Digital Records As Evidence
Towards a Digital Records Forensics

Luciana Duranti
Director, InterPARES & DRF Projects

International Conference on Surviving the Digital Wild Frontier
3 November 2010
Singapore, Singapore
Archival Science and the Law

Records are "much better than navy yards, much more efficacious than munitions factories, as it is finer to win by reason rather than by violence, by right than by wrong“ Baldassarre Bonifacio 1632

Archival concepts are grounded in Roman Law
- Archives as a place—trusted custody
- Authenticity based on a chain of trusted custody—wax tablets
- Reliability based on antiquity and on form (Justinian Code)

Archival methods are born out of legislative acts
- Swedish Law of 1766—freedom of information act
- Decree 25 July 1793—public records belong to the people
- Decree of 1841—principle of respect des fonds

Archival science is at its heart as a legal science
The rule of law was easily circumvented, authenticity and reliability needed to be tested using scientific methods, beyond Valla’s textual criticism.

**Diplomatics** (1681), a new science studying the nature, genesis, formal characteristics, structure, transmission and legal consequences of records, gave origin to **Palaeography**, **Sigillography**, **Heraldry**, **Philology**, **Exegesis**, **Semiotics**, etc.

The **Bella Diplomatica** gave origin to the **Law of Evidence**: by mid 18th century all faculties of law in Europe taught these “forensic” disciplines.
Teaching of Forensic Disciplines

In addition to Jurists, other professionals were educated in forensic disciplines:

1811 Naples—Scuola per Archivisti
1821 Paris—Ecole des Chartes
1821 Marburg—Archivschule
1854 Vienna—School on Auxiliary Disciplines of History
1925 Roma—Scuola Speciale per Archivisti e Bibliotecari

They all taught the forensic disciplines, legislation, and the archival methods rooted in legislation.
Traditional Archivist's Education

- Diplomatics
- Palaeography
- Administrative Law and Law of evidence
- Archival Knowledge
- Archival Methods
- History of the Law
- Other auxiliary disciplines

Digital Records Forensics Project
Digital Forensics is 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.
Records Managers, Archivists and Digital Forensics Experts

Records Managers and Archivists are 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 business records, and keeping them intact for as long as needed. They are also called to attest to and sometimes provide quality assurance for digital systems that produce and/or contain records, to assess whether spoliation has occurred, to fulfill e-discovery requirements.
We Need Each Other’s Knowledge

Digital forensic experts need our knowledge on

• Concepts of Record and Recordkeeping
• Records Trustworthiness

We need digital forensic experts’ knowledge on

• Types of integrity
• Processes of access, reproduction, identification and extraction
The issue of **what is a record in the digital environment** keeps coming up at trials and in political discussions.

- **British Columbia Rail case:** the judge pointed out that legislation speaks of preserving “records,” and the Liberal MLA Ralph Sultan asked “What is the definition of a record?” referring “to the controversy over to what extent e-mails qualify”

- The Supreme Court of Canada is deciding whether hyperlinks in a text are akin to footnotes or make of the material to which they connect the reader a component of the document being read
Record: Legal and Archival Views

- A document made or received in the usual and ordinary course of business and kept for the purposes of such business at a time close to the fact at issue by a person responsible for doing so (laws of evidence)
- A document made or received in the course of activity as a by-product of or instrument for it and kept for action or reference (diplomatics/archival science)
Digital Record: the Computer Science View

- Any document that is exclusively machine readable.
- Comprised of Binary Data – stored in Base2
- One value is a bit – short for Binary digit

0 = 0
1 = 1
2 = 10
3 = 11
4 = 100
5 = 101
6 = ???
7 = 110
Bits and Bytes

- Bits bundled into 8-bit collections
- 256 values ranging from 0 to 255
  
  0 = 00000000
  1 = 00000001
  2 = 00000010
  255 = 11111110
  256 = 11111111

- ASCII characters given value 1-127

- Example:
  
  Four and seven
  70 111 117 114 32 97 110 100 32 115 101 118 101 110 32 = 00100000
### Big Bytes

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbr</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilo</td>
<td>K</td>
<td>$2^{10} = 1,024$</td>
</tr>
<tr>
<td>Mega</td>
<td>M</td>
<td>$2^{20} = 1,048,576$</td>
</tr>
<tr>
<td>Giga</td>
<td>G</td>
<td>$2^{30} = 1,073,741,824$</td>
</tr>
<tr>
<td>Tera</td>
<td>T</td>
<td>$2^{40} = 1,099,511,627,776$</td>
</tr>
<tr>
<td>Peta</td>
<td>P</td>
<td>$2^{50} = 1,125,899,906,842,624$</td>
</tr>
<tr>
<td>Exa</td>
<td>E</td>
<td>$2^{60} = 1,152,921,504,606,846,976$</td>
</tr>
<tr>
<td>Zetta</td>
<td>Z</td>
<td>$2^{70} = 1,180,591,620,717,411,303,424$</td>
</tr>
<tr>
<td>Yotta</td>
<td>Y</td>
<td>$2^{80} = 1,208,925,819,614,629,174,706,176$</td>
</tr>
</tbody>
</table>
Digital Record: Our View

- **Act**: an action in which the record participates or which the record supports
- **Persons Concurring to Its Creation**: author, writer, **originator**, addressee, and creator
- **Archival Bond**: explicit linkages to other records inside or outside the system through **metadata** (e.g. classification code)
- **Identifiable Contexts**: juridical-administrative, provenancial, procedural, documentary, **technological**
- **Medium**: necessary part of the technological context, not of the record
- **Fixed Form and Stable Content**
Digital Record: Digital Forensics View

- A combination of the legal view (i.e. a document made or received in the usual and ordinary course of business) and the computer science view (i.e. Bits and bytes)
- A view that is problematic in relation to the hearsay rule: in common law, documents are hearsay because they contain human statements made outside the court—if they are records, they fall under the business records exception to the rule
- It identifies three types of documents: only two can be records
Digital Record: Digital Forensics View (cont.)

- **Computer Stored Documents**: Contain human statements and are considered hearsay; if created in the course of business, they are records (testable for truthfulness and accuracy under the business records exception to the hearsay rule): e.g. e-mail messages, word processing documents, etc.

- **Computer Generated Documents**: Do not contain human statements, but are the output of a computer program designed to process input following a defined algorithm (testable for authenticity, based on the functioning of the computer program): e.g. server log-in records from Internet service providers, ATM records.

- **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).

Substantive Evidence vs Demonstrative Evidence
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, integrity and on the reliability of the 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).
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 or a reliable software.

This would be an open source software, 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 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 diplomacy, authenticity implies integrity, but the opposite is not true, that is, integrity does not imply authenticity.
Integrity: Our View

The quality of being complete and unaltered in all essential respects. With identity, a component of authenticity

The same for data, documents, records, copies, systems
Data integrity: the fact that data are not modified either intentionally or accidentally “without proper authorization.”

Based on Bitwise Integrity
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
Loss of Fidelity (cont.)

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

- Same bits, but Different value
Data Alteration

- Intentional alteration preventable through permission and access controls
- Accidental alteration avoidance requires that additional hardware and/or software be in place
Data Alteration (cont.)

• Requires method of determining if the record has been altered, maliciously or otherwise

• Cannot rely on file size, dates or other file properties

• Requires audit logs and strong methods
Checksum

- Form of data authentication
  - If checksum doesn’t match, data corrupted or incomplete
- Add up value of bits in a packet
  - If less than 255 actual value used
  - More than 255 value after divided by 256
- Example:

<table>
<thead>
<tr>
<th>Byte1</th>
<th>Byte2</th>
<th>Byte3</th>
<th>Byte4</th>
<th>Byte5</th>
<th>Byte6</th>
<th>Byte7</th>
<th>Byte8</th>
<th>Total</th>
<th>Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>212</td>
<td>232</td>
<td>54</td>
<td>135</td>
<td>244</td>
<td>15</td>
<td>179</td>
<td>80</td>
<td>1151</td>
<td>127</td>
</tr>
</tbody>
</table>

1,151 / 256 = 4.496 (round to 4)
4 x 256 = 1,024
1,151 - 1,024 = 127
HASH Algorithms

• Computed from the base number using an algorithm
• Nearly impossible to derive without original data
• Typically use 128bit or greater algorithms, that’s $2^{128}$
  – Like trying to find a grain of sand in the Sahara
• Example:

<table>
<thead>
<tr>
<th>Input</th>
<th>Hash</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,667</td>
<td>Input x143</td>
<td>1,525,381</td>
</tr>
</tbody>
</table>
HASH Values

- Compresses bits of a message into a fixed-size value
- Extremely difficult to come up with original record based on hash value
- Common Hash functions
  - SHA-1 160 bit
  - RIPEMD-160 160 bit
  - MD5 – 128 bit
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.
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.
  - Data
  - Users/permissions
  - Passwords
  - Logs
  - Firewalls
System Logs and Auditing

Set 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
– Access logs
– Transaction logs
Typical Web Log

- Client IP Address
- Request Date/Time
- Page Requested
- HTTP Code
- Bytes Sent
- Browser Type
- OS Type
- Referrer
Typical Access Log

- User account ID
- User IP address
- File Descriptor
- Bind record results
- Actions taken upon record
- Unbind record
- Closed connection
Typical Transaction Log

- History of actions taken on a system to ensure ACID over crashes (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’
Auditing Logs (cont.)

• Ability to determine accountability of people, resources for measure events
• Provide the necessary snapshot for post-event reconstruction (‘black-box’)
• Answers Who-What-Where-When
• Only if retained for sufficient time (space vs. money vs. risk vs. knowledge)
Digital Forensics Integrity

**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 Integrity
Digital Forensics View

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

Inference of system integrity derives from the fact 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 Integrity
Digital Forensics View (cont.)

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 to medieval seals (not signatures): verifies origin (identity); certifies intactness (integrity); makes record indisputable and incontestable (non-repudiation)

But, medieval seals were 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 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.

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 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 Means of Authentication (cont.)

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

Inference of system integrity: 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.
Other Relevant Concepts

- **Chain of Custody vs. Chain of Documentation** (the conditions under which the evidence is gathered, the identity of all evidence handlers, duration of evidence custody, security conditions while handling or storing the evidence, and the manner in which evidence is transferred to subsequent custodians each time such a transfer occurs)
- **Prevention vs. Preparation** (for detection and response)
- **Identification and Acquisition vs. Search and Seizure**
- **Cryptography vs. Steganography** (a covert form of information hiding)
- **Copy vs. Image**
**Forensic View: 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)
Forensic View: 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
Why should we care?

In a records management context:

- **Discovery**: the compulsory disclosure of pertinent facts or documents to the opposing party in a civil action, usually before a trial begins. The discovery process is the process of identifying, preserving, collecting, reviewing, analyzing and producing information during legal actions.

- **E-discovery**: the extension of the discovery process to information stored electronically (ESI), including email, instant messages, word processing files, spreadsheets, social networking content, and anything else stored on desktops, laptops, file servers, mainframes, smartphones, employees’ home computers or on a variety of other platforms.

- **Information Assurance and Cybersecurity**
Why should we care? (cont.)

In an historical archives context:

- 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
Digital Records Forensics

Archival science

Records management

Diplomastics

Digital Forensics

Law of evidence

Information Assurance and Cybersecurity

Digital Records Forensics Project
www.digitalrecordsforensics.org

Director, Luciana Duranti
luciana.duranti@ubc.ca