Technical Challenges in Digital Preservation

Dr. Jens-Martin Loebel

2014-04-24

Spokesperson Special Interest Group "Long-Term Preservation"
German Informatics Society (Gesellschaft für Informatik e. V.)

Gesellschaft für Informatik

nestor

http://digitale-bewahrung.de/

http://skriptorium.org/

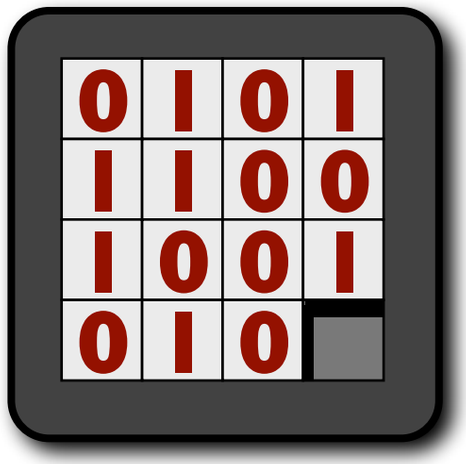
Our Digital Cultural Heritage

The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital object. [...] They are frequently ephemeral, and require purposeful production, maintenance and management to be retained.

UNESCO (Ed.): *Guidelines for the Preservation of Digital Heritage*. Doc CI-2003/WS/3, 2003.
See: http://unesdoc.unesco.org/images/0013/001300/130071e.pdf

Analog vs. Digital Objects

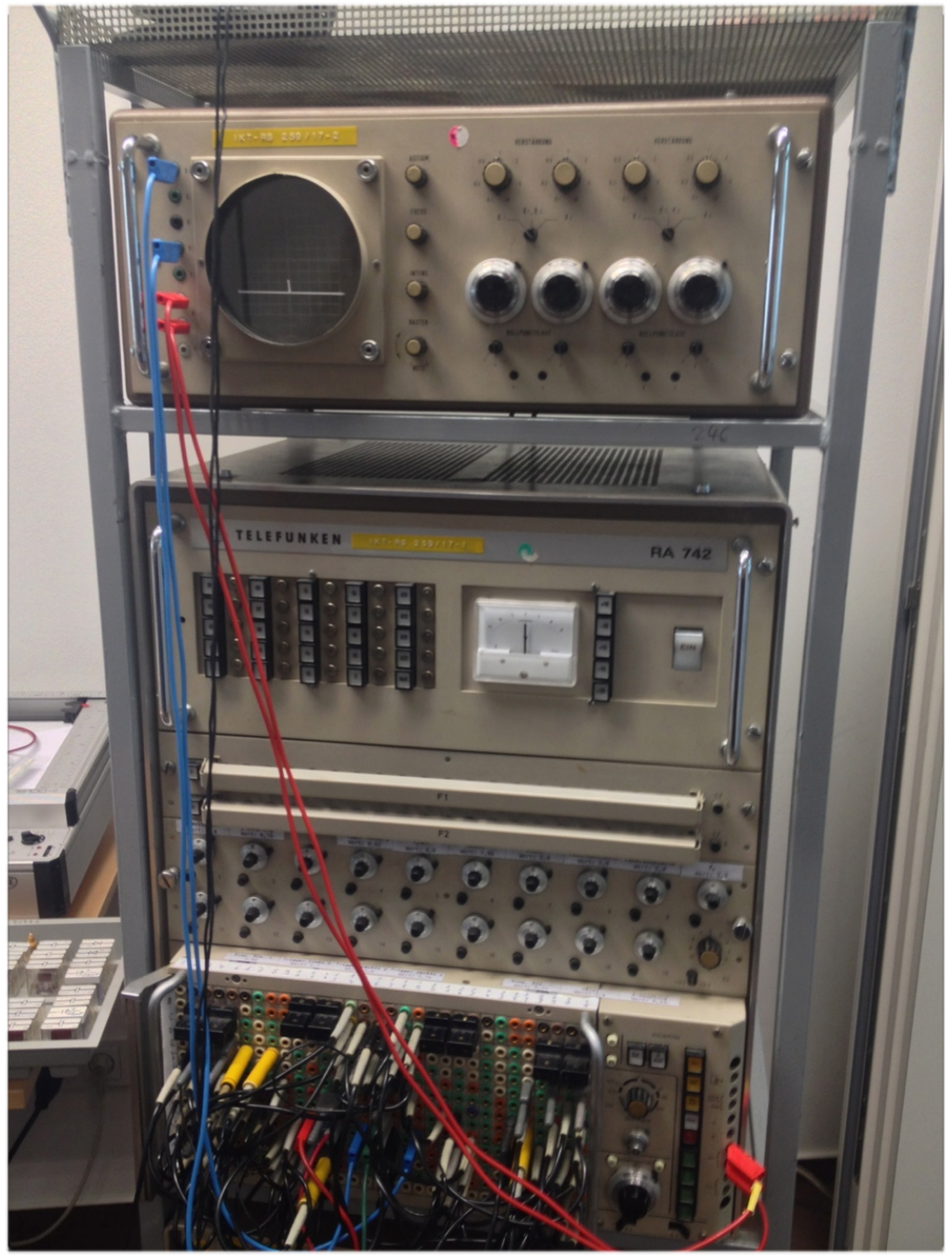
- directly readable
- no lossless copy
- aging, wear, deterioration



- requires interpretation through software and hardware
- universal bit stream of „ones“ and „zeros“
- lossless copy of bitstream
- no deterioation of bitstream

The Structure of Digital Objects

- **Physical**: the bitstream stored on a particular medium
- **Logical**: defines how the bits are used by a software application, data types / file formats, interpreted by software
- **Conceptual**: "real world" objects, visible result to the user, representational



see Thibodeau, K.: *Overview of technological approaches to digital preservation and challenges in coming years*. In: „The State of Digital Preservation: An International Perspec- tive" – conference proceedings, Juli 2002. S. 4-31.

Longevity of Storage Media

- longevity and durability of digital media is drastically lower than analog media
- high turnover in storage media technology
  - needs to be copied at regular intervalls
    - checksums needed to ensure copy = original
  - reading devices and other hardware may no longer be available

„Digital Information Lasts Forever – Or Five Years, Whichever Comes First.“

Jeff Rothenberg
Senior Researcher, RAND Corporation

Longevity of Storage Media

medium	lifespan
stone tablets	more than 1000 years
books (on acid-free paper)	several hundred years
microfilm	about 500 years (realistically 50 years)
CD-ROM / DVD (optical media)	25-100 years (realistically 3-5 years)
floppy disks / hard drives	5-10 years
magnetic tapes	up to 30 years

Media Formats

- increased rate of obsolescence of storage media and systems
  - Result: obsolete file formats and file systems
  - display/authoring software no longer available
- hardware no longer available
- different requirements for each data type

(styled) Texts, Fonts
Databases, Research Data
Software Programms
Images, Photos
Videos, Animation
Audio, Speech, Samples, Music
Multimedia, Mixed-Media
...

Software Challenges

- usually proprietary formats often inadequately documented
- variety of non-interchangable file formats
- few standards
- Digital Rights Management (DRM): technical and legal restrictions

Multimedia Publications in Memory Institutions

- the proportion of „multimedia“ publications in libraries, archives and museums is on the rise

  - Deutsche Nationalbiliothek (2011)
    > 57000 multimedia publications

  - Koningklijke Bibliotheek
    > 5500 publications

  - Bibltiothèque nationale de France
    > 65000 publications on over 100000 disks

  - Computerspielemuseum Berlin
    > 11000 games

Preservation Strategies I
Analog Backup

- Printouts
- Computer Output on Microfilm (COM)
- resource and time consuming, only possible for certain static media types (text and images)
- limited accessability
- high demand storage requirements

Barbarastollen underground archive

photo: Wikipedia user:Joergens.mi

Technology Preservation (Computer Museums)

- preservation of original content, hardware and peripherals
- resource consuming
- limited accessibility
- no replacement parts
- knowledge and repair skills for obsolete systems required

Restored Virtual Reality System at Computerspielemusem Berlin

Photo: Winfried Bergmeyer

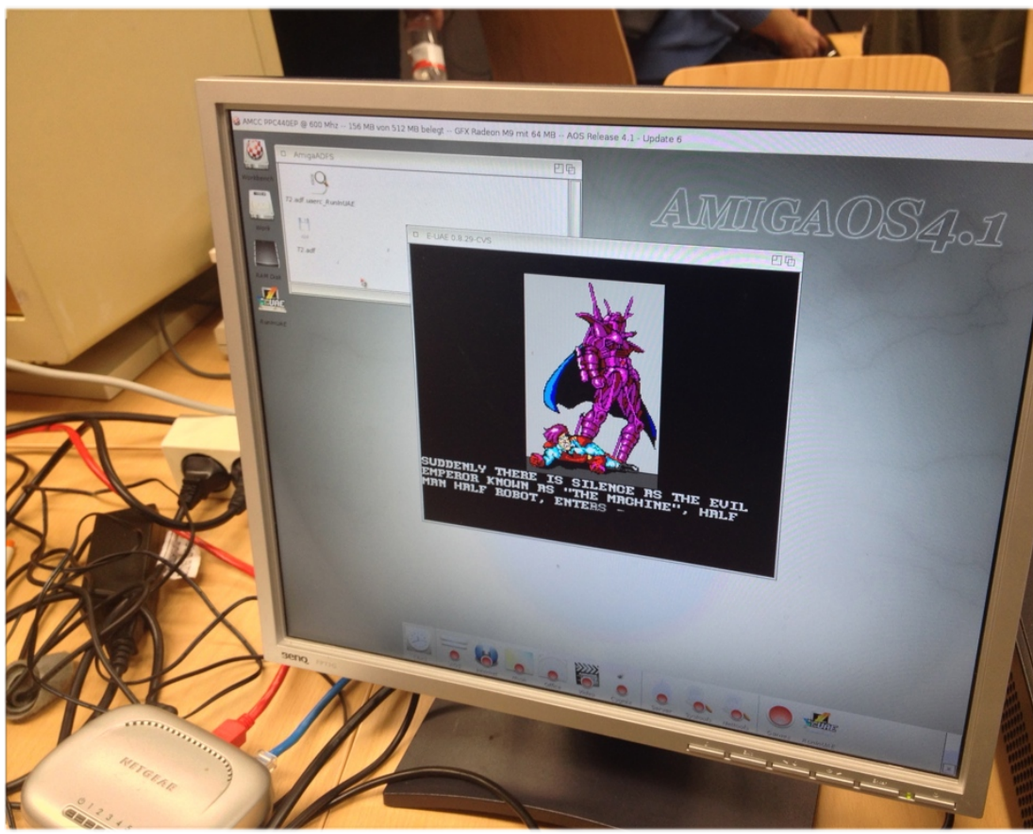
Preservation Strategies II
Content Migration

- transcoding of file formats to current generation
- lossless vs. lossy compression
- conversion of content, loss/change of appearance
- changes „original“ file
- may need manual post-processing
- not viable for interactive / multimedia formats
- not possible for proprietary formats

Preservation Strategies III
Emulation

„Emulation involves using software that makes one technology behave as another“ UNESCO, Guidelines for the preservation of digital heritage

- long-term preservation entails the emulation of obsolete hardware with software
- used first by IBM in 1962
- does not change original file(s)
- high development costs
- able to preserve „look-and-feel“
- higher authenticity



Digital Repositories

- Persistent Identifiers, Metadata
- Open Archival Information System (OAIS)
- Open Source Examples: DSpace, Fedora

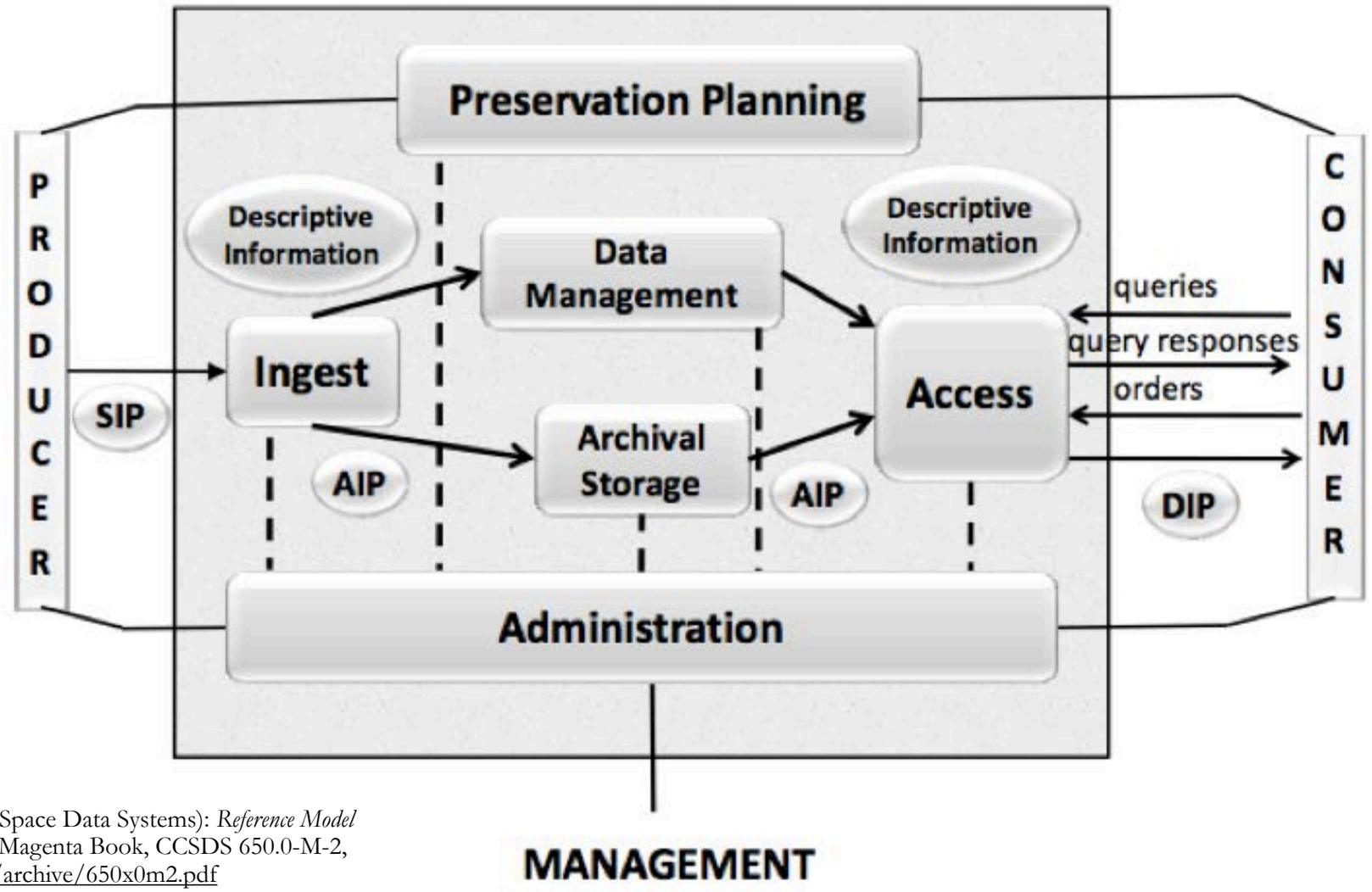


Image: CCSDS (Consultative Committee for Space Data Systems): *Reference Model for an Open Archival Information System (OAIS)*. Magenta Book, CCSDS 650.0-M-2, p. 4-1 - http://public.ccsds.org/publications/archive/650x0m2.pdf

Conclusion

- many technical and legal challenges remain
- emulation strategy is the most promising
- scalable preservation strategies necessary to deal with ongoing obsolescence
- authenticity and integrity
- preservation of system knowledge
- accessibility enhancements / technologies needed

Gesellschaft
für Informatik

Thank You

Special Interest Group "Long-Term Preservation"
German Informatics Society (Gesellschaft für Informatik e. V.)

http://digitale-bewahrung.de/

Twitter: @DigiBewahrung

http://skriptorium.org/