From Classic Diplomatics to Digital Diplomatics

Luciana Duranti
The University of British Columbia
Oslo, 26 September 2013

Archival Concepts

Archival concepts are grounded in Roman Law

- Records preserve perpetual memory of the acts which they carry out or attest to
- Deposit of the records in a public place guarantees the continuing trustworthiness of records as witnesses of action
- Antiquity provides records with the highest authority
- Unbroken legitimate custody ensures records authenticity
- Authentication is based on record form

(Justinian Code A.D. 565)

Archival Methods

Archival methods are embedded in legislative acts

- Swedish Law of 1766—freedom of information act
- French Decree 25 July 1793—public records belong to the people and must be kept for the people
- French Decree of 1841—principles of provenance and respect for original order

Archival knowledge was codified as a modern science in 1632 by Baldassarre Bonifacio. In his words: "documents 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"

Diplomatics

The rule of law was easily circumvented: the trustworthiness of records needed to be tested using scientific methods.

De Re Diplomatica (1681) , Dom Jean Mabillon

Trustworthiness is assessed on the basis of the process of formation of documents, and on their formal characteristics, structure, and transmission through time and space.

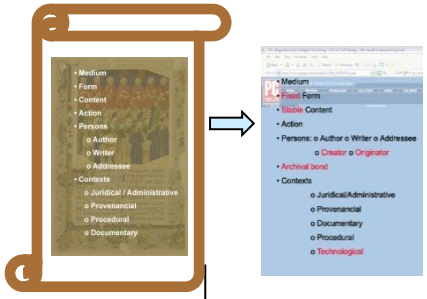
The **Bella Diplomatica** (judicial disputes on authenticity of documents based on diplomatic concepts gave origin to the **Law of Evidence**

By mid 18th century all faculties of law in Europe taught archival science and diplomatics as “forensic” disciplines

Archival Diplomatics of Digital Records

Dr. Luciana Duranti
The University of British Columbia

The Concept of Record



Digital Record Characteristics

On the face Of the Record

Formal Elements

Attributes

Digital Components

Lifecycle of Digital Records
Phase 1: Records of the creator
Phase 2: Authentic copies of the records of the creator

Archival Diplomatics

The integration of archival and diplomatic theory about the genesis, inner constitution, and transmission of documents; and about their relationship with the facts represented in them, and with other documents produced in the course of the same function and activities, and with their creators.

Retrospective Use

Prospective Use

Dynamic and Interactive Records
Stable Content
Fixed Documentary Form
Bounded Variability

Genesis of the Digital Records Workflow: actio et conscriptio

Initiative

Inquiry

Consultation

Deliberation

Deliberation Control

Execution

Functions of Records
Probative/Dispositive
Supporting/Narrative
Instructive/Enabling

Categories of Records
- Manifested:
- Stored:

} Form, Content, and Composition Data

Application: Research Projects
UBC Project (1994 – 1997)
InterPARES 1 (1999 – 2001)
InterPARES 2 (2002 – 2006)
InterPARES 3 (in application)

Status of Transmission
Draft
Original
Authenticated original
Copy (e.g., authentic copy)

The Concept of Trustworthiness

Reliability

The trustworthiness of a record as a statement of fact. It exists when a record can stand for the fact it is about.

Accuracy

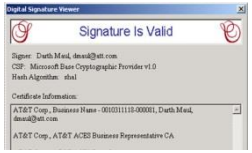
The degree to which data, information, documents or records are precise, correct, truthful, free of error or distortion, or pertinent to the matter.

Authenticity

- identity
- integrity

The trustworthiness of a record as a record; i.e., the quality of a record that is what it purports to be and that is free from tampering or corrupt

Digital Signature



✓As a Means of Authentication

Metadata
Identity Metadata
Integrity Metadata

Authentication:
A means of declaring the authenticity of a record at one particular moment in time

Luciana Duranti
Email:
luciana@interchange.ubc.ca
www.interpares.org

The Concept of Record

- **Record**: any document made or received by a physical or juridical person in the course of activity as an instrument and by-product of it, and kept for action or reference
- **Document**: recorded information (i.e., information affixed to a medium in an objectified and syntactic form)
- **Information**: "intelligence given," or a message intended for communication across time and space
- **Data**: the smallest meaningful piece of information

Digital Record Components

- **Act**: an action in which the records participates or which the record supports (naturalness and impartiality)
- **Persons Concurring to Its Creation**: author, writer, originator, addressee, and creator (authenticity)
- **Archival Bond**: explicit linkages to other records inside or outside the system (interrelatedness)
- **Identifiable Contexts**: juridical-administrative, provenancial, procedural, documentary, technological (uniqueness)
- **Medium**: necessary part of the technological context, not of the record
- **Fixed Form and Stable Content**

Fixed Form

- An entity has 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 (different digital presentation: Word to .pdf)
- An entity has fixed form also if the same content can be presented on the screen in several different ways in **a limited series of possibilities:** we have a different documentary presentation of the same stored record having stable content and fixed form (e.g. statistical data viewed as a pie chart, a bar chart, or a table)

Stable Content

- An entity has stable content if the data and the message it conveys are **unchanged and unchangeable**, meaning that data cannot be overwritten, altered, deleted or added to

- **Bounded Variability**: when changes to the documentary presentation of a determined stable content are limited and controlled by fixed rules, so that the same query or interaction always generates the same result, and we have different views of different subsets of content, due to the intention of the author or to different operating systems or applications

Digital Record Characteristics (cont.)

- **Formal Elements**: constituent parts of the record documentary form as shown on its face, e.g. address, salutation, preamble, complimentary close
- **Metadata**: the attributes of the records that demonstrate its identity and integrity
- **Digital Components**: stored digital entities that either contain one or more records or are contained in the record and require a specific preservation measure

Stored and Manifested Records

- **Stored record**: it is constituted of the digital component(s) used in re-producing it, which comprise the data to be processed in order to manifest the record (content data and form data) and the rules for processing the data, including those enabling variations (composition data)—e.g. instructive and enabling records

- **Manifested record**: the visualization or instantiation of the record in a form suitable for presentation to a person or a system. Sometimes, it does not have a corresponding stored record, but it is re-created from fixed content data when a user's action associates them with specific form data and composition data (e.g. a record produced from a relational database)

Types of Digital Records

Static: They do not provide possibilities for changing their manifest content or form beyond opening, closing and navigating: e-mail, reports, sound recordings, motion video, snapshots of web pages

Interactive: They present variable content, form, or both, and the rules governing the content and form of presentation may be either fixed or variable

Interactive Entities

- **Non-dynamic**: 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. Ex. Interactive web pages, online catalogs, records enabling performances—**they are records**
- **Dynamic**: the rules governing the presentation of content and form may vary—**they are either information systems or potential records**

Records Functions
(the way a record relates to an action)

- ***Ad substantiam*** (dispositive, e.g., contracts)
- ***Ad probationem*** (probative, e.g., registries)
- **Supporting**: generated to be used in the course of activity (ies) as a source of information, often by multiple users (e.g., GIS)
- **Narrative**: generated on a purely discretionary basis only as a means of communication (e.g., most e-mails, memos, web sites)

Records Functions

- **Instructive**: provide guidance on the way in which external data or documents are to be presented (e.g., scores, scripts, regulations, manuals of procedure, instructions for filling out forms)
- **Enabling**: enable the performance of artworks (software patches), the execution of business transactions (interacting business applications), the conduct of experiments (a workflow generated and used to carry out an experiment of which it is instrument, byproduct and residue), the analysis of observational data (interpreting software), etc. Most of them are stored only records.

Interactive Information Systems

- Entities produced in dynamic computing applications that select different sets of rules to produce documents, depending on user input, sources of content data, and characteristic of content (weather sites)
- Entities produced by evolutionary computing where the software generating them can change autonomously (scheduling and modeling of financial markets; edutainment sites)

Interactive Potential Records

- Entities where the variation is due to data that change frequently, because the design permits updating, replacement or alterations; allows data collection from users or about user interactions or actions; or uses these data to determine subsequent presentation
- Entities where the variation is due to data received from external sources and not stored within the system

They are presently not records but should be made into records if they fulfill one of the records functions.

Trustworthiness

Reliability

The trustworthiness of a record as a statement of fact,

based on:

- the competence of its author
- the controls on its creation

Accuracy

The correctness and precision of a record's content

based on:

- the competence of its author
- 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:

- identity
- Integrity
- reliability of the system

Authenticity: Identity

The whole of the attributes of a record that characterize it as unique, and that distinguish it from other records.

Identity metadata:

- names of the persons concurring in its creation
- date(s) and time(s) of issuing, creation and transmission
- the matter or action in which it participates
- the expression of its archival bond
- documentary form
- digital presentation
- the indication of any attachment(s)
- digital signature
- name of the person responsible for the business matter

Authenticity: Integrity

A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered.

Integrity metadata:

- name(s) of handling persons over time
- name of person responsible for keeping the record
- indication of annotations
- indication of technical changes
- indication of presence or removal of digital signature
- time of planned removal from the system
- time of transfer to a custodian
- time of planned deletion
- existence and location of duplicates outside the system

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

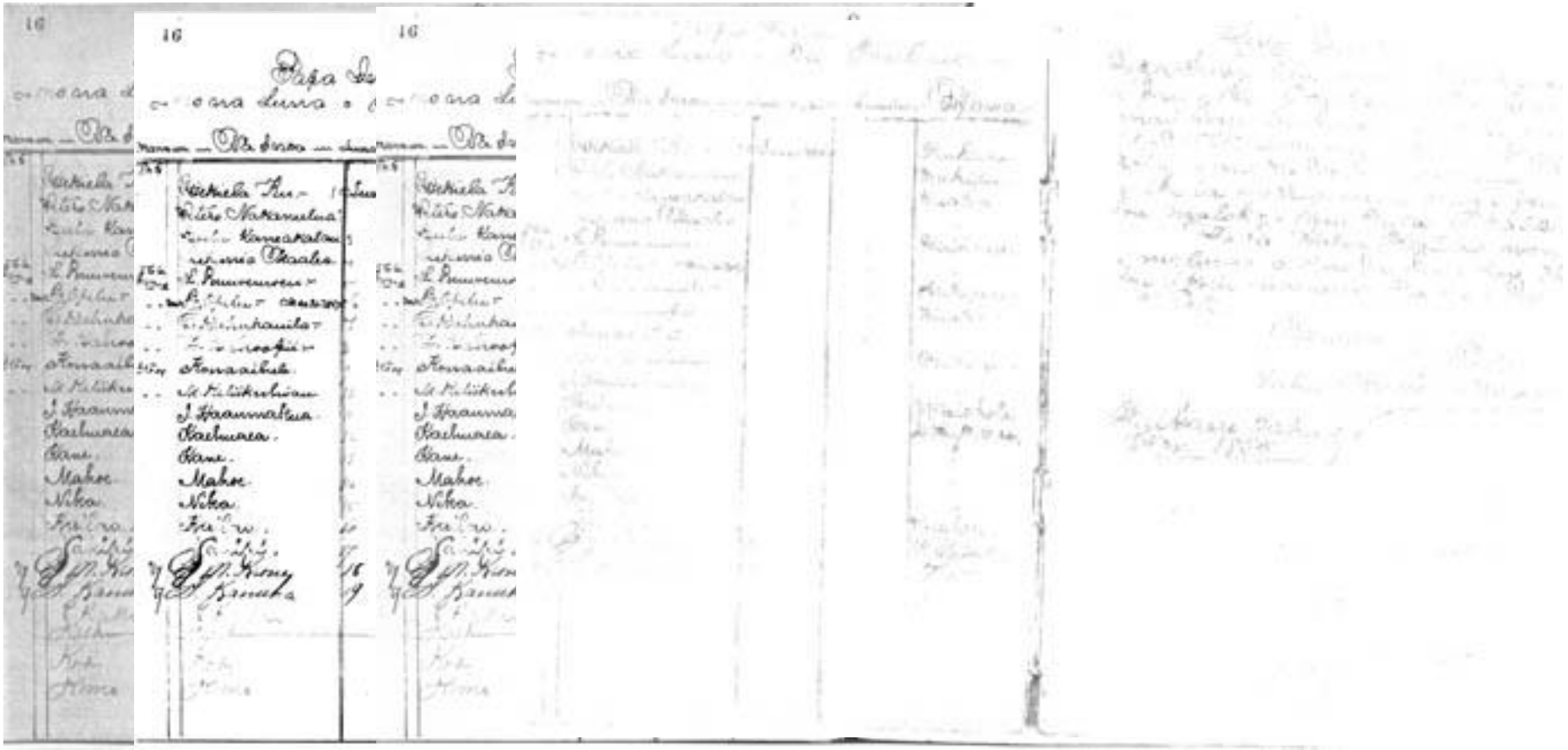
As long as it was good enough...but how good is good enough in the digital environment?

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.

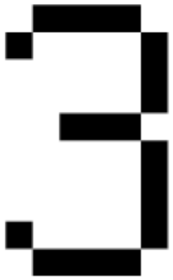
Loss of Integrity: Analog vs. Digital



Loss of Integrity (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 and strong methods like Checksum and HASH Algorithms
- Accidental alteration avoidance requires that additional hardware and/or software be in place
- We also need methods of determining if the record has been altered, maliciously or otherwise
- Cannot rely on file size, dates or other file properties
- We need logs: 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

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, of the empty spaces and the deleted files

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

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.
  - Users/permissions
  - Passwords
  - Firewalls
- **System and Auditing Logs:** 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 (Who-What-Where-When/the black-box)

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 medieval seals (not signatures):

- verifies origin (identity)
- certifies intactness (integrity)
- makes record indisputable and incontestable (non-repudiation)

The analogy is not perfect, because the medieval seal was associated exclusively with a person, while the digital signature is associated with a given person and a specific record, and because the former is an expression of authority, while the latter is only a mathematical expression

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

The Daubert rules:

- the theory, procedure or process for making or keeping the record has been tested or cannot be tampered with
- it has been subjected to peer review or publication
- the known or potential error rate is acceptable
- it is generally accepted within the relevant scientific community

Repeatability, verifiability, objectivity and **transparency**

Trusted Systems

Rules, and tools and methods to implement rules, for

Making reliable and accurate records

- record-identity metadata schemes
- business and documentary procedures integrated in a workflow structure linked to classification schemes and filing plans
- specifications of record forms
- record-making access privileges

Maintaining and keeping authentic records

- record-integrity metadata schemes
- classification schemes and filing plans
- linked retention schedule
- registration system
- retrieval system
- record-keeping access privileges

www.interpares.org

www.ciscra.org

Director, Luciana Duranti

luciana.duranti@ubc.ca