Trust & Authenticity in the Digital Environment:
An Increasingly Cloudy Issue

Luciana Duranti
University of British Columbia
Director, InterPARES & DRF Projects

San Bernardo, Chile 11 April 2012

Digital Records Forensics Project

"*Whatever* matters to human beings, trust is the atmosphere in which it thrives"

Sissela Bok

Trust & Its Rules

Trust involves acting without the knowledge needed to act. It consists of substituting the information that one does not have with other information.

The rules of trust refer to those who give trust as well as to those who receive trust:

trusters [givers] and *trustees* [receivers]

The trust-bond between trusters and trustees is usually based on four

characteristics of the trustees

Characteristics of Trustees

- *reputation*, which results from an evaluation of the trustee’s past actions and conduct;
- *performance*, which is the relationship between the trustee’s present actions and the conduct required to fulfill his or her current responsibilities as specified by the truster;
- con*fidence*, which is an assur-ance of expectation of action and conduct the truster has in the trustee; and
- compe-*tence*, which consists of having the knowledge, skills, talents, and traits required to be able to perform a task to any given standard

Sztompka P (1999) *Trust*. Cambridge University Press, Cambridge

Trust & Authenticity

- In the digital environment authenticity is an inference based on foundation evidence and, in some measure, on *confidence* in the *performance* and *competence* of the keeper of the material, based on its *reputation*.
- The level of trust required is proportional to the sensitivity of the material to be trusted as authentic and the adverse consequences of its lack or loss of trustworthiness.
- To guarantee the authenticity of **digital records** requires intentional action or intervention by trusted entities imbued with accountability, but also an adequate framework of policies, procedures, and technologies. This has always been the case.

Digital vs. Traditional Records

In the digital environment:

- Content, structure and form are no longer inextricably linked
- The stored entity is distinct from its manifestation and its digital presentation has to be considered as well as its documentary one
- When we save a record, we take it apart in its digital components, and when we retrieve it, we reproduce it (it is not possible to preserve a digital record, only the ability to reproduce or recreate it)

Therefore, we can no longer determine authenticity on the object-record, which is composite (stored + manifested) and permanently new (re-production), but must make an **inference of authenticity from its environment of creation, maintenance & use and preservation**.

Types of Digital 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** (of its content)
- **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** (of the action from which they result)
- **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.

Trustworthy Record:
More Than An Authentic Record

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

Reliability

Reliability: the *source* of the record is the key, defined in a way that points primarily to a reliable person and procedure (for computer stored documents) or a reliable process and 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 the process or system used to produce a result or
- by showing that the process or system produces an accurate result

Accuracy

Digital entities are guaranteed accurate if they are repeatable.

Repeatability, which is one of the fundamental precepts of digital forensics, is supported by the documentation of each and every action carried out on the record.

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

Authenticity

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

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).

Integrity

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 definition used with respect to data, documents, records, copies, records systems

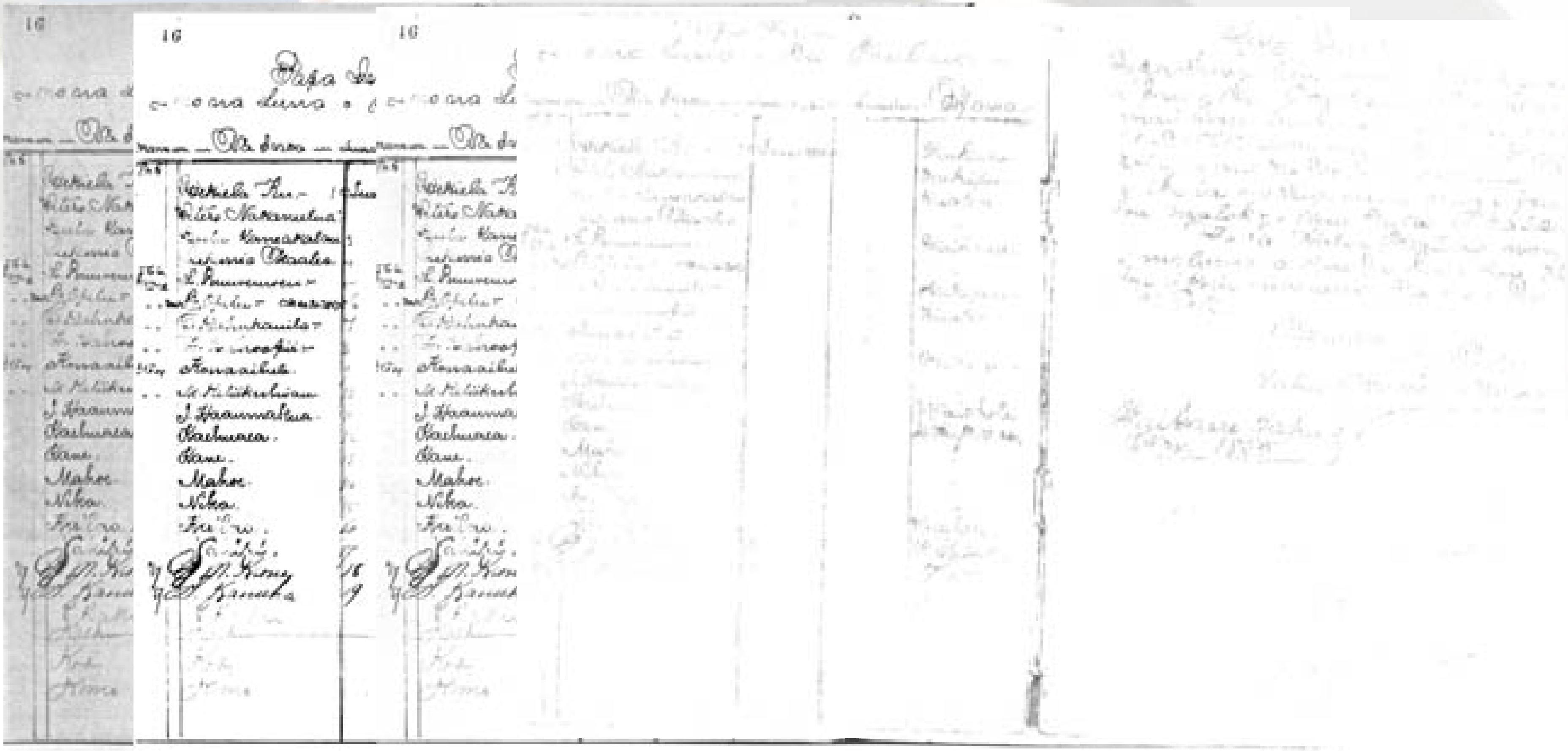
As long as it was good enough...but how good is good enough in the digital environment?

Data Integrity

Based on **Bitwise Integrity**: the fact that data are not modified either intentionally or accidentally "without proper authorization."

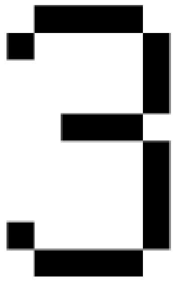
- The original bits are in a complete and unaltered state from the time of capture, that is, they have the exact and same order and value
- Small change in a bit means a very different value presented on the screen or action taken in a program or database.

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

Protecting Records 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

Duplication Integrity

The fact that, given a data set, the process of creating a duplicate of the data does not modify the data, and the duplicate is an exact bit copy of the original data set. Time stamps are useful to support it.

Disk Image: a bit by bit reproduction of the storage medium. A full disk copy of the data on a storage device

Different from a copy: a 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

Issues linked to images: deleted files? Is the image a record?

Computer and System Integrity

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 performs its intended functions in an unimpaired manner, free from unauthorized manipulation whether intentional or accidental, and it did so when the evidence was generated and used.

Both imply **hardware and software integrity**

Computer or System Integrity

Protected by:

- 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)

Assessment of Computer/System Integrity

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

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

- the theory, procedure or process on which the 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

Process Integrity

Non-interference: the method used to gather, capture, use, manage and preserve digital data or records does not change the digital entities

Identifiable interference: the method used does alter the entities, but the changes are identifiable

These principles, which embody the ethical and professional stance of records and information managers, archivists, and digital forensics experts, are consistent with the impartial stance of a neutral third party, a trusted custodian

Authentication

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 seals (not to signatures): verifies record's origin (identity); certifies record's 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. They are not a preferred means of authentication through time: they are preferred only across space.

Preferred 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 (**Information governance and quality assurance**).

So, In Whom Shall We Trust?

Shall we Trust in Cloud Computing Providers?

Definition of Cloud Computing:

“Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

National Institute of Standards and Technology

In other words…

The Cloud is bringing us back to the 60s, when we had dummy terminals and a mainframe...but in this case it isn’t the company mainframe, it is providers like Google, Microsoft or Amazon in whom we trust!

Let’s look at the situation as it stands now

What We Use the Cloud For

- Communication (e-mail #1 use)
- Backup
- Collaboration
- Distribution
- Recordkeeping
- Long-term storage

Reliability

- ✓ We have no real control on the processes of records creation and maintenance on the cloud.
- ✓ No control over whom we share our cloud with.
- ✓ Terms of service or policy may change.
- ✓ Backup may be done without us knowing and may not be disposed of as needed
- ✓ Records might be deleted without us knowing or may not be deleted according to the retention schedule.
- ✓ Audit usually is not allowed

Authenticity

- Chain of custody is not demonstrable
- Authenticity cannot be inferred from circumstantial evidence
- Tampering is possible in the cloud, so authenticity must always be verified. How?
- Do records in the cloud have bitwise and duplication integrity?
- Do systems and processes have integrity (repeatability, verifiability, objectivity, transparency)?
- Can records made, kept and used in the cloud be admissible as evidence (best evidence, authentication, hearsay exception)?
- What happens when cloud hardware/software become obsolete (conversion, migration)?

Authenticity (cont.)

- ✓ The cloud is targeted by hackers more than any records system. Would we be told when it happens?
- ✓ Tasks are given to sub contractors. Would we be told?
- ✓ Records can be stored anywhere and moved at any time
- ✓ Encryption might not be done-in transit or in the cloud.
- ✓ Shared server could intermingle information.
- ✓ FBI seized servers at Dallas Data Center for 1 person-50 businesses used it, and it took days to get servers back up.
- ✓ Can’t always move or remove records (e.g. for transfer to archives).

Legal Risks

- Geographic location of information-jurisdiction issues (Patriot Act-FBI gets court order under Section 215; Privacy Act; HIPAA--US health information, and Gramm-Leach Bliley Act--US financial information, have to be protected through specific agreement).
- Trade secrets-are they still secret in the cloud?
- Legal privilege-is it still applicable if the cloud can access it?
- Can we isolate records for legal hold—e-discovery issue?
- Are records preserved properly? Can they be accessed? Are there multiple copies in different locations, and which is used?

Conclusion

As Leslie Johnson has stated in the Signal (LoC):

“We can’t be afraid of cloud computing. Given the volumes of data coming our way and mounting researcher demands for access to vast quantities of data, the cloud is the only feasible mechanism for storing and providing access to the materials that will come our way. We need to focus on developing authentication, preservation and other tools that enable us to keep records in the cloud.”

In other words, when it comes to digital records,

trust is best given to a good contract!

www.interpares.org
www.digitalrecordsforensics.org

Director, Luciana Duranti

luciana.duranti@ubc.ca