Authenticating Digital Records: The Archivist as a Forensics Expert

Authenticating digital records means assessing their trustworthiness as records and as statements of fact. The archival concept of trustworthiness is rooted in Roman law, which defined an archives as a trusted place, and based records authenticity on a chain of trusted custody and records reliability on their antiquity and form.[1] This archival concept also relies on the understanding of authenticity provided by diplomatics, which in the 18th century entered European universities and provided the foundation of the law of evidence as we know it today both in civil law and common law countries. Diplomatics, consistently with Roman law, also assesses trustworthiness by examining records form.[2]

Archival science and diplomatics have served archives and archivists quite well in the past. However, as digital technology has separated content and structure from form, we can no longer determine authenticity on the exclusive basis of the form of the object-record, which is composite and permanently new,[3] but must make an inference of authenticity from its environment. For this we need the help of a relatively new body of knowledge, Digital Forensics. In fact, archivists are increasingly called to act as forensics experts, for example, by being asked to ensure the identity and integrity of digital records through time and attest to it, and to acquire such records, often from obsolete systems or portable media, without altering them in the process.

Also digital forensic experts are called to act as archivists, for example, when they are asked to identify what digital materials fall under the definition of records, and to keep them intact for as long as needed. They are also called to attest to the integrity of digital systems, to provide quality assurance for digital system that produce, contain or preserve records, to assess whether spoliation (i.e., fraudulent disposal) has occurred, and to ensure that e-discovery requirements are fulfilled.

Digital forensics is defined as "the use of scientifically derived and proven methods toward the collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events, or helping to anticipate unauthorized or inappropriate actions."[4] Its methods are based on conceptual assumptions about records, trustworthiness, and recordkeeping which are slightly different from the archival and diplomatic ones, but not necessarily conflicting. This paper will focus on the assumptions regarding trustworthiness and, specifically, integrity.

---

[1] Justinian, *Corpus Juris Civilis*, Novella I, 5, 5 and Codices I, 4, 30.

[2] Luciana Duranti, "Diplomatics," in *Encyclopedia of Library and Information Science*. Marcia Bates, Mary Niles Maack, Miriam Drake eds.(New York, Basel, Hong Kong: Marcel Dekker, INC., 2009).

[3] We know that we cannot keep digital records, but only maintain the ability to reproduce them, which are therefore always new objects, made up of several digital components.

[4] *Digital Forensics Research Workshop*, 2001, online at http://www.dfrws.org/2001/dfrws-rm-final.pdf, p. 16.

In our archival view, records trustworthiness is composed of three qualities: reliability, accuracy and authenticity. Reliability is defined as the trustworthiness of a record[5] as a statement of fact, based on the competence of its author, its completeness, and the controls on its creation; accuracy is defined as the correctness and precision of a record's content, based on the above and on the controls on content recording and transmission; and authenticity is defined as the trustworthiness of a record that is what it purports to be, untampered with and uncorrupted, based on its identity and integrity, and on the reliability of the records system in which it resides. Authenticity is on turn composed of identity and integrity, where identity is the whole of the attributes of a record that characterize it as unique and distinguish it from other records (e.g. date, author, addressee, subject, classification code), and integrity is the quality of a record that is capable of transmitting exactly the message it is meant to communicate in order to achieve its purpose (e.g. text and form fidelity, absence of technical changes). Both identity and integrity are assessed in context, that is, in light of the administrative-juridical, provenancial, procedural, documentary and technological environment in which the record was created (i.e., made or received and set aside for further action or reference) and used overtime.

In contrast with the archival view, the digital forensics view of trustworthiness is linked to the type of document that is the object of its consideration. Digital forensics divides documents in three groups: 1) Computer Stored Documents, which contain human statements and, if created in the course of business, are records or archival documents (e.g. e-mail messages, word processing documents) and can be used in a court of law as substantive evidence; 2) Computer Generated Documents, which do not contain human statements, but are the output of a computer program designed to process input following a defined algorithm (e.g. server log-in records from Internet service providers, ATM records) and in a court of law can only be used as demonstrative evidence;[6] and 3) Computer Stored & Generated Documents, which are a combination of the two (e.g. a spreadsheet that has received human input followed by computer processing, that is, by the mathematical operations of the spreadsheet program) and can be used in a court of law in either way.

According to digital forensics, reliability is the trustworthiness of a record as to its *source*, defined in a way that points to either a reliable person (for computer stored documents) or a reliable software (for computer generated documents), or both. If the source is a software, it should be open source, because the processes of records creation and maintenance can be forensically authenticated either by describing a process or system used to produce a result or by showing that the process or system produces an accurate result, and open source allows for both types of authentication.

Accuracy is instead a component of authenticity and, specifically, integrity. Digital entities are guaranteed accurate if they are repeatable, that is, if the same process carried out on them

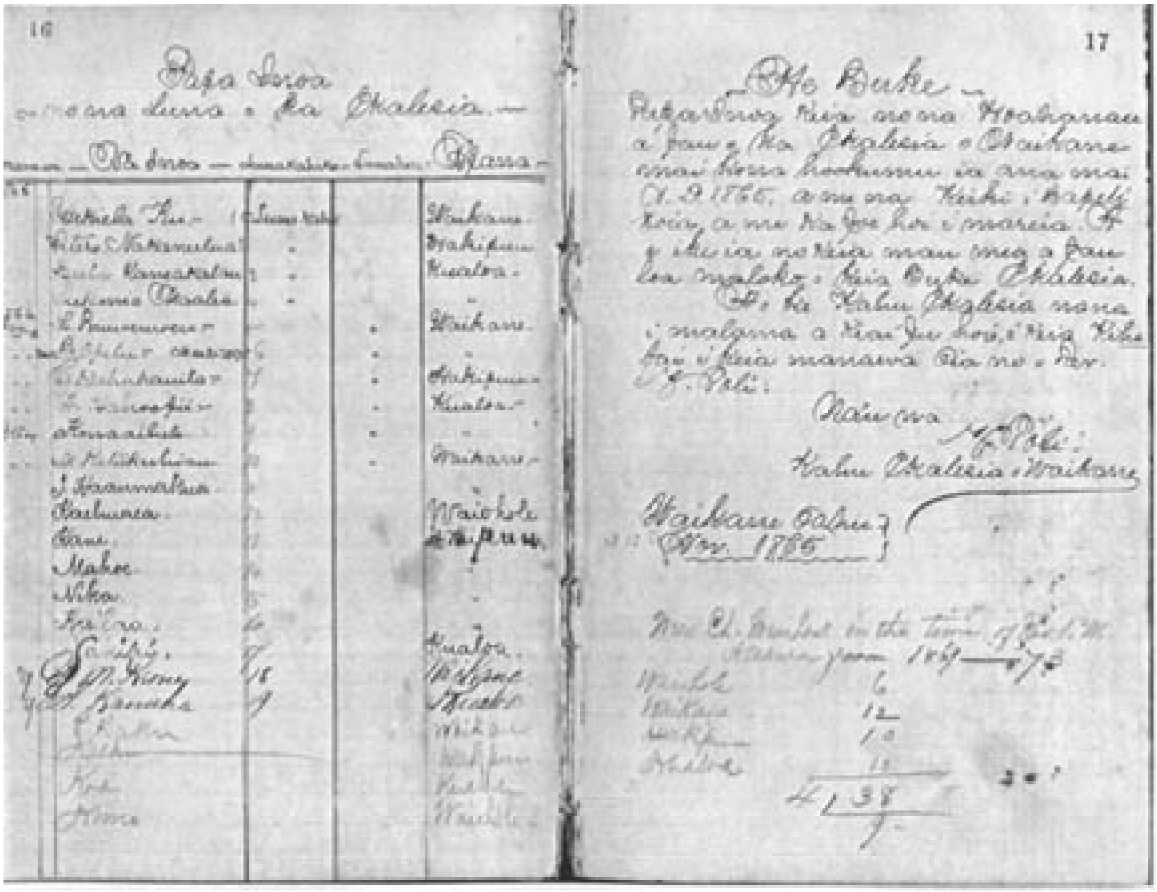
---

[5] In the context and for the purposes of this paper the terms record and archival document are used interchangeably.
[6] The difference between substantive and demonstrative evidence is that the former is admitted for its content, while the latter only for the mere fact of its existence, in support of other substantive evidence.

produces the same outcome. Repeatability, which is one of the fundamental precepts of digital forensics practice, is supported by the documentation of each and every action carried out on the digital evidence. Open source software is also the best choice for assessing accuracy, especially when conversion or migration occur, because it allows for a practical demonstration that nothing could be altered, lost, planted, or destroyed in the process.

Authenticity, to digital forensics experts, means that the data or content of the record are what they purport to be <u>and</u> were produced by or came from the source they are claimed to have been produced by or come from. Again, the term “source” is used to refer to a person (physical or juridical), a system, software, or a piece of hardware. As in the archival concept, authenticity implies integrity, but the opposite is not true, that is, integrity does not imply authenticity (because identity must also be certain). In fact, the digital forensics view of integrity is much more nuanced than the archival view, for which integrity is simply the quality of being complete and unaltered in all essential respects, a definition that equally applies to data, documents, records, copies, or records systems.

In digital forensics, integrity is distinguished in several types. The first type is data integrity, which is the fact that data are not modified either intentionally or accidentally “without proper authorization,” and is based on bitwise integrity, that is, on the fidelity not only of the bits but of their order. To clarify, let’s consider loss of fidelity in the analogue environment, where a document may fade to the point of being unreadable but maintains the same content/data:



16

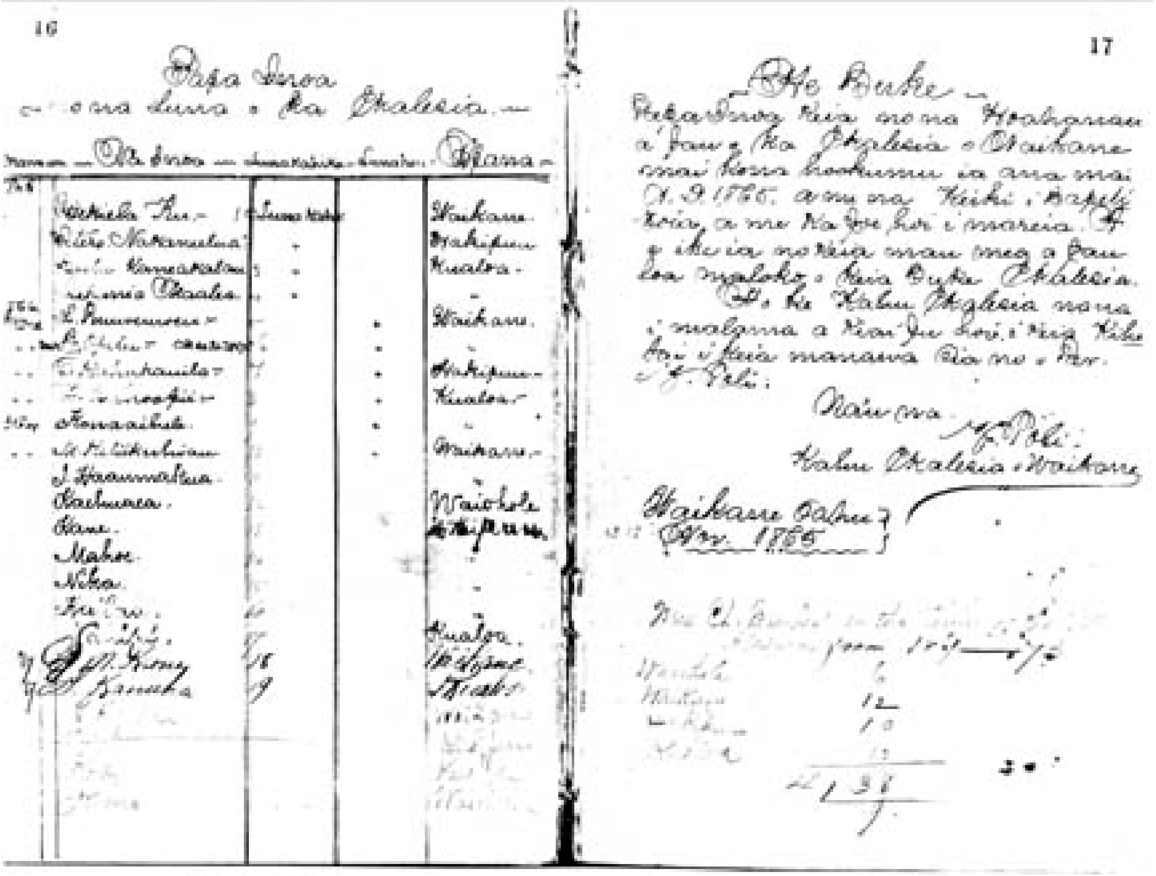
17

Ka Buke

Waikane Oahu
Nov. 1865

Waiahole 6
Waikane 12
Waikapu 10
Kualoa 10

4 | 38



16

Papa Inoa

o na lala o ka Ekalesia

Ka Inoa	[illegible]	[illegible]	Wahi
[illegible]	[illegible]		Waikane
[illegible]			Hakipuu
[illegible]			Kualoa
[illegible]			
L. [illegible]			Waikane
[illegible]			
[illegible]			Hakipuu
[illegible]			Kualoa
[illegible]			
[illegible]			Waikane
J. [illegible]			
[illegible]			Waiahole
[illegible]			Hakipuu
Mahoe			
Nika			
[illegible]	16		
[illegible]	17		Kualoa
[illegible]	18		Hakipuu
[illegible]	19		Kualoa

17

He Buke

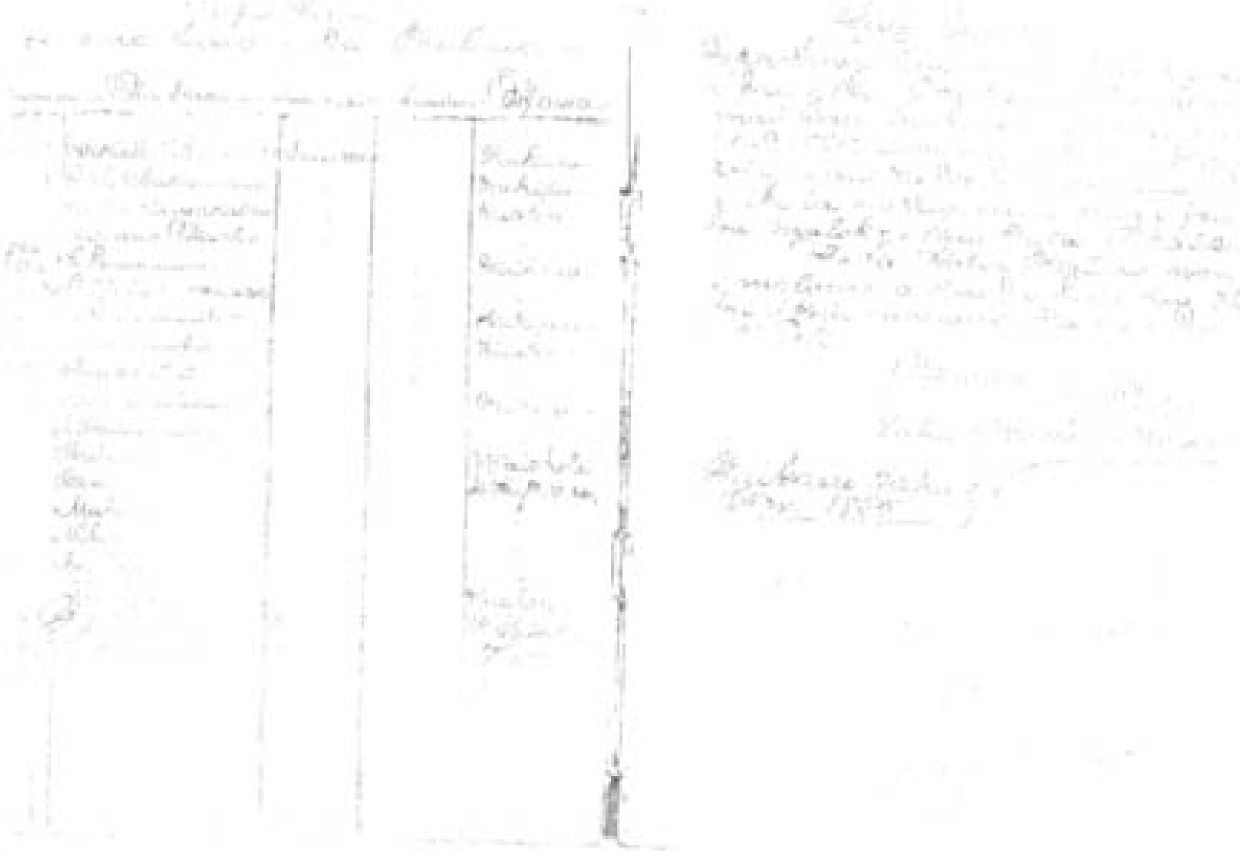
Papa Inoa keia no na Hoahanau a pau o ka Ekalesia o Waikane mai kona hookumu ia ana mai A.D. 1865. a me na Keiki i bapetizo ia, a me na Poe hoi i mareia. A e ike ia no keia mau mea a pau loa maloko o keia Buke Ekalesia.
A. he Kahu Ekalesia nona i malama a [illegible] hoi i keia [illegible] i keia manawa oia no o Rev. J. Poli;

Owau no
J. Poli
Kahu Ekalesia o Waikane

Waikane Oahu
Nov. 1865

[illegible] from 1865 — 1875

Waiahole	6
Waikane	12
Hakipuu	10
Kualoa	13
	41



In contrast, in the digital world, if the original bits are, for example, 101, the value conveyed is 5, but if we change the order to 110, the value is 6, and, if we change again to 011, the value is 3. The same bits have different value if their order changes. Thus, loss of fidelity implies different content. We as archivists are today responsible for preventing loss of data integrity. How can we do that? Intentional alteration is preventable through permission and access controls, but accidental alteration avoidance requires that additional hardware and/or software be in place. Both type of alteration require, in addition to methods for preventing them, methods of determining whether the record has been altered, maliciously or otherwise. For this we cannot rely on file size, dates or other file properties, but need audit logs and strong methods like Checksum and HASH Algorithms.

A second type of integrity digital forensics experts are concerned with is duplication integrity, that is, the fact that, 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. This type of integrity is extremely important to archivists because we can only preserve digital records by reproducing them. However, when we archivists talk about duplication, we usually refer to making "copies," while forensic experts refer to "taking images." The difference is fundamental.

A copy is a selective duplicate of files. One can only copy what one can see. Therefore copying provides an incomplete picture of the digital device. Furthermore, it rarely includes confirmation of completeness and it mostly involves moving individual files. In contrast, an image is a bit by bit reproduction of the storage medium, a full disk copy of the data on a storage device—regardless of operating system or storage technology—made prior to performing any analysis of the disk. Creating a disk image is important in forensics to ensure that disk information is not inadvertently changed, to reproduce forensic test results on the original evidence, and to capture information normally invisible to the operating system when in use (including memory, page files, boot sector, BIOS). In addition, digital forensics experts link duplication integrity to time and have considered the use of time stamps for that purpose. The reason is that every time one accesses a computer something changes, thus, not two images taken at different times—even in a close sequence—are identical.

Whether we choose a reproduction process involving copying or imaging in order to preserve digital records must depend not only on the technological advantages presented by the one or the other method, but also and foremost on archival considerations, such as those embedded in our deontological code, considering the imaging involves reproducing also deleted files. Thus, while duplication integrity, as well as all other types of integrity, is a concept we need to appropriate and use, as is the case with all concepts taken from other disciplines, we need to bring it to bear on our own discipline by adapting it to the disciplinary, scientific, ethical and social context in which we carry out our functions.

Another type of integrity is computer integrity, which means that the computer produces accurate results when used and operated properly, and that it was so employed when the evidence was generated. Very similar is the concept of system integrity, which means that the system in question would perform its intended function in an unimpaired manner, free from unauthorized manipulation, whether intentional or accidental. Both integrities imply hardware and software integrity. To be able to establish computer and system integrity one needs to verify that 1) sufficient security measures are in place to prevent unauthorized or untracked access to the computers, networks, devices, or storage, and 2) stable physical devices will maintain the value they were given until authorized to change: users/permissions, passwords, firewalls, and system logs. The latter are sets of files *automatically* created to track the actions taken, services run, or files accessed or modified, at what time, by whom and from where. They are categorized in Web logs (Client IP Address, Re quest Date/Time, Page Requested, HTTP Code, Bytes Sent, Browser Type, etc.), Access logs (User account ID, User IP address, File Descriptor, Actions taken upon record, Unbind record, Closed connection), Transaction logs (History of actions taken on a system to ensure Atomicity, Consistency, Isolation, Durability; Sequence number; Link to previous log; Transaction ID; Type; Updates, commits, aborts, completes), and Auditing logs. The latter are increasingly required by law to demonstrate the integrity of the system and, when properly configured and restricted, provide checks and balances, are able to determine effective security policies, to trap errors that occur, to provide instantaneous notification of events, to monitor many systems and devices through ‘dashboards,’ to support the determination of the accountability of people, to provide the necessary snapshot for post-event reconstruction (‘black-box’), and to answer Who-What-Where-When, but only if retained for sufficient time.

Regardless of the elements of the computer/system that are examined to verify it, computer/ system integrity can be inferred on the basis of repeatability, verifiability, objectivity and transparency. More generically, an inference of system integrity can be made if the theory, procedure or process on which the system design is based 1) has been tested or cannot be tampered with; 2) has been subjected to peer review or publication (standard); 3) its known or potential error rate is acceptable; and4) is generally accepted within the relevant scientific community.

The final type of integrity is process integrity, that is, the respect of formalized legal requirements for the collection, recovery, interpretation and presentation of digital evidence. The assessment of process integrity is based on two fundamental principles, the principle of non-interference and the principle of identifiable interference. The former means that the method used to gather and analyse [or acquire and preserve] digital data or records does not change the digital entities; the latter means that, if the method used does alter the entities, the changes are identifiable. These principles, which embody the ethical and professional stance of digital forensics experts, are consistent with the traditional impartial stance of archivists, as well as with their new responsibility of neutral third party, or trusted custodian.

If trustworthiness embodies the qualities of reliability, accuracy, and authenticity (with its sub-qualities of identity and integrity), its assessment gives origin to authentication. To archivists, authentication is a means of declaring the authenticity of a record at one particular moment in time. In the digital environment authentication is often entrusted to a digital signature. The digital signature is functionally equivalent to seals rather than to signatures in that it verifies origin (identity), certifies intactness (integrity), and makes record indisputable and incontestable (non-repudiation). However, seals are associated with a person while digital signatures are associated with a person and a record.

Also for digital forensics authentication is proof of authenticity by means of an authoritative declaration, but such declaration 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 digital forensics expert 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. In digital forensics, the strength of circumstantial digital evidence could be increased by metadata which record (1) the exact dates and times of any document sent or received, (2) which computer(s) actually created them, and (3) which computer(s) received them. Also a chain of legitimate custody is ground for inferring authenticity and authenticate a record, and so is a digital chain of custody, that is, the information preserved about the record and its changes that shows specific data was in a particular state at a given date and time. Additionally, a declaration made by an expert who bases it on the trustworthiness of the recordkeeping system and of the procedures controlling it (quality assurance) is recognized as valid authentication, and so is circumstantial evidence that a system would perform its intended function in an unimpaired manner, free from unauthorized manipulation of the system, whether intentional or accidental. Biometric identification systems and cryptography are not considered by digital forensics the most reliable means of authentication.

Clearly there are several concepts, methods and practices related to authentication of digital records that we can take from digital forensics and adapt for our archival purposes. It is important that we go through this process of identification of useful knowledge and its elaboration because we are increasingly given responsibility for extracting digital materials from obsolete hardware and software in order to access and acquire it, authenticating digital material of uncertain provenance for various purposes, documenting the technological context of digital records we preserve as we move it from a system to another, protecting the authenticity of digital materials over the long term, and responding to e-discovery requests on the part of the legal system, as we acquire records earlier in their life cycle. Certainly such an endeavour requires much research, of which the Digital Records Forensics Project carried out at the University of British Columbia[7] is only a ground breaker, but it is necessary work that will greatly enrich our discipline as well as the digital forensics one.[8]

---

[7] See www.digitalrecordsforensics.org

REFERENCES

Carrier, Brian. (2003). "Defining Digital Forensic Examination and Analysis Tool Using Abstraction Layers." *International Journal of Digital Evidence*, 1(4). Available online: http://www.utica.edu/academic/institutes/ecii/publications/articles/A04C3F91-AFBB-FC13-4A2E0F13203BA980.pdf (accessed 30 November 2010).

Casey, Eoghan. (2007). "What Does 'Forensically Sound' Really Mean?" *Digital Investigations*, 4: 49-50.

Cohen, Fred. (2011). *Digital Forensic Evidence Examination* 3rd edition.

Digital Forensics Research Workshop. (2001). online at http://www.dfrws.org/2001/dfrws-rm-final.pdf

Duranti, Luciana. (2009). "From Digital Diplomatics to Digital Records Forensics." *Archivaria*, 68: 39-66.

Duranti, Luciana and Barbara Endicott-Popovsky. (2010). "Digital Records Forensics: A New Science and Academic Program for Forensic Readiness," *Journal of Digital Forensics, Security and Law*, 5(2): 45-63.

Government of Canada. (2005). "CAN/CGSB-72.34-2005: Electronic Records as Documentary Evidence." Gatineau, Canada: Canadian General Standards Board.

Kirschenbaum, Matthew G., Richard Ovenden, and Gabriela Redwine. (2010). *Digital Forensics and Born-Digital Content in Cultural Heritage Collections.* Washington, D.C.: CLIR.

MacNeil, Heather. (2000). "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-term Preservation of Electronic Records." *Archivaria*, 50: 52-78

Mason, Stephen (ed.). (2010). *Electronic Evidence.* 2nd edition. LexisNexis Butterworths. Page 56 of 56.

Moses, Lyria Bennett. (2007). "Recurring Dilemmas: The Law's Race to Keep Up With Technological Change." *University of New South Wales Faculty of Law Research Series*, 21, available at http://www.austlii.edu.au/cgi-bin/sinodisp/au/journals/UNSWLRS/2007/21.html (accessed on July 8, 2010).

---

[8] Examples of such cross-fertilization are Fred Cohen, *Digital Forensic Evidence Examination* 3rd edition (Cohen, 2011); Matthew G. Kirschenbaum, Richard Ovenden, and Gabriela Redwine, *Digital Forensics and Born-Digital Content in Cultural Heritage Collections* (Washington, D.C.: CLIR, 2010); and Luciana Duranti and Barbara Endicott-Popovsky, "Digital Records Forensics: A New Science and Academic Program for Forensic Readiness," *Journal of Digital Forensics, Security and Law*, 5(2) (2010): 45-63.

Murray, Daniel R., Timothy J. Chorvat, and Chad E. Bell. (2008). "Taking a Byte out of Discovery: How the Properties of Electronically Stored Information Have Shaped E-Discovery Rules." *Uniform Commercial Code Law Journal*, 41(1): 35-49.

Paul, George L. (2004). "The 'Authenticity Crisis' in Real Evidence." *Practical Litigator*, 15(6): 45-52.

Paul, George. (2008). *Foundations of Digital Evidence*. Chicago: American Bar Association.

Peritz, Rudolph J. (1986). "Computer Data and Reliability: A Call for Authentication of Business Records Under the Federal Rules of Evidence." *Northwestern University Law Review*, 80(4): 956-1002.

Pollitt, Mark and Sujeet Shenoi (eds.). (2006). *Advances in Digital Forensics: IFIP International Conference on Digital Forensics*. New York: Springer.