Digitizing News and Information: Consequences for Preservation


in saeculis authenticus
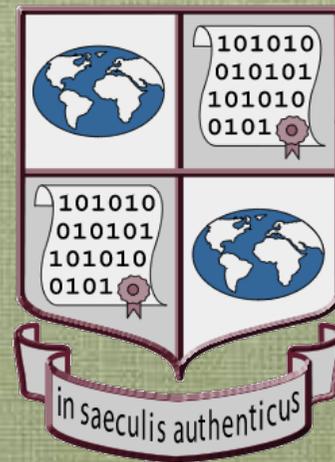
Dr. Luciana Duranti
InterPARES Project Director

Advantages of Digitizing

- Digital documents do not fade or become yellow and brittle
  If the documents we are digitizing are already so, digitization can give them a lift

- They occupy very little storage space

- They can be copied an infinite number of times

- They can be made accessible over the Internet

- They can be sent and received across the world within seconds

- They can be magnified for easy reading, compared on the screen, accessed randomly

- The allow to protect the paper documents from deterioration

Disadvantages of Digitization
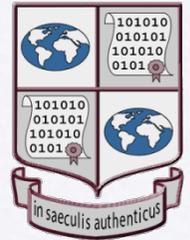
- The Internet makes intellectual property increasingly difficult to protect

- 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

- Authenticity of the source document is not verifiable

- The information provided by the materiality of the document does no longer exist, because digitization does not reproduce a document, like a facsimile does, but visually imitates its form

- Technological obsolescence makes documents inaccessible very fast

made worse by…
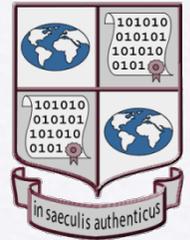
- Inadequate or non-existent metadata

- Lack of standards, policies and best practices

- Competition for resources with born-digital material

- Trusting brand names

- Confusing storage with preservation

- Downsizing of archives staff because of the expected productivity gains

See McCargar Victoria, "Following the trail of disappearing data," *The Seybold Report* 4(21) February 9, 2005, 7-14

Consequences
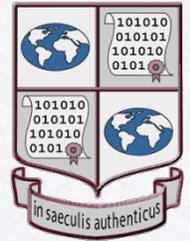
- Digitized documents are not always

    - Accurate

    - Retrievable

    - Accessible (because of incompatibility and lack of interoperability),

    - Readable or intelligible (because of obsolescence)

- It is difficult to prove their accuracy and authenticity without the original

- It is difficult to provide for the long-term preservation of their authenticity

What to keep in mind
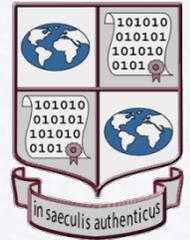
- European Commission (Guidelines 2010 for digital libraries): "To put digitized material online does not imply that the user will easily retrieve it and even less that such material will be accessible and usable for research. It is necessary to provide adequate services that allow the user to retrieve content and work with it. This presupposes a structured and high quality description of the contents and proper support for its use."

- A digital news library is a "conversation," an interaction between users and a distribution community, organized in services, and requires flexibility: "plurality" and "convergence" are the key words.

- A digital news library is not a system, a great systematic narration, but many conversations held together by a common language, thus the "common language" becomes the crux of the problem

What to keep in mind (cont.)
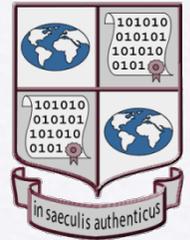
- The common language is "inter-operability"

- "Technological interoperability": the use of standard formats for the representation of content and related metadata, and the use of standard protocols for their transmission

- "Semantic interoperability": the definition of the same entities using the same terms within the various digital domains

- "Legal, administrative or organizational interoperability": the removal of legal, regulatory, or organizational obstacles to the free circulation of content among the various digital domains

What to keep in mind (cont.)
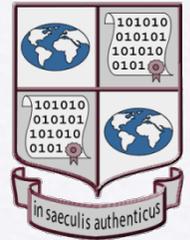
- Ultimately, the focus of the work of digitization is the user

- If the work is well done, it will be the user (also inter-operating) who will build his/her own news library, following his/her own route

- The other two entities involved with the user in this conversation are the content owners and the service providers (content distributors)

- To support the conversation is very useful the model of the Open Archival Initiative, with its Protocol for Metadata Harvesting: a technological model that has become an organizational and service model

but…what about preservation?
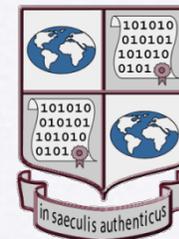
- Digital preservation involves maintaining the technical accessibility over time to
--the digitized images,
--the born digital information about the images that allows
    for their identification, retrieval and integrity, and
--to the web site and specific pages containing the images

- Digital preservation involves assigning specific responsibilities to a trusted custodian on the basis of a clearly defined mandate—the Open Archival Information System standard (ISO 2003) is a good enough model

- Digital preservation requires constant monitoring, technological and financial sustainability, and organizational commitment

Key points concerning preservation
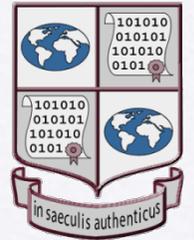
- Technology cannot determine the solution to the long-term preservation of digitized documents: organizational needs define the problem and preservation and access principles must establish the correctness and adequacy of each technical solution

- Solutions to the preservation of digitized documents are inherently dynamic and specific

- Preservation is a continuous process that **begins with the initial digitization choices** and whose purpose is to transmit authentic copies of the digital documents across time and space

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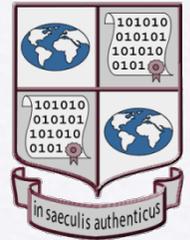
1. Select software with a track record of on-going compatibility with its earlier versions and with wide interoperability.

2. Organize digitized material into logical groupings consistent with the organization of the corresponding paper documents and as much as possible linked to retention periods.

3. Assign to each image the attributes or metadata necessary to establish and maintain its identity and integrity, in addition to ensuring its retrievability

4. Ensure that this born digital information, which needs to be kept with the images, has stable and complete content and fixed form.

Images Creation Principles

5. Establish who will be responsible for preservation activities, and determine the preservation strategy (e.g. migration) before digitization begins or as soon as possible afterwards.

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 digitized material from non-authorized action with security measures and audit trails

8. Protect your digitized material from accidental loss and corruption, through regular copying and back-up
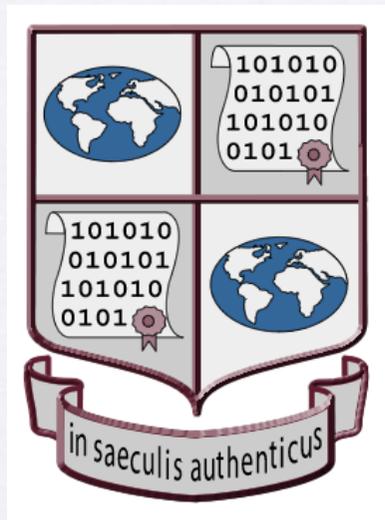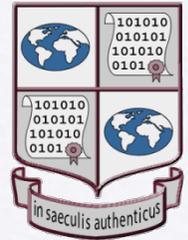
…and then what?
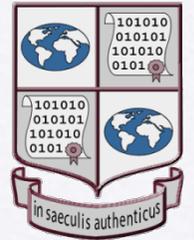
KEEP THE PAPER DOCUMENTS!!!

(if you really have to destroy them, microfilm them first)

Where to find more information InterPARES Web Site

www.interpares.org