Authenticating Digital Records: The Archivist as a Forensics Expert

Luciana Duranti
Director, InterPARES & DRF Projects

Louvain, 25 March 2011

Digital Records Forensics Project

Archival Science, Diplomatics & the Law

Archival concepts of trustworthiness are rooted in Roman Law

- Archives as a place of trusted custody
- Authenticity based on a chain of trusted custody—wax tablets
- Reliability based on antiquity and on form (Justinian Code)

Diplomatics' understanding of authenticity is the foundation of the law of evidence as we know it

Archival science and Diplomatics have served us well in the past, but...

As **digital technology has separated content and structure from form**, we can **no longer determine authenticity on the object-record**, which is **composite and permanently new**, but must make an **inference of authenticity from its environment**. For this we need help.

Archivists and Digital Forensics Experts

Archivists are increasingly called to act as forensics experts, e.g. ensuring the identity and integrity of digital records through time and attesting to it, and acquiring such records, often from obsolete systems or portable media, without altering them in the process

Digital forensic experts are called to act as archivists, e.g. identifying what digital materials fall under the definition of records, and keeping them intact for as long as needed. They are also called to

- attest to the integrity of digital systems
- provide quality assurance for digital system that produce, contain or preserve records,
- assess whether spoliation (fraudulent disposal) has occurred
- ensure that e-discovery requirements are fulfilled.

Digital Forensics

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"

Its methods are based on conceptual assumptions about records, trustworthiness, and recordkeeping

What Knowledge We Should Share

Digital forensic experts need our knowledge on

- Concepts of Archival Document (or Record) and Recordkeeping
- Concept of Trustworthiness

We need digital forensic experts' knowledge on

- Types of integrity
- Processes of access, reproduction, identification and extraction

Today I will focus on Trustworthiness and Integrity

Records Trustworthiness: Our View

Reliability: The trustworthiness of a record as a statement of fact, *based on* the competence of its author, its completeness, and the controls on its creation

Accuracy: The correctness and precision of a record's content, *based on* the above, and on the controls on content recording and transmission

Authenticity: 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: Our View

Identity: The whole of the attributes of a record that characterize it as unique, and that distinguish it from other records (e.g. date, author, addressee, subject, identifier).

Integrity: A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered (e.g. text and form fidelity, absence of technical changes).

Context: The administrative-juridical, provenancial, procedural, documentary and technological environment in which the record was created and used overtime

Digital Forensics View: Linked to Type of Documents

- **Computer Stored Documents:** Contain human statements; if created in the course of business, they are records; e.g. e-mail messages, word processing documents, etc. Used as **Substantive Evidence**
- **Computer Generated Documents**: 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. Used as **Demonstrative Evidence**
- **Computer Stored & Generated:** A combination of the two: e.g. a spreadsheet record that has received human input followed by computer processing (the mathematical operations of the spreadsheet program). Used both or either way.

Records Trustworthiness: Digital Forensics View. Reliability

Reliability: the trustworthiness of a record as to its *source*, defined in digital forensics in a way that points to either a reliable person (for computer stored documents) or a reliable software (for computer generated documents), or both.

The software should be open source, because the processes of records creation and maintenance can be 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

Records Trustworthiness: Digital Forensics View: Accuracy

A component of authenticity and, specifically, integrity. Digital entities are guaranteed accurate if they are repeatable.

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 evidence.

Open source software is again the best choice for assessing accuracy, especially when conversion or migration occurs, because it allows for a practical demonstration that nothing could be altered, lost, planted, or destroyed in the process

Records Trustworthiness: Digital Forensics View: Authenticity

The data or content of the record are what they purport to be and 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 either a person (physical or juridical), a system, software, or a piece of hardware.

Like in diplomatics/archival science, authenticity implies integrity, but the opposite is not true, that is, **integrity does not imply authenticity (as identity must also be certain)**.

Integrity: Our View

The quality of being complete and unaltered in all **essential** respects. We were never fussy about it. What if a letter had holes, or was burned on the side or the ink passed through?

The same for data, documents, records, copies, records systems

As long as it was good enough...but how good is good enough in the digital environment?

Integrity
Digital Forensics View

Data integrity: the fact that data are not modified either intentionally or accidentally "without proper authorization."

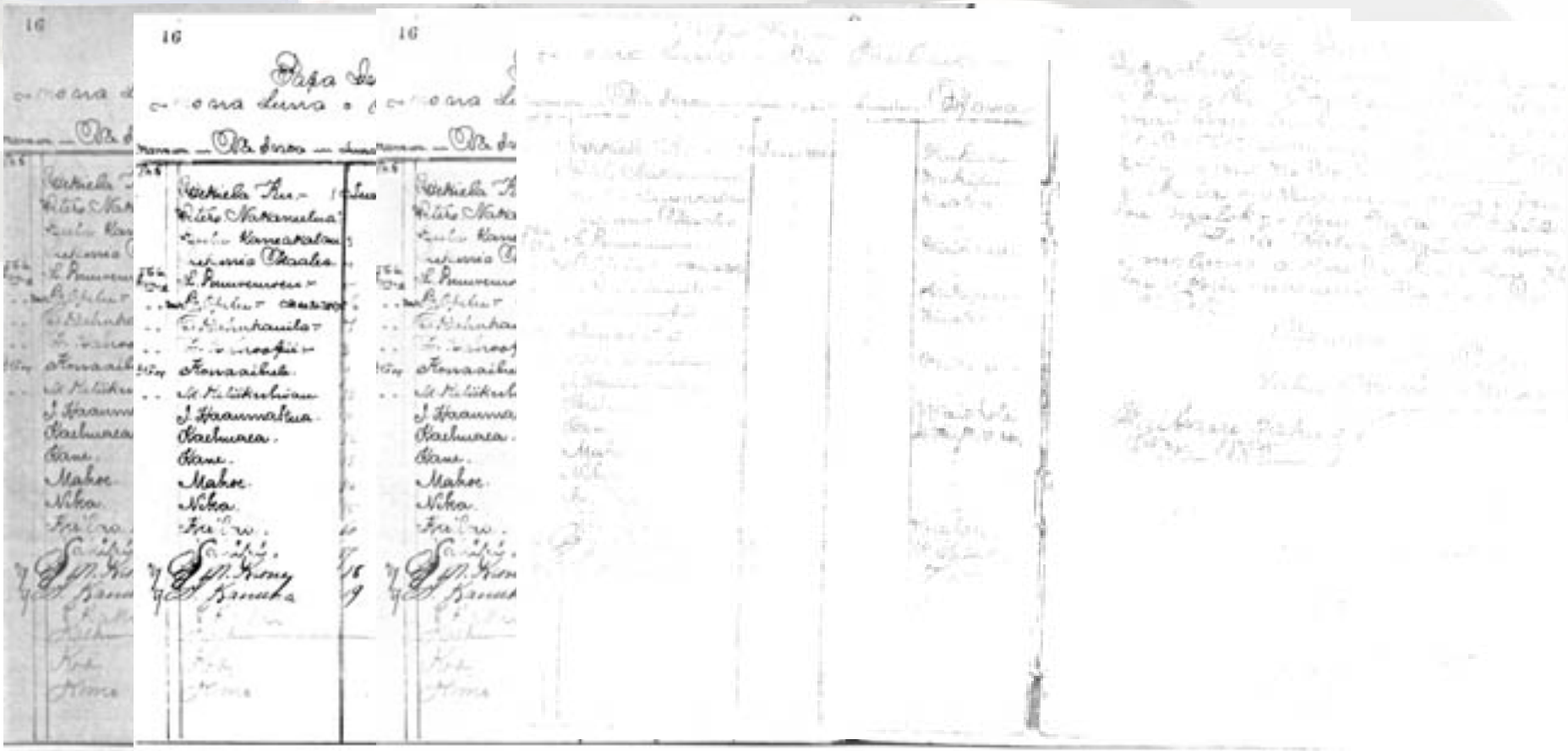
Based on **Bitwise Integrity**

Integrity
Digital Forensics View (cont.)

Bitwise Integrity

- The original bits are in a complete and unaltered state from the time of capture
- Exact and same order and value of the bits
- Small change in a bit means a very different value presented on the screen or action taken in a program or database.

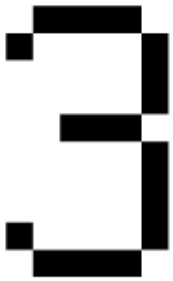
Loss of Fidelity: Analog vs. Digital



Digital Records Forensics Project

Loss of Fidelity (cont.)

- If Original Bits 101
- Change state to 110
- Continues to a 011



- Same bits, but
  Different value

Protect From Data Alteration

- Intentional alteration preventable through permission and access controls
- Accidental alteration avoidance requires that additional hardware and/or software be in place
- Requires method of determining if the record has been altered, maliciously or otherwise
- Cannot rely on file size, dates or other file properties
- We need audit logs and strong methods like Checksum and HASH Algorithms

Integrity
Digital Forensics View (cont.)

Duplication integrity: 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.

Digital forensics experts also link duplication integrity to time and have considered the use of time stamps for that purpose.

But, when we say duplicate...

Archival Duplicate: Copy

Copy: selective duplicate of files

- You can only copy what you can see
- Rarely includes confirmation of completeness
- Moved as individual files
- Provides incomplete picture of the digital device

Forensic Duplicate: Disk Image

Image: 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 forensic analysis of the disk.

Creating a disk image is important in forensics to:

- ensure that disk information is not inadvertently changed.
- reproduce forensic test results on the original evidence.
- capture information normally invisible to the operating system when in use (including memory, page files, boot sector, BIOS)

Integrity
Digital Forensics View (cont.)

Computer integrity: the computer process produces accurate results when used and operated properly and it was so employed when the evidence was generated.

System Integrity: a system would perform its intended function in an unimpaired manner, free from unauthorized manipulation whether intentional or accidental

Both imply **hardware and software integrity**

Computer or System Integrity

Inferred from:

- Sufficient security measures to prevent unauthorized or untracked access to the computers, networks, devices, or storage.
- Stable physical devices that will maintain their 'statefulness' – the value they were given is maintained until authorized to change.
  - Users/permissions
  - Passwords
  - Firewalls
  - **Logs**

System Logs and Auditing

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

- 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)

Auditing Logs

- Increasing **required by law to demonstrate integrity of the system**
- Properly configured, restricted, provide checks and balances
- Ability to determine effective security policies
- Ability to trap errors that occur
- Provide instantaneous notification of events
- Monitor many systems and devices through ‘dashboards’
- Allow to determine accountability of people
- Provide the necessary snapshot for post-event reconstruction (‘black-box’)
- Answer Who-What-Where-When, but only if retained for sufficient time (space vs. money vs. risk vs. knowledge)

Integrity
Digital Forensics View (cont.)

Process Integrity: Formalized legal requirements for the collection, recovery, interpretation and presentation of digital evidence.

Example: UK ACPO:

- No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court.
- In exceptional circumstances, where a person finds it necessary to access original data held on a computer or on storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.
- An audit trail or other record of all processes applied to computer based electronic evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
- The person in charge of the investigation (the case officer) has overall responsibility for ensuring that the law and these principles are adhered to.

Assessment of Computer/System Integrity

The assessment is based on **repeatability**, **verifiability**, **objectivity** and **transparency**

An **inference** of system integrity can be made based on the facts that:

- the theory, procedure or process on which the system design is based has been tested or cannot be tampered with
- it has been subjected to peer review or publication (standard)
- its known or potential error rate is acceptable
- it is generally accepted within the relevant scientific community

Assessment of Process Integrity

Non-interference: the method used to gather and analyse [or acquire and preserve] digital data or records does not change the digital entities

Identifiable interference: 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 the archivist, as well as with his/her new responsibility of neutral third party, of trusted custodian

Authentication: Our View

A means of declaring the authenticity of a record at one particular moment in time -- possibly without regard to other evidence of identity and integrity.

Example: the **digital signature**. Functionally equivalent seals (not signatures): verifies origin (identity); certifies intactness (integrity); makes record indisputable and incontestable (non-repudiation)

But, seals are associated with a person; digital signatures are associated with a person and a record

Authentication: The Digital Forensics View

Proof of authenticity 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.*

The strength of circumstantial digital evidence could be increased by metadata which records (1) the exact dates and times of any messages sent or received, (2) which computer(s) actually created them, and (3) which computer(s) received them.

Other Forensic Means of Authentication

A chain of legitimate custody is ground for inferring authenticity and authenticate a record.

Digital chain of custody: the information preserved about the record and its changes that shows specific data was in a particular state at a given date and time.

A declaration made by an expert who bases it on the **trustworthiness of the recordkeeping system** and of the procedures controlling it (quality assurance).

Other Forensic Means of Authentication (cont.)

Biometric identification systems and cryptography are not considered the prevalent means of authentication.

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.

Again, Why should we care?

- Extraction of digital materials from obsolete hardware and software
- Authentication of digital material of uncertain provenance
- Documentation of the technological context of records
- Protection of digital material over the long term
- E-discovery (as we acquire records earlier in their life cycle)

www.digitalrecordsforensics.org

Director, Luciana Duranti

luciana.duranti@ubc.ca

Digital Records Forensics Project