TIB | German National Library of Science and Technology

Selection criteria for digitization and digital preservation

A comparison

Thomas Bähr and Michelle Lindlar

Digital Library of Today and Challenges of Tomorrow

January 24th 2013, Jagiellonian Library, Kraków

TIB facts
A brief introduction

- German National Library of Science and Technology…
- … for engineering, architecture, chemistry, computer science, mathematics and physics
- Founded in 1959
- 212 members of staff
- Holdings in shelf length: 125 km
- 24.700 current journals (17.700 print, 12.000 electronic)
- 15.75 million patents and standards
- Special collections: specialist literature from Eastern Europe and East Asia, research reports, conference proceedings
- Goportis digital preservation system since 2010
- In-house digitization since 2012

- OPF member and nestor cooperation partner

Open Planets Foundation

nestor

Digitization and digital preservation
Definition

*digitization **is not** digital preservation !*

digitization

- reformatting analogue material to a digital form
- to improve access, to protect the original, to preserve content
- analogue content on a variety of carriers inlcuding text, (moving) images and sound

digital preservation

- monitor and counteract dangers of digital decay and digital obsolence
- to ensure long-term access over time
- digitized content and born-digital content

Image: „Preservation Hall Sign“ by swampier
http://www.flickr.com/photos/32425581@N03/3252065466/

Digitization and digital preservation
Correlation

„Scan all you want, but think about preservation, too.“

Bill LeFurgy – Blog „The Signal“

… and start thinking about preservation **before** you scan, e.g.

- When choosing a format
- When deciding on a resolution or bit depth
- When thinking about compression
- When deciding which metadata to create [automatically]

But is there a correlation between the selection process for digitization and the selection process for digital preservation?

Quote: http://blogs.loc.gov/digitalpreservation/2011/07/digitization-is-different-than-digital-preservation-help-prevent-digital-orphans/

Image: http://www.flickr.com/photos/sixteenmilesofstring/1989059615/

Selection processes

Conservatory aspects

- close to selection processes in „traditional [analogue] preservation“ workflows
- based on physical factors

Contextual aspects

- close to selection processes in „traditional acquisition“ workflows
- based on aspects derived from the content/context of the work

Organizational aspects

- close to library management processes
- usually function as contributing factors to conservatory or contextual aspects

Image: „This one looks good“ by Kate Thomlison
www.flickr.com/photos/katetomlinson/3992576181/

Selection based on conservatory criteria

Key factor is the physical state of the object

- for analogue materials: physical state of the carrier
- for digital materials: physical state of the carrier and state of the bitstream

Preventive selection

- digitization:
  usually conducted to conserve the original by facilitating use to the digital fascimile
  example: Rare book digitization
  original kept: yes

- digital preservation:
  media migration from one carrier to another is an established process in digital preservation
  example: migration of dissertations on CD-R to HDD
  original kept: usually not

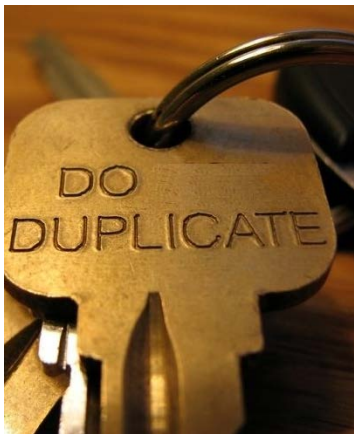

Image: http://www.flickr.com/photos/wlef70/5681028120/in/set-72157626498029089/lightbox/

Selection based on conservatory criteria

Remedial selection

- digitization:
  heavily damaged material or obsolete AV-carriers
  examples: in the case of audiovisual materials, especially certain reel-to-reel tapes, digitization is the only valid strategy
  original kept: no

- digital preservation:
  obsolete data carriers or data formats
  examples: Amstrad CPC 3“ floppy discs
  original kept: no

Image: http://www.flickr.com/photos/49730818@N04/5985620512/

Selection based on contextual aspects

Key factors are based on content / context of the work

- content / context based criterion require closer definition, e.g. usage, intrinsic value or uniqueness
- parameters may contradict each other

- digitization:
  most common selection process
  example: dissertations between 1820-1850
  original kept: varies

- digital preservation:
  technical environment and format as contextual factor ?
  example: event-based webharvests, e.g. Library of Congress US election 2012 webarchive
  Original kept: varies

Image: http://commons.wikimedia.org/wiki/File:Compact_Floppy.jpg

Selection based on organizational criteria

Key factors are based on overarching management/policy criteria

- usally do not form a selection on their own (exception: mass digitization)
- play a major role in every other selection process

- in digitization:
  example: based on conservatory selection reel-to-reel tapes need to be digitized but the funds are not available

- in digital preservation:
  example: media needs to be transfered from floppy disks to be saved but the heritage institution does not have the legal right to do so

Image: http://www.flickr.com/photos/vaguery/2210098036/

Conclusion

- main selection criteria work for digitization and digital preservation alike

- the processes differ in the decision to keep the original
    - is this due to the fact that the „digital original“ is impossible to define?
    - how important is „the original look and feel“ to us?

- digital preservation workflows benefit greatly from knowledge about the selection process behind the digitized objects

- a close coordination between selection for digitization and digital preservation enables optimal use of resources

Image: http://www.flickr.com/photos/titoperez/3437322060/

TIB | German National Library of Science and Technology

Thank you for your attention!

Contact details:
Thomas Bähr
Head of Digital Preservation
German National Library of Science and Technology (TIB)
thomas.baehr@tib.uni-hannover.de