InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Canada

The Trustworthiness of Digital Records

Luciana Duranti

International Congress on Digital Records Preservation

Beijing, China
16 April 2010

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 Characteristics

- **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
- **Archival Bond**: explicit linkages to other records inside or outside the system (uniqueness)
- **Identifiable Contexts**: juridical-administrative, provenancial, procedural, documentary, technological (interrelatedness)
- **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)

- **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**

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 presentations (e.g. Land Registry)
- Entities where the variation is due to data received from external sources and not stored within the system (e.g. GIS)

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

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

- **Dispositive**, e.g., contracts
- **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.

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**

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

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

Digital Forensics

The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events, or helping to anticipate unauthorized actions

Digital Forensics View

- **Computer Stored:** They contain human statements and are considered hearsay (tested for truthfulness and accuracy under the business records exception to the hearsay rule): e.g. e-mail messages, word processing documents, and Internet chat room messages.
- **Computer Generated**: They do not contain human statements, but they are the output of a computer program designed to process input following a defined algorithm (tested for authenticity on the basis of the functioning of the computer program): e.g. server log-in records from Internet service providers, ATM records.
- **Computer Stored & Generated:** e.g. a spreadsheet record that has received human input followed by computer processing (the mathematical operations of the spreadsheet program).

Digital Forensics View (cont.)

The rules used for assessing scientific and technical evidence (not substantive but demonstrative evidence—intention vs. capability) are:

- 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

The forensic experts look for **repeatability, verifiability, objectivity** and **transparency** in the system (proof of authenticity by inference)

Digital Forensics View (cont.)

- **Authenticity** implies integrity
- To prove integrity of the records the computer process or system must produce **accurate results** when used and operated properly and it must be possible to prove that **it was so employed** when the records were generated
- The forensic experts look for **repeatability**, **verifiability**, **objectivity** and **transparency** in the system (open source software can be tested for those, not the proprietary software)
- **Integrity** however does not imply authenticity

Digital Forensics View (cont.)

What else id needed to infer authenticity?

- **Native format or repeatability of conversion**
- The **exact dates and times** of record creation
- **Data integrity**, inferred on proper authorization to access them
- **A chain of legitimate custody**
- **A trusted third party custodian**

Digital Forensics View (cont.)

Principles useful to justify migration for purposes of preservation

- **Principle of non-interference**: the method used to gather and analyse digital data or records does not change the original digital entities
- **Principle of identifiable interference**: if the method used does alter the original entities, the changes are identifiable

Digital Forensics Process

1. **Location and recovery** of the digital evidence
2. **Prioritizing the examination** of the potential evidence
3. **Examination** of the recovered material
4. **Evaluation and interpretation** of the findings
5. **Presentation of the results** in a report that should include factual findings, interpretation, and expert opinion
6. **Technical and administrative review** by a third party

Why is It Important to Know What a Digital Record Is and If It Is Trustworthy?

- Evidence
- Accountability
- Protection of Rights
- Preserving Identity
- Understanding the Past
- Relying on Sources
- ***Quod non est in actis, non est in mundo*** - What is not in the records does not exist

How Can We Use Digital Forensics?

- To design digital systems that create and maintain trustworthy digital records that can be regarded as material evidence of facts and acts, serving at the same time transparency, accountability and users needs
- To ensure that the authenticity of digital records can be verified when its presumption is weak
- To determine how records should be reliably extracted from the systems in which they reside and maintained in long-term storage in such a way that their authenticity can be presumed
- To determine how records should be authentically reproduced to keep them accessible and usable

How Can We Use Digital Forensics? (cont.)

- To determine how the features of the records, the actions conducted over them, and the changes caused by such actions should be documented
- To establish how the records submitted to court as evidence should be kept after the conclusion of court proceedings for as long as needed, so that they remain trustworthy
- To determine how long-term preservation activities can be conducted in such a way that the records will continue to be considered authentic for as long as they exist

InterPARES Web Site

www.interpares.org

Digital Records Forensics Web Site

www.digitalrecordsforensics.org