InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Due Diligence & Professionals

Keeping Records in a World Gone Digital

Due Care and Due Diligence

- Professionals have **a duty of care towards their records**, ethically and legally, because they are accountable for their actions to their profession (ethically), to their organization (administratively and legally), to society (legally and morally) and to the next generation (historically and morally)

- To fulfill **due care**, they need to exercise **due diligence**, that is, they need to comply with existing legal requirements, standards, best practices, and leading research recommendations. In order to comply, they need to be informed about current issues.

Advantages of the Digital Medium

- Digital materials do not fade or become yellow and brittle
- It is easy to alter them without leaving a trace for editing or repurposing
- They occupy very little storage space
- They can be copied an infinite number of times
- They can be shared over the Internet
- They can be sent and received across the world within seconds

Disadvantages of the Digital Medium

- A computer is needed to read digital materials: The medium does not contain any given record or work but only bit-strings

- It is not possible to preserve digital materials but only the ability to reproduce them

- There is no longer an original

- Authenticity is no longer verifiable on the work itself

- The easiness of reproduction makes it difficult to identify the final version

…and more

- The Internet makes intellectual property increasingly difficult to protect
- Viruses and technology failures make it easy to lose everything
- Technological obsolescence makes digital materials inaccessible very fast
- The information provided by the materiality of the object no longer exists

…and bad habits make it worse

- Hybrid systems
- Creating materials in different applications and leaving them there
- Not doing regular back-up and upgrading of files
- Not keeping media in the right climatic environments
- Not refreshing the media

InterPARES Principles

- Technology cannot determine the solution to the reliable and accurate creation of digital records or to their authentic preservation over the long term: **the professional's needs** define the problem and **archival principles** must establish the correctness and adequacy of each technical solution
- Solutions to the digital records challenges are inherently **dynamic** and **specific** to the cultural, disciplinary, administrative and legal situations

InterPARES Principles (cont.)

- Preservation is a **continuous process that begins with records creation**
- We must be able to **presume records trustworthiness (reliability, accuracy and authenticity)**, till proof to the contrary is established

Records Trustworthiness

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

More specifically...

Reliability

Reliability: the trustworthiness of a record as to its *source*, that is, a reliable person, a reliable software, and a reliable process.

The software should be **open source**, because the processes of records creation and maintenance can be demonstrated either by describing a 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 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

Authenticity

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

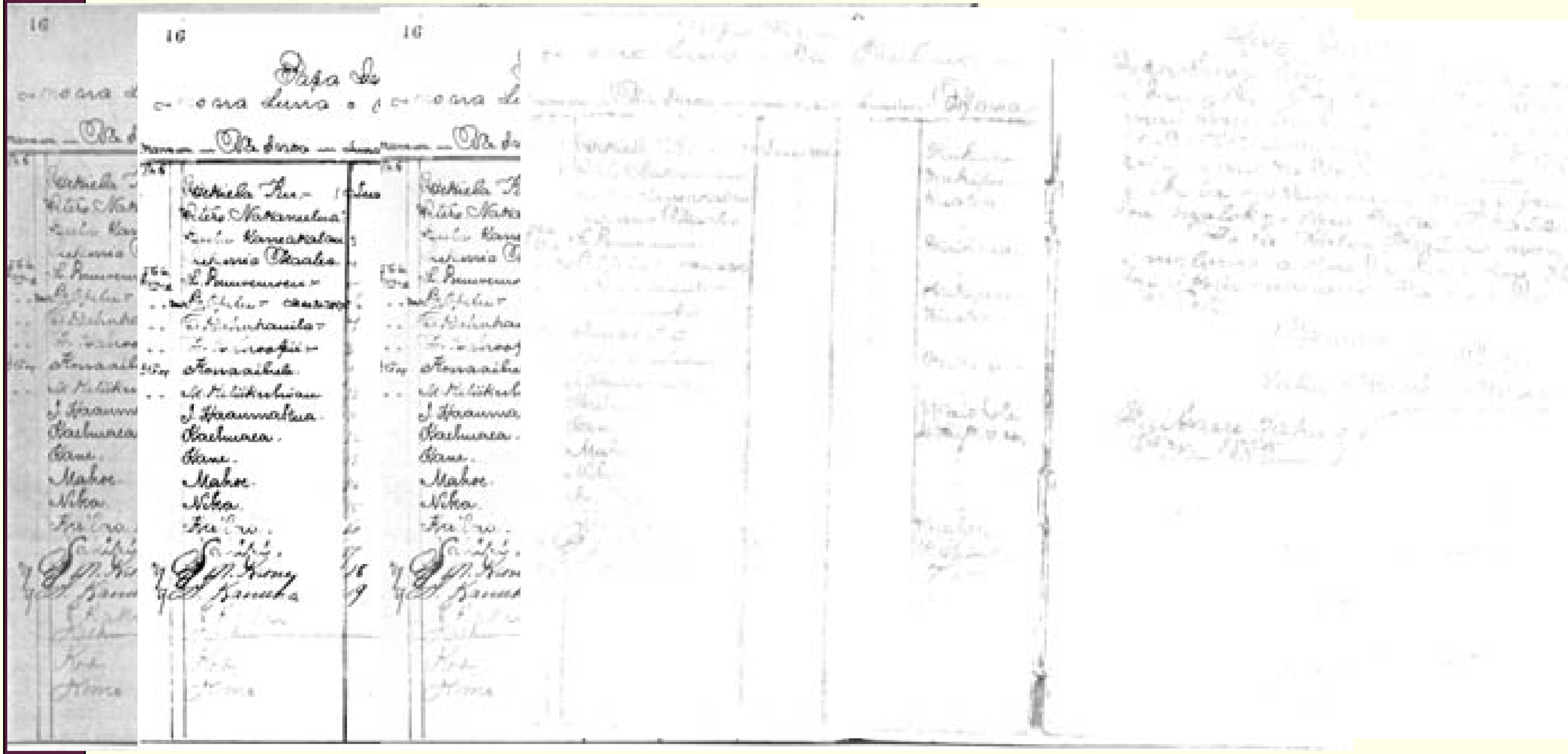
Data integrity: the fact that data are not modified either intentionally or accidentally “without proper authorization.”

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
  - additional hardware and/or software;
  - method of determining if the record has been altered, maliciously or otherwise;
  - audit logs and strong methods (e.g. checksum, hash algorithms), as it cannot rely on file size, dates or other file properties

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

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 (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’
- Ability to determine accountability of people
- Provide the necessary snapshot for post-event reconstruction (‘black-box’)
- Answers Who-What-Where-When, but only if retained for sufficient time (space vs. money vs. risk vs. knowledge)

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.

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

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

Assessment of Integrity

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

Non-interference: the method used to maintain, use 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

How to go about it?

1) Selecting Software & Hardware
2) Ensuring Stability of Digital Materials
3) Identifying Digital Material
4) Supporting the Presumption of Integrity
5) Organizing Materials into Logical Groupings
6) Authentications
7) Protecting Your Materials from Unauthorized Action
8) Protect Records from Accidental Loss & Corruption
9) Protecting Materials Against Hardware & Software Obsolescence
10) Planning for Long-Term Preservation

1. Selecting Software & Hardware

- Choose software that presents your old materials as they originally appeared (**compatibility**) and allow you to share materials easily (**interoperability**)
- Use software that adheres to **standards** (*de jure* or *de facto*)
- Maintain the **specifications** of software
- **Document changes** when software is customized (include comments in the software code)
- **Document** the **construction** of your system
- Choose **non-proprietary**, platform independent, **uncompressed** formats with freely available specifications (open format) and software whose code is made freely available and can be modified (**open source**)

2. Ensuring Stability of Digital Materials

- **Stability** is the essential characteristic of every record and means:
  - **fixed content** (content cannot be overwritten, altered, deleted or expanded—if it can, make sure you have a log of changes and an audit trail and you consider your system the record rather than its content) and
  - **fixed documentary form** (the rules governing the presentation on the screen are unchangeable)
  - manifested and stored records must be clearly **linked** (the number of variations are limited)
- The **documentary form** of each record associated with each activity should be defined at the outset
- **Do not compress** your digital material

3. Identifying Digital Materials

- Record essential information about the record so that it may be uniquely identified (**identity metadata**=properties), retrieved, and contextualised:
  - names of **author**, writer, originator, addressee
  - name of **action**, matter, subject, or simply the title
  - **dates** of compilation, transmission, receipt, filing
  - documentary **form**
  - digital **presentation** (format, wrapper, encoding, etc)
  - **attachments**, if applicable
  - intellectual **rights**, if applicable
  - presence or removal of digital signature, or other form of **authentication**
  - name of **person responsible** for the record, if applicable
  - name of or code for the file or **group** of records in which the record belongs

4. Supporting the Presumption of Integrity

Record information that helps to infer that the record is the same as when created (**integrity metadata** or properties):

- name(s) of **handling persons** over time
- name of person primarily **responsible for keeping** the record
- indication of additions (**annotations**) made to the record
- indication of **technical changes** (e.g. format, encoding, upgrading, changes to digital components, migration)
- indication of presence or removal of **digital signature**
- **planned removal from the system**, transfer to a custodian, deletion—business records exception to hearsay applies here
- existence and location of **duplicates** outside the system

The latter two bullets establish a proper **chain of custody** (more on security later)

5. Organizing Materials into Logical Groupings

- Separate **records** from other types of materials, the **business records** from the personal ones, ephemeral from **official** (not on the basis of form: an e-mail may be considered a contract)
- Create a classification scheme or **filing plan** or a structured directory to provide a logical place for each record
- Ensure that such scheme, plan or directory **corresponds to the way your non-digital records are organized**
- Provide each new digital record with an **identifier** showing its proper place in the scheme (e.g. a code or the name of the file and of the higher groups in which the file belongs) and include this among the record's identity metadata
- Identify **how long records need to be kept**—big buckets theory...not a good idea
- Make **decisions at the group or file level**, not the individual record (maintaining consistency between the records on different media belonging to the same file or group)

6. Authentication

Technology dependent authentication

- Nature of authentication techniques
  - digital signature
- Obsolescence and authentication
  - preservation issues
- Managing documents with digital signatures
  - integrity metadata

Technology independent authentication

7. Protecting Your Materials from Unauthorized Action

- You must be able to demonstrate that it is **impossible to tamper** with your materials without being identified
- **Do not download your records to computers that are not under your direct control** (privacy issues compound problems of trustworthiness)
- **Restrict physical access** to your computer(s)
- **Create access permissions** for all legitimate users of your system
- **Maintain an audit trail of access** to your system and the materials in it
- If you cannot prove the **authenticity** of your material, it is irrelevant that it is authentic
- Authenticity of a record can be challenged by **challenging the identity of its author**
- Evidence that **your record system is trustworthy** lays the foundation for the authenticity of your records
- If you are an organization or community of practice, have a **procedures manual**

Records Procedures Manual

In the event of legal proceedings, the procedures manual can be the most important support to satisfy the legal requirements for electronic records .The procedures manual can be used by a witness as evidence to prove that

- an authorized Records Management System (RMS) program is followed;
- the RMS program provides proof of the system's integrity, i.e., the system reliably records, stores and processes electronic records; and, therefore,
- the electronic records are authentic (have identity and integrity), so trustworthy and reliable documentary evidence can be produced from them at any time and for any purpose.

If you create a manual you should also include in it the activities related to what is detailed in items 1-10 of this presentation

8. Protect Records from Accidental Loss & Corruption

- Ensure that **your system is backed up** at least once a day; use the best backup technique for your situation; ensure your backup system include an audit trail
- The **purpose of your backup is to recover the system in case of failure**, not to keep records. Destroy your backups on a regular basis (e.g., every third day)
- **Duplicates of your material should be kept on additional hard drives**. If they are on tapes or discs, remember to refresh and upgrade them periodically
- The **integrity of the electronic record system** guarantees the trustworthiness of the record
- The reliability of a computer program can be established by showing that **users of the program actually do rely on it** on a regular basis, such as in the ordinary course of business

9. Protecting Materials Against Hardware & Software Obsolescence

- **Eliminate dependence** on specific hardware by transferring its functionalities to the software (ask IT to help you)
- **Plan for regular technology upgrades** (keeping in mind the need for backward compatibility)
- **Consider external storage** for infrequently used records (including computer output microfilm for textual records that do not require random search)
- If you remove materials from your live system, associate with it the **system documentation and all the necessary information about the material** to be able to maintain accessibility and to understand the material itself

10. Planning for Long-Term Preservation

- Identify the materials that need **to be preserved for the long term**—”total archiving” is a bad idea
- Identify a **trusted custodian for the records** (in-house or external)
- Establish a **preservation strategy** early and in consultation with the designated trusted custodian
- Follow this set of recommendations

Conclusion

You should be aware of the risks of neglecting to manage your digital records properly, and:

- Adopt measures that **work best in your situation**
- **Consult** with professional archivists
- Review other **InterPARES documents**, such as the Framework for Policies and the Preserver Guidelines
- Review standards such as the **Canadian Standard "Electronic Records as Documentary Evidence"** (CAN/CGSB-72.34-2005) *(Canadian General Standards Board)*
- For complex recordkeeping systems, review the **DoD 5015.2** (2007) standard and the **MoReq2** standard (2008)

Findings and Products

All findings and products of InterPARES are available on the InterPARES Web site:

www.interpares.org