i.e., content data and form data) and the rules for processing the data, including those enabling variations (i.e., composition data). A "manifested record" is the visualization or materialization of the record in a form suitable for presentation to a person or system. Sometimes, a manifested record does not have a corresponding stored record, but is re-created from fixed content data when a user's action associates these data with specific form and composition data (e.g., a record produced from a relational database). If the same user's action always results in the same documentary presentation with the same content, the manifested entity is considered to have fixed form and stable content, even when it does not have a corresponding stored record, and, if all other requirements for the existence of a record are present, it is a record. In contrast, when one stored record may be manifested in several documentary presentations, the creator has to determine

emphasized that traditional records also present these characteristics, but they are often implicit. For example, the relationship between a record and the other records of the same creator or resulting from the same business activity, was often revealed by the physical location of the record – which in the digital environment is meaningless – rather than by a classification code.

28 Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES," *Archival Science*, vol. 6, no. 1 (2006), p. 16, http://dx.doi.org/10.1007/s10502-006-9021-7 (accessed on 25 August 2009).

29 "Documentary presentation" is a more generic term than "documentary form" and encompasses all possible documentary forms. Similarly, "digital presentation" is a more generic term than "format" and encompasses all possible formats, in addition to digital components and their interrelationships.

30 "Digital components" are digital entities that either contain one or more records, or are contained in the record and require a specific preservation measure.

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

whether the official record is the stored one, or one or more of the manifested ones by assigning to the chosen entity a classification code and/or a retention period. There might be situations in which a stored record is never manifested, as is the case with software patches that enable the playing of electro-acoustic music, or with interacting business applications, workflow generated and used to carry out experiments, analyses of observational data carried out by interpreting software, etc. Also in this case, the creator determines which entities should be retained with other records of the same activity, manifested or not. Clearly, these decisions are based on the functions and activities in which the records participate, both as aggregates and as individual entities. This point is a key one in relation to the contributions that digital diplomatics and digital forensics can offer to each other, and will be discussed later on. Now it is important to clarify the concept of stable content.

A digital entity has stable content and can be considered a record, if all other conditions are satisfied, if the data and the message in it are unchanged and unchangeable, meaning that data cannot be overwritten, altered, deleted, or added to. However, there are cases in which entities that demonstrate "bounded variability" can be said to have stable content. A digital entity has bounded variability when changes to its form are limited and controlled by fixed rules, so that the same query or interaction always generates the same result, and when the user can have different views of different subsets of content, due to the intention of the author or to the character of the operating systems or applications. While the first definition of stable content applies to static digital entities, the second is significant when the entities we are looking at are interactive.

A "static digital entity" is one that does not provide possibilities for changing its manifest content or form beyond opening, closing, and navigating; for example, emails, reports, sound recordings, motion videos, and snapshots of Web pages. These entities, if all other requirements are satisfied, are records, because they have fixed form and stable content. By contrast, an "interactive digital entity," presents variable content, form, or both, and the rules governing the content and form of presentation may be either fixed or variable. Interactive entities may or may not be records, depending on whether they are non-dynamic or dynamic. "Non-dynamic entities" are those for which the rules governing the presentation of content and form do not vary, and the content presented each time is selected from a fixed store of data. Examples are interactive Web pages, on-line catalogues, and entities enabling performances: if the other conditions exist, they are records. "Dynamic entities" are those for which the rules governing the presentation of content and form may vary: these entities may be components of information systems or "potential records," in that they can become records if the digital system in which they exist, given the purpose that it fulfills, is supposed to contain records and is therefore redesigned in such a way that it will produce and manage records,

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

or if the entities that should exist as records are moved to another system that only maintains digital records (i.e., static or non-dynamic entities).

Digital forensics has a similar understanding of static and dynamic entities. Mocas writes:

> For example, imaging a single hard drive and then performing a search on that image provides a static technical environment. In contrast, a dynamic technical environment is one in which one or more of the components from which data are retrieved have a potential for modification, independent of any system changes that might be introduced during the investigative process. In other words, "live" systems and systems connected to the Internet qualify as dynamic.[31]

Digital diplomatics is, however, more specific about dynamic entities and allows for an easier identification by describing their behaviour. Examples of dynamic entities are: entities whose variation is due to data that change frequently (e.g., the design permits updating, replacement, or alterations; it allows data collection from users or about user interactions or actions; or it uses these data to determine subsequent presentations); entities whose variation is due to data continually received from external sources and not stored within the system; entities produced in dynamic computing applications that select different sets of rules to produce documents, depending on user input, sources of content data, and characteristics of content (e.g., weather sites); entities produced by evolutionary computing where the software generating them can change autonomously (e.g., scheduling and modelling of financial markets; edutainment sites, etc.).[32]

In order to establish whether entities of the kind described above are records,[33] it is essential to establish in which way they participate in activities, if at all, in the context of the functions of their creator. There are different ways in which a record may participate in an action, and, depending on its function with respect to the action in which it takes part, a record may acquire a specific qualifier. Thus, if a record is meant to provide evidence of an act that came into existence and was complete before being manifested in writing, it is qualified as a *probative* record, while if it is meant to put the act into being and therefore constitutes the essence and substance of the act, it is qualified as a

31 Mocas, pp. 62–63.

32 This articulation of the concepts of fixed form and stable content, static and interactive records, and non-dynamic and dynamic records was developed by this author with Ken Thibodeau in the context of an article summarizing the findings of InterPARES 2 with respect to the concept of record. See Duranti and Thibodeau, pp. 13–68.

33 In the case of a positive decision, a trusted recordkeeper should indicate the need for a redesign of the system in which the entity exists, while a digital forensic expert would consider the digital entity as one that cannot be subject to the business records exception to the hearsay rule for admissibility.

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

dispositive record. Examples of probative records are certificates, registrations, transcripts, and receipts. Examples of dispositive records are contracts, grants, applications, and money orders. These types of records all have in common the fact that their existence and written form are required by the juridical-administrative system within which they are created, and therefore they are all legal records. Traditionally, these records have a very formal documentary presentation, but in the digital environment they are increasingly becoming informal, to the point that a simple email can have the effects of a contract.

Records whose existence is not required by the juridical-administrative system for carrying out actions and the written form of which is discretionary are considered non-legal records. They have been distinguished in two categories: *supporting* records, whose function is to inform the activity in which they take part; and *narrative* records, whose function is one of free-form communication of information. While both categories of records participate in some kind of act, neither is able to provide evidence of such act by itself or to carry it out. Examples of supporting records are teaching notes and maps; examples of narrative records are notes, unsolicited reports, and informal accounts of events. In the digital environment, we find two additional categories of records, *instructive* records and *enabling* records. The former indicate the way in which data, documents, or records are to be presented (e.g., forms with embedded instructions for filling them out and formatting), and the latter enable a presentation, such as the performance of artworks (e.g., software patches), execution of business transactions (e.g., interacting business applications), conduct of experiments (e.g., a workflow generated and used to carry out the experiment of which it is instrument, by-product, and residue), or analysis of observational data (e.g., interpreting software). The salient characteristic of instructive records is that the record as it is stored differs from the record as it is manifested on the computer screen, while the salient characteristic of enabling records is that they usually do not have a corresponding manifested record.[34]

Digital forensic experts have been concerned with the different types of records that can be found in digital systems, although only in relation to their ability to be admissible evidence. In fact, the courts have generally admitted digital records under the business records exception to the hearsay rule. This rule, which is shared by common law countries, considers all written evidence hearsay because it contains statements made by humans who cannot be cross-examined on the stated matter. However, written documents which 1) are produced in the usual and ordinary course of business, 2) at or near the time of the facts or acts of which they are offered as evidence, 3) by someone who is under a duty to create them for the purposes of the business and is familiar with the procedure for doing so, are considered "business records" and admis-

34 Duranti and Thibodeau, pp. 49–52.

sible under an exception to the hearsay rule. The term "business" is used in a general sense and includes, in addition to business proper, the work conducted by any institution, association, profession, occupation, and calling of every kind, whether or not carried out for profit.[35] The identification of what constitutes a business record according to the criteria listed above has never been problematic with traditional records, but, with computer records, experts are beginning to make distinctions.

Digital forensics experts distinguish "computer-stored" records from "computer generated" records. The former category, *computer-stored* records, includes the writings of physical persons that happen to be in electronic form. Examples are email messages, word processing documents, and Internet chat room messages. Like any traditional document containing human statements, computer-stored records are regarded by the law of evidence as hearsay and must fall under the business records exception to the hearsay rule to be admissible in court. The latter category, *computer-generated* records, contains the output of computer programs, produced without direct human interference. These records can be computer-recorded events, like transactions recorded by ATM machines; data sets produced by a system that performs analysis or calculations of input provided by humans; simulations modelling behaviours or predicting outcomes of events; decision trees, which lead a user to a destination through a path that depends on questions provided along the way; and, of course, records that digital diplomatics includes in the "stored only" and "enabling" records category.[36] Unlike computer-stored records, computer-generated records are regarded as containing no human statements, as they are produced by a computer program designed to process input following a defined algorithm. The idea that computer-generated records do not contain human statements is very significant because the fact that a computer rather than a human being has generated the record has dramatic effects with respect to the nature of the evidence: if the record can no longer be considered hearsay on the grounds that it does not contains statements made by human beings outside the court, it cannot fall under the business records exception to the hearsay rule. As a consequence, the evidentiary issue is no longer whether a human's statement made out of court was truthful and accurate – a question of reliability – but whether the computer program that generated the record was functioning properly – a question of authenticity.[37]

35 John Henry Wigmore, *Evidence in Trials at Common Law*, ed. and rev. by James H. Chadbourn, Vol. 9 (Boston, 1978).

36 George L. Paul, *Foundations of Digital Evidence* (Chicago, 2008), pp. 116–18.

37 *Canada Evidence Act*, http://laws.justice.gc.ca/eng/C-5/index.html (accessed on 5 January 2010); Computer Crime and Intellectual Property Section Criminal Division, United States Department of Justice, *Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations*, http://www.cybercrime.gov/

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

The characteristics of computer-generated records make them similar to the scientific evidence mentioned earlier in relation to the *Daubert* rules or guidelines. These guidelines recommend assessing the trustworthiness of scientific evidence on the basis of four criteria: 1) whether the theory, procedure, or process generating it has been tested or can be tampered with; 2) whether it has been subjected to peer review or publication; 3) what is the known or potential error rate; and 4) whether it is generally accepted within the relevant scientific community.[38] In 1999, *Kumho Tire v. Carmichael* extended the *Daubert* guidelines to non-scientific evidence. The Court in fact decided that the four factors used to measure the reliability of scientific evidence could be applied just as effectively to evaluate technical or specialized evidence, such as computer-generated records.[39] This decision is not universally supported because many believe that, since software programs are designed and selected by humans, their output results from human intervention. Furthermore, others believe that, by considering computer-generated records equivalent to scientific evidence – and therefore by focusing on their source in terms of capability rather than intentions – the legal system degrades them from substantive evidence, that is, evidence offered to prove a factual issue, to demonstrative evidence, that is, evidence that has in itself no probative value but is used to illustrate and clarify the factual matter at issue.[40]

Digital forensics experts identify a third category of computer records: those that are both *computer-generated and computer-stored*. An example is a spreadsheet that has received human input followed by computer processing (the mathematical operations of the spreadsheet program). This very common category (e.g., most student records in a student registration system) falls both under the hearsay rule (and its business records exception) and the authenticity rules (such as the *Daubert* test) used for scientific/technical evidence.[41] The categorization of digital records made by forensics experts is clearly based exclusively on evidentiary purposes, but it is useful to all profession-

s&smanual2002.htm (accessed on 25 August 2009). See also Fred Galves and Christine Galves, "Ensuring the Admissibility of Electronic Forensic Evidence and Enhancing Its Probative Value at Trial," *Criminal Justice Magazine*, vol. 19, no. 1 (Spring 2004), http://www.abanet.org/crimjust/cjmag/19-1/electronic.html (accessed on 25 August 2009).

38 Carrier, p. 3.

39 Erin Kenneally, "Gatekeeping Out Of The Box: Open Source Software as a Mechanism to Assess Reliability for Digital Evidence," *Virginia Journal of Law and Technology*, vol. 6, no. 3 (2001), n.p., para. 34–35, http://www.vjolt.net/vol6/issue3/v6i3-a13-Kenneally.html (accessed on 25 August 2009).

40 Ibid., para. 59–63. To avoid this legal pitfall, George Paul insists that computer-generated records be considered hearsay as "the statements made by information systems might be dubious, since a human programmer told the computer what to say under certain circumstances. The true maker of the statement the system outputs is thus twice removed from the judicial proceeding." Paul, p. 142.

41 See footnote 37 above.

als who fulfill the function of trusted recordkeeper or custodian, in that it supports their responsibility of guaranteeing and perpetuating the authenticity of the records placed in their trust. Equally important are the ideas that digital forensics experts hold about the larger concept of trustworthiness. These ideas at times coincide with digital diplomatics concepts and at other times complement them, and for this reason it is best to start explaining the view of digital diplomatics.

The Concept of Record Trustworthiness in the Digital Environment

Diplomatically, a digital record is trustworthy if it is accurate, reliable, and authentic. Record *accuracy* has never been a consideration in general diplomatics because, as a concept, it was subsumed under both reliability and authenticity. Accuracy is the trustworthiness of the data (i.e., the smallest, meaningful, indivisible pieces of information) within a record, and is defined as their truthfulness, exactness, precision, or completeness. In the digital environment, it is necessary to consider and assess accuracy as a separate quality of a record because of the ease with which data can be corrupted during transmission across space (between persons and/or systems) and time (when digital systems are upgraded or records are migrated to a new system). Consequently, accuracy is a shifting responsibility that moves over time from the creator's trusted recordkeeper to the trusted custodian.

Reliability is the trustworthiness of a record as a statement of fact, as to content. It is assessed on the basis of 1) the completeness of the record, that is, the presence of all the formal elements required by the juridical-administrative system for that specific record to be capable of achieving the purposes for which it was generated; and 2) the controls exercised on the process of creation of the record, among which are included those exercised on the author of the record, who must be the person competent, that is, having the authority and the capacity, to issue it. The reliability of a record is the exclusive responsibility of its creator and the trusted recordkeeper, that is, of the person or organization that made or received it and maintained it with its other records.

Authenticity is the trustworthiness of a record as a record, and is defined as the fact that a record has not been tampered with or corrupted, either accidentally or maliciously. An authentic record is one that preserves the same identity it had when first generated, and can be presumed or proven to have maintained its integrity over time. The identity of a record is constituted of the whole of those characteristics that distinguish it from any other record, and is assessed on the basis of the formal elements on the face of the record, and/or its attributes, as expressed for example in a register entry or as metadata. The metadata that attest to the identity of a record are the names of the five persons concurring in its creation; the date(s) and time(s) of its issuing, creation, and transmission; the matter or action in which it participates; the expression of its archival bond

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

(e.g., its classification code); its documentary form; its digital format; the indication of any attachment(s); the indication of the presence of a digital signature, if applicable; and the name of the person/office handling the business matter in which the record participates. The integrity of a record is linked to its ability to convey the message it was intended to communicate when generated. Thus, it does not matter if the ink is fading, the medium (i.e., the material support) is falling apart, or the bit-stream is not the same as in the first manifestation of the record, as long as the content is readable and is the same as it was originally intended, the medium does not have missing parts, or the manifestation we see on the computer screen is the same as it was the first time the record was saved. The integrity of a record is inferred not only from its appearance – which might be deceiving in the case of good forgeries – but also from the circumstances of its maintenance and preservation: an unbroken chain of responsible and legitimate custody is considered an insurance of integrity until proof to the contrary, and integrity metadata are required to attest to that. They are: the name(s) of the persons/offices handling the record over time; the name of the person/office responsible for keeping the record; the indication of annotations, if applicable; the indication of technical changes, if applicable; the indication of presence or removal of digital signature; the time of planned removal of the record from the digital system; the time of transfer to a trusted custodian; the time of planned deletion; and the existence and location of duplicates outside the system. In the absence of sufficient evidence linked to the record, authenticity can be inferred on the basis of the trustworthiness of the record system in which the record exists. The necessary requirements for a trusted record system will be discussed later on. The authenticity of a record is a movable responsibility, as it shifts from the creator's trusted recordkeeper, who needs to guarantee it for as long as the record is in its custody, to the trusted custodian, who guarantees it for as long as the record exists.

The trustworthiness of a record, consisting of its accuracy, reliability, and authenticity, should not be confused with one of the means of protecting and/or establishing it: authentication. *Authentication* is defined as a declaration of authenticity made by a competent officer, and consists of a statement or an element, such as a seal, a stamp, or a symbol, added to the record after its completion. While authenticity is a quality of the record that accompanies it for as long as it exists as is, authentication only guarantees that a record is authentic at one specific moment in time, when the declaration is made or the authenticating element or entity is affixed. In the digital environment, extreme authentication is usually provided by a digital signature. The digital signature has the function of a seal because it is attached to a complete record, allows verification of the origin and integrity of the record, and makes the record indisputable and incontestable by performing a non-repudiation function. As Heather MacNeil contends:

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).[42]

The authentication provided by a digital signature is considered "extreme" because the test fails if even one bit changes. Digital forensics experts are beginning to see that biometric identification systems and cryptography, not being in common use, cannot be considered the prevalent means of authentication. Mocas writes that, although "[w]ithout such mechanisms, authentication (computer security) can be difficult to positively establish based simply on digital evidence … there is interesting research on ways that digital information can be used to indicate a possible author. For example, there is work on establishing personal characteristics based on document features."[43] She refers to Del Vel, Corney, Anderson, and Mohay, who wrote in 2002 about formal characteristics of email, developing some sort of diplomatic analysis of this specific documentary form leading to the identification of the authors.[44]

Another form of authentication is a declaration made by an expert who bases it on the trustworthiness of the system containing the record and of the procedures controlling it. This raises the issue of when a record system can be trusted. Digital diplomatics assesses the trustworthiness of a record system according to the same criteria used by general diplomatics to assess the trustworthiness of chancery procedures and processes: the level of standardization of, and control on, record systems, that is, on the set of rules governing the making and keeping of records, and the set of tools and mechanisms used to implement these rules. In order to generate reliable and accurate records, every record-making system should include in its design integrated business and documentary procedures, record metadata schemes, records forms, and record-making access privileges, and should fulfill technological requirements that ensure the integrity of the system.

Integrated business and documentary procedures are business procedures linked to documentation procedures and the classification system established

42 MacNeil, "Providing Grounds for Trust," p. 62.

43 Mocas, p. 66.

44 Olivier De Vel, Malcom Corney, Alison Anderson, and George Mohay, "Language and Gender Analysis of E-mail Authorship for Computer Forensics," *Proceedings of Second Digital Forensic Research Workshop* (2002).

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

in the organization. This integration reinforces the control over record-making procedures by supporting the reliability and accuracy of records that are explicitly connected to the activities in which they participate and to the records organization system, thereby standardizing the procedures for creating and managing those records. The integration of business and documentary procedures also establishes the basis and central means for demonstrating ownership of, and responsibility for, the records. A record-making metadata scheme is a list of all metadata required for uniquely identifying each record, and enabling the maintenance of its integrity and the presumption of its authenticity. Such a scheme can also be used later on to verify authenticity when questioned. Record forms are specifications of the documentary presentation for the records generated in the system. Access privileges refer to the authority to compile, edit, annotate, read, retrieve, transfer, and/or destroy records in the record-making system, granted to officers and employees by the organization on the basis of job duties and business needs. Access privileges control access to the record-making system, and are established in the course of integrating business and documentary procedures through connecting specific classes of records to the office of primary responsibility for a business function or activity. The establishment and implementation of access privileges is the most important step toward ensuring that the reliability and accuracy of records can be presumed.

In its design, a trusted record-keeping system should include a record-keeping metadata scheme, a classification scheme, a retention schedule, a registration system, a record-keeping retrieval system and access privileges, and procedures for maintaining authentic records. A record-keeping metadata scheme is the list of all metadata required to ensure each record's continuing identity and integrity in the record-keeping system. The classification scheme should take the form of a plan for the systematic identification and arrangement of business activities and related records into categories according to logically structured conventions, methods and procedural rules, and should be linked to retention schedules specifying and authorizing the disposition of records series and/or classes as identified in the classification scheme. A registration system assigning a unique identifier to each created record, linked to its identity and integrity metadata, is an additional instrument of control that also keeps track of all destroyed records, as it can be kept as evidence of their past existence and proper deletion. Record-keeping access privileges refer to the authority to classify, annotate, read, retrieve, transfer, and/or destroy records in the record-keeping system, granted to officers and employees by the organization based on job duties and business needs. Typically, access to records for the purpose of classification, transfer, and destruction is given only to the record officer (records manager or archivist) of the organization. The record-keeping retrieval system should include a set of rules governing the searching and finding of records and/or information

about records in the system based on the identity metadata, the classification, and the registration systems. Similar to digital forensics – as we will see – digital diplomatics identifies the procedures for maintaining authentic records with the procedures designed to ensure that the identity and integrity of the records in the record-keeping system are protected. Also, consistent with digital forensics as well as general diplomatics, digital diplomatics considers the role of the recordkeeper as trusted custodian key to a presumption of trustworthiness for digital records.[45]

To demonstrate the point made above, having examined records trustworthiness from the perspective of digital diplomatics, we now proceed to examine it from the perspective of digital forensics. Digital forensics does not use the term trustworthiness in relation to records or records systems other than in a general way. The terms most commonly found in digital forensics literature are authenticity, accuracy, reliability, and integrity. Although these qualities are all required to support "circumstantial guarantees of trustworthiness" of digital materials presented as evidence of facts at issue, the first two are primarily used in relation to records, while the other two are applied to systems or media and only indirectly to records. The term *authenticity,* which is used in the context of admissibility rules, is a pre-condition to admissibility of evidence related to the matter in question, and refers to the fact that "the data or content of the record" are what they purport to be and were produced by, or came from, the "source"[46] they are claimed to have been produced by or come from. It is to be noted that, in digital forensics, like in diplomatics, while authenticity implies integrity, the opposite is not true, that is, integrity does not imply authenticity.[47] Proof of authenticity – or authentication of evidence – is provided by a witness who can testify about the existence and/or substance of the record on the basis of his/her familiarity with it, or, in the absence of such person, by a computer programmer showing that the computer process or system produces accurate results when used and operated properly, and that it was so employed when the evidence was generated.[48] Galves and Galves suggest that "to enhance the strength of circumstantial digital evidence" one could examine metadata "which records (1) the exact dates and times of any messages sent or received, (2) which computer(s) actually

45 See "A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-Term Preservation of Digital Records," in Duranti and Preston.

46 In digital forensics, the term "source" is used in a general way to refer to either a person (physical or juridical), a system, a software, or a piece of hardware.

47 Mocas, p. 66.

48 Fred Galves, "Where the Not-So-Wild Things Are: Computers in the Courtroom, the Federal Rules of Evidence, and the Need for Institutional Reform and More Judicial Acceptance," *Harvard Journal of Law and Technology*, vol. 13, no. 2 (Spring 2002), pp. 161, 230.

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

created them, and (3) which computer(s) received them." Although in a different context, these authors also argue the importance of a chain of legitimate custody for inferring authenticity.[49] A presumption of authenticity is afforded to evidence such as x-rays, photographs, tape recordings, computer-generated records, or scientific surveys produced by an automated process that is shown to render accurate results.[50] This presumption has been commonly extended to records managed by software performing data storage, collection, or retrieval functions, if the operation of the software can be proven to have been reliable. Indeed, a majority of the cases considering the admissibility of such evidence has done so in the context of computerized business records that are maintained or prepared by electronic computing equipment. However,

> … the evidentiary challenges to computer-derived evidence have gone to the weight of the evidence, rather than its admissibility. Thus, the evidence may be allowed to go before a jury, but its reliability is contested by attacking the chain-of-custody (human handling) of the digital data.[51]

Regardless of the fact that, in digital forensics, references to authenticity appear to focus on the data or content in the record rather that on its formal aspects, like diplomatics, the importance of protecting both the documentary and digital presentation of a record for purposes of authentication is implicit in the discussion of digital forensics practices. For example, Ghirardini and Faggioli state that, although conversion of digital evidence to forms and formats different from the original is a process useful to its accessibility and analysis, it "modifies its nature." This implies that converted records cannot be used as evidence and must always be accompanied by the records in the original presentation.[52] Although these authors write in the context of a civil law system, which does not consider records hearsay and rules them admissible if proven authentic, the issue they identify also relates to the "best evidence" requirement for admissibility in a common law context, according to which evidence must be submitted in the most authoritative status of transmission, which is the original or an authenticated copy of the original when the former is not accessible.[53] Indeed,

49 Galves and Galves, http://www.abanet.org/crimjust/cjmag/19-1/electronic.html (accessed on 25 August 2009).

50 See *People v. Lugashi*, 252 Cal. Rptr. 434 (Cal. Ct. App. 1988) (presuming a data collection software program accurate); *People v. Mormon*, 422 N.E.2d 1065, 1073 (1981) (presuming a data retrieval program accurate).

51 Kenneally, n.p., para. 45, endnote 50.

52 Andrea Ghirardini and Gabriele Faggioli, *Computer Forensics* (Milano, 2007), p. 178.

53 The best evidence rule stipulates the requirement to produce an original record when it exists. For an explanation from a record point of view, see Heather MacNeil, *Trusting Records. Legal, Historical, and Diplomatic Perspectives* (Dordrecht, 2000), pp. 48–50. For purely legal explanation, see also Wigmore, vol. 4, p. 396, para. 1172–78. "Status of transmission," according to general diplomatics, is the degree of perfection of a record.

in the digital environment, there are no originals in the diplomatics sense, that is, there are no records which, in addition to being complete and capable of reaching the purposes for which they were generated (i.e., effective) are also the first instance of each item under consideration, because when we close a digital record for the first time we destroy the original and every time we open it we create a copy. However, we can state that each digital record, in the last version used by the creator in the usual and ordinary course of business, is a copy in the form of original and, in any version kept by the preserver, is an authentic copy of the record of the creator. They are both authoritative and authentic if their identity is intact and their integrity can be either presumed or proven.

The concept of *accuracy* is not clearly defined in digital forensics, primarily because the law does not include it in its procedures or rules concerning the presentation or evaluation of evidence. However, both legal and digital forensics writers use accuracy as a component of authenticity and, specifically, integrity, in a meaning very similar to that given to the term by digital diplomatics, and it is one of the qualities of the evidence to which digital forensics practitioners pay more attention. In fact, their processes for extracting digital evidence must, first of all, avoid altering the data, and are guaranteed reliable in such sense by ensuring that they are repeatable. "Repeatability," which is one of the fundamental precepts of digital forensics practice, is supported by the accurate documentation of each and every action carried out on the evidence.[54] This is certainly an area in which digital diplomatics would have much to learn from digital forensics, especially when it is used in a prospective way to support the selection of the best software for a record-making or a record-keeping system, and the definition of transfer procedures from the creator to the preserver, as will be seen later.

There is a general agreement among legal and digital forensics experts that open source software is the best choice from an evidentiary point of view both as a records source, and as a tool for extraction and preservation. Their arguments are based on the fact that the judiciary, in assessing accuracy, integrity, and reliability, uses measurements such as objectivity, transparency, verifiability, and repeatability.[55] In addition, digital forensics experts value the availability of open source, which, at the same time, allows modification and encourages dissemination, thereby making it possible to submit the software together with the records presented as evidence, so that their accuracy can be tested promptly by anyone at any time. This is especially true when conversion or migration occurs, because it would allow a practical demonstration that the software could not simultan-

The "original" is the perfect record, in that it is the first complete and effective record. A "copy in the form of original" (i.e., a second original issued later) and a copy authenticated by a public officer, have the same legal authority as the original. Other types of copies and drafts are lower on the authority scale. See Duranti, "Diplomatics. Part I," pp. 19–21.

54 Ghirardini and Faggioli, p. 230.

55 See for example Carrier, p. 3; and Kenneally, n.p., para. 70.

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

eously manipulate the files' content while copying them, and that nothing could be altered, lost, planted, or destroyed. Finally, open source is preferred because of the possibility of exchange of evidentiary material between the parties in the course of e-discovery.[56]

The concept of *reliability*, used in reference to the source of the records, is defined in digital forensics in a way that points to a reliable software, measured by principles similar to those the courts use to determine evidentiary reliability, that is, empirical testing, subjection to peer review and publication, determination of error rate, and general acceptance within the relevant community.[57] Also these principles point to open source software because the processes of records creation and maintenance can be authenticated with evidence either by describing a process or system used to produce a result, or by showing that the process or system produces an accurate result.[58]

The concept of *integrity* is more nuanced. Digital forensics distinguishes *data integrity* from *duplication integrity*; clearly this distinction is very important for digital diplomatics, which concerns itself with the authenticity of the copies made in the course of digital records maintenance and preservation. Indeed, considering that it is not possible to preserve digital records, but only the ability to reproduce them, the concept of duplication integrity is key to digital preservation and the functions of the designated trusted custodian.[59] Carl E. Landwehr defines data integrity as the fact that data are not modified either intentionally or accidentally "without proper authorization."[60]

56 See for example Ghirardini and Faggioli, pp. 233–34. The Sedona Conference defines "discovery" as "the process of identifying, locating, securing, and processing information and materials for the purpose of obtaining evidence for utilization in the legal process." Similarly, it defines electronic discovery as the process of "collecting, preparing, reviewing, and producing electronically stored information (ESI) in the context of the legal process." Conor R. Crowley and Sherry B. Harris, eds., *The Sedona Conference Glossary: E-Discovery and Digital Information Management,* 2nd ed. (December 2007), http://www.thesedonaconference.org/dltForm?did=TSCGlossary_12_07.pdf (accessed on 25 August 2009).

57 Indeed, George Paul, having stated that computer-generated records are hearsay, in order to accommodate them, proposes to add a second exception to the business records exceptions to the hearsay rule, an exception called "system reliability." "*System reliability* is a two-pronged concept. At issue are both the concept of *reliability* (the accuracy and trustworthiness of the end product ...) and the concept of *authenticity* (primarily *integrity*, given that information in systems is subject to change)." Paul, p. 131. "Accordingly, proponents of computer-generated information, will need to lay a foundation to qualify statements as reliable under a system reliability exception.... The court will act as a trier of fact to determine the competency of the evidence." Ibid., p. 145.

58 Committee on Information Systems Trustworthiness, Computer Science and Telecommunications Board, Commission on Physical Sciences, Mathematics, and Applications, *Trust in Cyberspace* (Washington, DC, 1998), p. 154.

59 Duranti and Thibodeau, p. 19.

60 Carl E. Landwehr, "Computer Security," *International Journal of Information Security*,

Duplication integrity is ensured when "given a data set, the process of creating a duplicate of the data does not modify the data (either intentionally or accidentally) and the duplicate is an exact bit copy of the original data set."[61] Sarah Mocas believes that separating these two notions of integrity is important because the concept is too broad to be able to address the aspects of each given situation, and because most of digital forensics work, just like records maintenance and preservation work, is carried out over duplicates.

> It is possible to preserve data integrity over the duplicate, with respect to the original, by using a *trusted third party*. At the time the image is created, a copy of the hash can be given to a trusted third party to hold in escrow. Now changes to the duplicate can be detected even if the original is modified. One possibility is to have the same technology that is creating the duplicate transmit the hash and a device ID to a trusted database. The escrow device can then acknowledge receipt by returning the same information and adding a time stamp.[62]

Digital forensics experts also link duplication integrity to time and have considered the use of time stamps for that purpose.[63] A distinction between the integrity of a record as such and that of its duplicate may be useful to eliminate the conflict between the view of integrity held by diplomatists and that held by information technology experts, who tend to support the need for the extreme authentication provided by a digital signature. Indeed, one could further enrich the concept of integrity by also adopting the link between integrity and time proposed by digital forensics experts, and define record integrity differently in each phase of the record life cycle and/or custodial history.

Clearly, digital forensics has a very high stake in the trustworthiness of the evidence gathered, maintained, and submitted to court. Two principles that are at the foundation of forensic practice and could be very useful to a trusted preserver are those of *non-interference* and *identifiable interference*. The former means that the method used to gather and analyze digital data or records does not change the original digital entities. The latter means that, if the method used does alter the original entities, the changes are identifiable.[64] These principles, which embody the ethical and professional stance of digital forensics experts, are consistent with the traditional impartial stance of the

vol. 1, no. 1 (August 2001), pp. 3–13.

61 Mocas, p. 65.

62 Ibid., p. 66. Emphasis added.

63 Michael Duren and Chet Hosmer, "Can Digital Evidence Endure the Test of Time?" *Proceedings of the Second Digital Forensic Research Workshop 2002* (7 August 2002), n.p.

64 Eoghan Casey, "Error, Uncertainty and Loss in Digital Evidence," *International Journal of Digital Evidence*, vol. 1, no. 2 (Summer 2002), n.p.

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

archivist mentioned by Diamond and so important to Jenkinson, as well as with his/her new responsibility of neutral third party, of trusted custodian. These principles are at the core of digital forensics procedures, the knowledge of which could provide great support to the archivist working with digital records, especially with regard to the activity that represents the weakest link in the chain of records preservation,[65] the transfer of the records from the creator to the preserver. As well, digital forensics experts could derive useful input from diplomatic criticism, the method used by diplomatists to identify records among other types of materials and to assess their trustworthiness. Thus, the following section briefly outlines the methods of digital diplomatics and digital forensics, beginning with the former.

The Methods of Digital Diplomatics and Digital Forensics

General diplomatics uses criticism of the formal elements of a record to determine its identity and integrity. Although the type of analysis conducted and the elements that are the object of such analysis depend on the time and the geo-political-administrative area in which the record was generated, the concepts and the procedure guiding diplomatic criticism have remained stable over time.[66] While in the traditional environment it was easy enough to understand at a glance when the entity one was observing was a record or not, in the digital environment such an endeavour can be quite complex and requires additional steps in the procedure of analysis. The development of a template for conducting diplomatic criticism of digital entities is one of the outcomes of the second phase of the InterPARES Project that has been further refined in the ongoing third phase.[67]

65 The concept of "Chain of Preservation" was introduced by the InterPARES research project in its first phase and resulted in the model of the records life cycle produced by the second phase of the research. The InterPARES Terminology Database defines it as a "system of controls that extends over the entire lifecycle of records in order to ensure their identity and integrity over time," http://www.interpares.org/ip2/ip2_terminology_db.cfm. The Chain of Preservation model is available at http://www.interpares.org/ip2/ip2_model_display.cfm?model=cop (both accessed on 25 August 2009).

66 See Luciana Duranti, "Diplomatics: New Uses for An Old Science. Part V," *Archivaria* 32 (Summer 1991), pp. 16–21.

67 The first phase of InterPARES developed a Template for Analysis that for the most part reflected that of traditional diplomatics. Its application to databases and document management systems did not encounter major difficulties, but its use in the second phase of InterPARES, which focused on interactive and dynamic systems, proved to be impossible. Thus a new template was generated that worked with the very complex systems examined in the course of the various case studies, but was too high level for the situations dealt with in the third phase of InterPARES and required further refinement. See "Appendix 1. Template for Analysis," in Duranti, *The Long-term Preservation of Authentic Electronic Records*. See also "Appendix 7. Diplomatic Analysis Template," in Duranti and Preston.

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

Diplomatic criticism of digital entities starts with a description of the technological environment in which the entities exist, and of their digital and documentary presentations. It then proceeds to investigating the presence or absence of the six requirements for the existence of a record discussed in the first part of this article, paying particular attention to the last two requirements, fixed form and stable content. If the entity participates in, or is associated with, an action, this connection is examined and qualified, and so are the intellectual relationships among entities, and the relationship between each entity and each of the persons involved in its creation. Finally, the various contexts are analyzed in depth, especially the procedural context if the entity is linked to an action, the documentary context if the entity is explicitly linked to other entities inside or outside the system that qualify as documents, and the technological context if the entity has direct links with other digital entities.

At this point, the diplomatic criticism should conclude whether the entity in question is a record or not. If the answer is negative, the analysis should further illustrate and explain the status of the digital entity as a data set, a publication or a potential record, and its identifying characteristics, attributes and behaviour, and recommend the most appropriate actions depending on the purpose of the investigation. If the answer is positive, a more detailed analysis is conducted to determine whether the record is trustworthy, what are its salient characteristics that need to be protected for it to maintain its identity and integrity over time, by which formal elements they are expressed, and in which digital components they reside.

The deep understanding of digital materials gained through this analysis would undoubtedly be useful to digital forensics experts, whose primary and most delicate task is to extract potential evidence from a digital environment without interfering with it, that is, maintaining intact the identity of the evidence and protecting its integrity, as well as the integrity of the reproductions they make.

A brief description of the digital forensics procedure will make clear this point and also illustrate how useful some methodological aspects of such procedure could be to a trusted custodian of digital records. The digital forensics literature is not very consistent as it regards the terminology used to name the steps of the procedure, but the substance of each step and its sequence is very clear. The first phase consists of the location and recovery of the digital evidence. This involves taking anti-contamination precautions; searching the scene; collecting the evidence; packaging and labelling; and documenting at every step of the way what is done. The second phase consists of prioritizing the examination of the potential evidence determining which items are the most likely to serve the purposes of the investigation and which are more time sensitive, which are most at risk of being lost or corrupted, etc. The third phase is the examination of the recovered material. It involves considering

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

all the anti-contamination measures necessary for the specific case; a review of the integrity of the packaging; establishing precautions against external hazards such as electrical static, magnetism, etc.; reviewing the analysis protocol; and writing a record of what is being done:

> … the records should be in sufficient detail to allow another examiner, competent in the same area of expertise, to be able to identify what has been done and to assess the findings independently. Case records should include both administrative and examination documentation. Whenever appropriate standardised forms should be used to document examinations.... casework involving digital evidence should include details of case records such as notes, work sheets, photographs, printouts, charts, spectra and other data or records which support findings should be generated during the course of the examination, and kept.[68]

The fourth phase is the evaluation and interpretation of the findings, and the fifth phase is the presentation of the results in a report that should include factual findings, interpretation, and expert opinion. Finally, all the work done undergoes technical and administrative review. The technical review is carried out by a qualified person designated by the organization, agency, or institution, who considers the validity of the raw data, of the findings, and of the conclusions drawn from them, and produces a report that will remain with the case file. The administrative review ensures that the needs of the investigation have been properly addressed and served, and that the applicable policies have been observed.[69]

There are many ways in which digital diplomatics and digital forensics procedures could be integrated in support of both disciplines and professions.[70] However, this author envisions the development of a new science

68 International Organization on Computer Evidence, Digital Evidence Standard Working Group, *Guidelines for Best Practice in the Forensic Examination of Digital Technology* (IOCE, 2002), http://www.ioce.org/fileadmin/user_upload/2002/ioce_bp_exam_digit_tech.html (accessed on 25 August 2009).

69 See also Mark Reith, Clint Carr, and Gregg Gunsch, "An Examination of Digital Forensic Models," *International Journal of Digital Evidence*, vol. 1, no. 3 (Fall 2002), n.p.; Association of Chief Police Officers (ACPO), *Good Practice Guide for Computer-based Electronic Evidence*, version 4.0 (London, UK, n.d.), pp. 1–36, http://www.7safe.com/electronic_evidence/ACPO_guidelines_computer_evidence.pdf (accessed on 25 August 2 009). British Columbia, Supreme Court, *British Columbia Electronic Evidence Project* (1 July 2006), http://www.courts.gov.bc.ca/supreme_court/Practice_directions_and_notices/electronic_evidence_project.aspx (accessed on 25 August 2009). Computer Crime and Intellectual Property Section Criminal Division, United States Department of Justice, http://www.cybercrime.gov/s&smanual2002.htm (accessed on 25 August 2009).

70 Indeed, digital forensics experts have already had this idea. See for example K.A. Ferguson-Boucher and B. Endicott-Popovsky, "Digital Forensics and Records Management: What We Can Learn From the Discipline of Archiving," *Information Security Compliance and Risk Management Institute Conference Proceedings* (Seattle, 2008); and Alastair

that will serve records managers and archivists, the law enforcement and the legal professions, the information technology profession and any scholar who is interested in the analysis and understanding of digital records: a Digital Records Forensics.[71]

Toward a Digital Records Forensics

Both diplomatics and forensics were developed as practices for the purpose of investigating existing material evidence, assessing its status of transmission, its authenticity, and its ability to provide proof of facts at issue. In order to do so, they have articulated concepts, principles, methods, and procedures that support their purposes, admittedly at a very different level of sophistication, diplomatics being a centuries-old science and forensics a decades-old practice. However, they both use as primary terms of reference the legal system of the geo-political contexts in which they operate and in which the object of their study originated; they both aim to see the universal beyond the particular in each situation; they both are retrospective in their outlook in that they examine what exists rather than propose what should exist; and they both have sought the intellectual support of disciplines that study the same kind of material that is the object of their analysis: records in the case of diplomatics; various objects, substances, and traces in the case of forensics. Thus, whereas over time diplomatics merged its body of theory with archival science and used the support of philological and historical sciences, forensics has relied on the support of the disciplines that best studied the material under investigation, such as medicine, mathematics, engineering, and computer science. The latter was mostly used to develop the tools necessary to process the collected evidence and the information about it, but the increasing use of its body of knowledge changed the name of the practice from forensics to computer forensics. More recently, with the widespread use of digital technologies in every kind of activity, much of the evidence that came to be examined by computer forensics began taking digital form (as opposed to physical form), and a new branch of forensics practice developed – digital forensics – in the same way in which, confronted with the abundance of digital records, digital

Irons, "Computer Forensics and Records Management – Compatible Disciplines," *Records Management Journal*, vol. 16, no. 2 (2006), p. 111.

71 The name Digital Records Forensics was created by this author as the title of a research project she is conducting together with Anthony Sheppard (Professor of Law of Evidence, Faculty of Law, University of British Columbia), and in collaboration with Inspector Kevin McQuiggin (Vancouver Police Department, Forensics Division) and Victoria Lemieux (Professor of Archival Studies, University of British Columbia, and Director, Centre of Investigation for Financial Electronic Records [CiFER]), funded by the Social Sciences and Humanities Research Council of Canada for the period 2008–2011. See http://www.digitalrecordsforensics.org (accessed on 25 August 2009).

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

diplomatics arose out of general diplomatics.

Conclusion

This article has endeavoured to show how the concepts, principles, methods, and procedures developed in the context of the two knowledge areas are consistent with each other and complementary. While the focus of diplomatics is records by definition (*de re diplomatica* means about records), the focus of digital forensics is any kind of digital evidence; but this field is increasingly wrestling with the issues of 1) what digital evidence can be regarded as hearsay and fall under the business records exception to the hearsay rule, especially in the context of e-discovery, and 2) how to maintain evidence extracted and/or reproduced from computer hardware and networks over the long term in a trustworthy manner. At the same time, records professionals are in need of more sophisticated methods of analyzing, acquiring, reproducing, and preserving digital records in accurate and reliable ways. A digital records forensics would serve both forensic work, and records management and preservation not only by integrating the two bodies of knowledge and developing new knowledge based on them, but also by deriving from them a new "prospective" approach which would be added to their traditional "retrospective" approach. Instead of simply looking at what exists, this new discipline will grow by using the knowledge acquired through retrospective examination to develop a theory of what should exist. Ideally, a digital records forensics would articulate:

- how a variety of digital systems should be designed to create and maintain trustworthy digital records that can be regarded as material evidence of facts and acts, serving at the same time transparency, accountability, and users' needs;
- how the authenticity of digital records can be verified when its presumption is weak;
- how the records of the creators should be reliably extracted from the systems in which they reside, and maintained in long-term storage either with the creator or with the preserver in such a way that their authenticity can be presumed;
- how records should be authentically reproduced in the course of their long-term preservation;
- how the features of the records, the actions conducted over them, and the changes caused by such actions should be documented;
- how the records submitted to court as evidence should be kept after the conclusion of court proceedings for as long as needed, so that they remain

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved

trustworthy[72];
- how long-term preservation activities can be conducted in such a way that they would not interfere with the applicability of the business records exception to the hearsay rule[73];
- and much more.

This is a very ambitious and challenging project that will require the collaboration of scholars and professionals in all the areas involved, but it is one worth undertaking, as the challenge presented by digital records is one that archivists cannot and should not meet in isolation. If records are – as Jenkinson stated and Elizabeth Diamond so passionately reminded us – the "material evidences" of actions and events, then we need to protect them as such, building research alliances that foster the development of new knowledge in the forensic arena, and using our combined expertise and imagination to create a new science that can support the role that we are repeatedly called to take on, that of trusted keepers of the authentic record of our past.

72 The final report by Commissioner MacCallum into the wrongful conviction of David Milgaard issued in September 2008 contains the following recommendations: "9. In all indictable offence cases, documentary exhibits should be scanned and stored electronically, unless a court orders otherwise. 10. All prosecution and police files, including police notebooks, relating to indictable offences should be retained in their original form for a year, then scanned and entered into a database where a permanent, secure electronic record can be kept." See http://www.milgaardinquiry.ca (accessed on 25 August 2009). See also Christopher Sherrin, "Preserving Evidence for the Innocent," *The Lawyers Weekly*, vol. 28, no. 23 (17 October 2008), p. 7. Sherrin says the issue has already been debated in the US and "numerous pieces" of American legislation have already been enacted for preserving evidence. He also says the police, Crown prosecutors, and forensic labs in Canada lack consistent policies.

73 In *Ak-Chin Indian Community v. United States*, 85 Fed. Cl. 397 (2009), records arranged differently from the way they were filed prior to transfer to long-term storage were held not to have been kept in the ordinary course of business. For documents transported to storage to still be considered as kept in the ordinary course of business, the court said (quoting RCFC 34(b)(2)(E)(i)), the documents must be stored in the same way they were originally kept: one more argument in support of respect for original order!

Archivaria, The Journal of the Association of Canadian Archivists – All rights reserved