Keeping Digital Documents Authentic Over The Long Term

in saeculis authenticus

Dr. Luciana Duranti
InterPARES Project Director

Advantages of the Digital Medium

- Digital documents do not fade or become yellow and brittle
- It is easy to alter them without leaving a trace for editing purposes, for repurposing or just for reading them better
- 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 documents: The medium does not contain documents but only bit-strings

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

- There is no longer an original

- Authenticity is no longer verifiable on the document

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

- With databases, especially GIS, and with interactive and dynamic systems, often we have only views, not documents, or fluid data

…and more

- Documents including text, images, graphics, etc. are broken down and stored in different parts of the memory
- Images are very limited in their variety of colours
- The Internet makes intellectual property increasingly difficult to protect
- Viruses and technology failures make it easy to lose everything
- Technological obsolescence makes documents inaccessible very fast
- The information provided by the materiality of the document does no longer exist

…and bad habits make it worse

- Hybrid systems without proper connections
- Creating documents in different applications and leaving them there
- Not doing any back-up of files
- Not keeping media in the right climatic **environments**
- Not refreshing the media

and worse…

- Not migrating the documents
- Not protecting the documents from malicious or accidental tampering—by access
- Using protection systems—encryption or digital signatures—that do not allow for preservation
- Trusting brand names
- Confusing storage with preservation

Consequences

- Documents are less
  - reliable (manipulability),
  - retrievable (incongruence of classifications),
  - accessible (incompatibility and lack of interoperability),
  - readable or intelligible (obsolescence)
- It is difficult to prove their accuracy and authenticity
- It is difficult to provide for the long-term preservation of their authenticity
- It is difficult for the creating body to maintain its accountability

Reliability and Accuracy

Reliability is the trustworthiness of a document as a statement of facts or as content.

It is the responsibility of the author/creator.

Accuracy is the degree to which the data in the document are precise, correct, truthful, free of error or distortion. To establish it, one has to verify the controls exercised on the creation, transmission and preservation process.

Over time, the responsibility for it moves from the author, to the creator, the keeper and the preserver of the document.

Authenticity

- Refers to the fact that a document is what it purports to be and has not been tampered with or otherwise corrupted.

- Authenticity is the trustworthiness of a document as a document.

- Over time, the responsibility for it moves from the keeper to the preserver of the document. It is at risk during transmission across space and time

- To establish it, one must verify the identity and integrity of a document.

- Authenticity of the data in the document is also related to their identity and integrity

Authentication

- A declaration of authenticity, resulting either by the insertion or the addition of an element or a statement to a document, and the rules governing it are established by legislation.

- A means of proving that a document is what it purports to be at a given moment in time.

Authenticity vs. authentication

- Certain mathematical techniques are said to provide **incontrovertible** mechanisms for ensuring authenticity of digital objects (e.g., cryptographic digital signatures)
- Such technologies have been given legal or regulatory value (e.g., European Directive on electronic signatures, Security and Exchange Commission on hash functions).
- Digital signatures are enabled through complex and costly public-key infrastructures (PKI)

Digital signatures and preservation

- Digital signatures are great tools for ensuring/verifying authenticity of documents across **space …**

- **…** but not across **time**!

- Digital signatures are subject to obsolescence, and thus, only compound the preservation problem

- Preserving institutions have announced they will not attempt to maintain encrypted or digitally signed documents transferred to them

Conceptual Framework for Authenticity

- In archival theory and jurisprudence, documents that are relied upon by their creator in the usual and ordinary course of business are presumed authentic

- In digital systems, the presumption of authenticity must be supported by **evidence** that a document is what it purports to be and has not been modified or corrupted in essential respects.

- To assess the authenticity of a document, the keeper or preserver must be able to **establish its identity** and **demonstrate its integrity**

Identity of a Document

- It refers to the characteristics of a document that distinguish it from other documents. They include: the names of the persons concurring to its formation (e.g., author); its date(s) of creation and transmission; its title; an indication of the matter or subject; classification code or other unique identifier; as well as an indication of any attachment(s).

- These characteristics should be explicitly expressed either on the face of the document, or in metadata related to it.

Integrity of a Document

- Its wholeness and soundness. A document has integrity if it is intact and uncorrupted, that is, if the message that it is meant to communicate in order to achieve its purpose is unaltered

- Data are intact and uncorrupted if they are as accurate as they were when generated

- A document’s physical integrity, such as the proper number of bit strings, may be compromised, provided that the content and its required elements of form remain the same

- A document’s integrity can be verified either on the face of the document from its formal elements, or on the metadata expressing the responsibility for it and its technological changes

Key Documentary Features

A document **formal element** is a constituent part of its form.

A document **attribute** is a defining characteristic of each given document (e.g. name of author) or of a formal element in it (e.g. title)

A document **digital component** is a digital object that may contain all or part of a document, and/or the related metadata, or more than one document, and that requires specific methods for preservation.

Other Features

- The relation between a document and a file can be one-to-one, one-to-many, many-to-one, or many to many
- The same formal aspect of a document can be created by a variety of digital presentations and viceversa, from one digital presentation a variety of documentary forms can derive
- It is possible to change the way in which a document is contained in a file without changing the document

Implications for Authenticity

- The identity of a document can be demonstrated by its formal elements and attributes expressed as metadata

- The integrity of a document may be demonstrated by formal elements found on its face, or in attributes linked to it as metadata, or in one or more of its contexts

- Both depend on the existence of a trusted custodian also responsible for the reproduction process

- A **trusted custodian** is a professional who is educated in recordkeeping and preservation, who has no stake in the content of the document and no interest in allowing others to manipulate or destroy the records

Key points concerning preservation

- Technology cannot determine the solution to the long-term preservation of digital documents: organizational needs define the problem and principles derived from the nature of the documents must establish the correctness and adequacy of each technical solution

- Solutions to the preservation problem are inherently dynamic and specific

- Preservation is a continuous process that **begins with document creation** and whose purpose is to transmit authentic documents across time and space

Document Creation Principles

1. Select software with a track record of on-going compatibility with its earlier versions and with wide interoperability.

2. Ensure that digital information that needs to be kept as a document has stable and complete content and fixed form.

3. Establish who will be responsible for preservation activities, and determine the preservation strategy before document creation begins or as soon as possible afterwards.

4. Assign to each document the attributes or metadata necessary to establish and maintain its identity and integrity

Document Creation Principles

5. Organize digital documents into logical groupings consistent with the organization of the paper documents and as much as possible linked to retention periods.

6. If it is necessary to use some form of digital authentication, ensure that the kind of authentication selected does not hamper the maintenance and preservation of the document

7. Protect your documents from non-authorized action

8. Protect your documents from accidental loss and corruption

Where to find more information
InterPARES Web Site

www.interpares.org