TIB | German National Library of Science and Technology

Non-textual materials at TIB – an emulation use case?

Michelle Lindlar

OPF Hackathon“Emulation, learn from the experts“

17.11.2012 Freiburg im Breisgau

Disclaimer: We do not emulate (yet?).
The use cases illustrate possible emulation strategies based on our materials (and not concrete plans).
This presentation includes questions.

Image: http://iconbug.com/detail/icon/1128/blue-question-mark/

What I‘m going to talk about

- Who we are
- What we do
- Use case 1: AV-portal
- Use case 2: 3D objects
- Questions, questions, questions …….

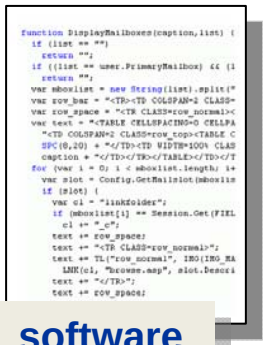
Who we are

Main Building - Hannover, Germany

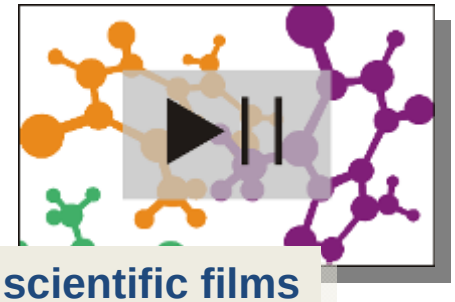
Who we are - TIB facts

- German National Library of Science and Technology…
- … for engineering, architecture, chemistry, computer science, mathematics and physics
- Founded in 1959
- Financed by Federal Government and all Federal States (regional:national = 70:30)
- 212 members of staff
- Member of Leibniz Association
- 6.1 m items (2.8 m books, 3.3 m non-electronic materials)
- 24,700 current journals (17.700 print, 12.000 electronic), 51% in exclusive possession, 84% from abroad
- 15.75 m patents and standards
- Grey literature comprises around 1/3 of the total TIB holdings
- GetInfo: over 150 million data sets, including 20 million of TIB‘s own

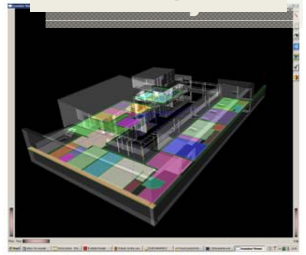
What we do
Vision and Strategy



software

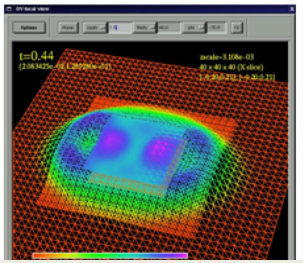


scientific films

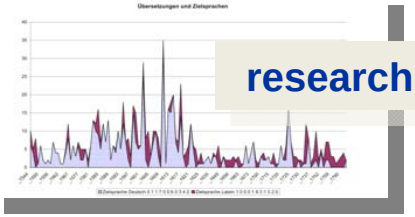


3D-objects

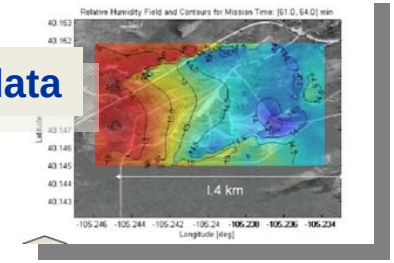
TIB



simulation



research data



text

What we do - Non-textual materials at TIB

- 2005 – first DOI registration agency for research data sets in the fields of science, technology and medicine
- 2009 – founding members of DataCite
- 2006-2011 PROBADO (DFG funded project)
- 2011 establishment of a Competence Centre for Non-Textual Materials at TIB
  - „[…] to fundamentally improve access to, and use of, non-textual materials and to enable new forms of usage for existing inventories.“
  - AV webportal design and development with Hasso-Plattner Institut for software system engineering at Potsdam University (HPI) as a technology partner
- 2013 – start of EU-funded DuraArk project (Durable Architectural Knowledge)

What we do - Digital Preservation at TIB

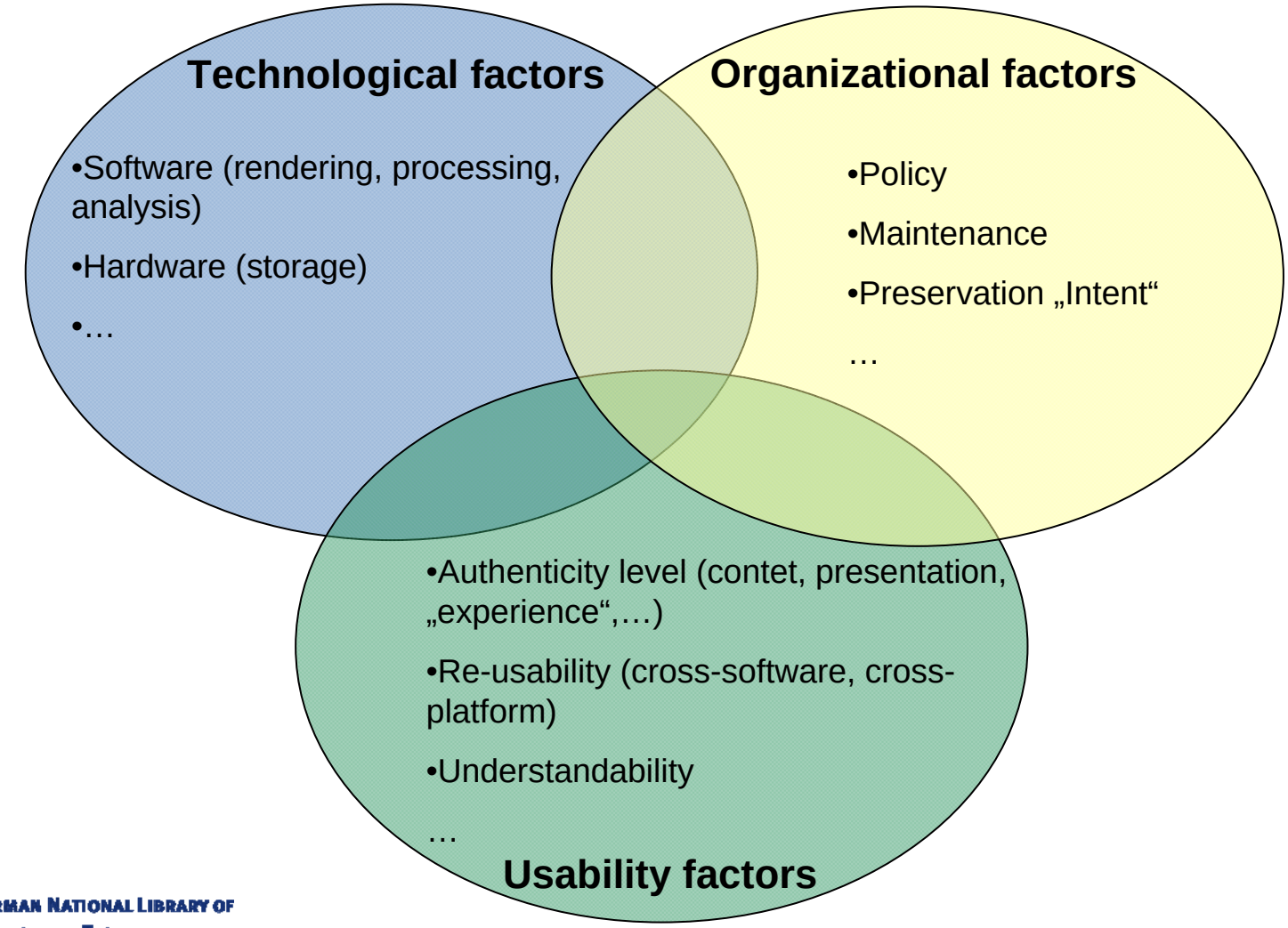
Within Goportis:

- Goportis is the consortia of the three German subject libraries (TIB, ZB MED, ZBW)
- 2009 – 2011 Goportis pilot project to evaulate technological and organizational needs for a cooperatively operated digital preservation system
- OPF member and nestor cooperation partner

On an institutional level:

- TIB hosts the digital preservaton system within the consortia
- Risk assesment, content and workflow analysis
- Strategies for non-textual materials

What we do - Implications of a Preservation Strategy



What we do – Preservation Strategy

- Software basis of the Goportis Digital preservation system is Rosetta by Ex Libris
- Rosetta allows capturing **extensive preservation/technical metadata** (e.g. output of technical metadata extraction through jhove, mediainfo, etc.)
- No normalization or limitation to certain formats in the (pre-) Ingest

We currently do not migrate.

We currently do not emulate.

To not preform an action on an object is an action in itself.

Is „covering all your bases“ a future proof approach? What can we do to achieve that?

Is there such a thing as „too much information“ about an object?

Use Cases

TIB Marstall Building - Hannover, Germany

Use Case 1: AV-portal

Starting point for this use case is the portal.

- AV-portal features are based on user-centred studies conducted by TIB
- semantic search, visual indexing and automatic classification are examples for key features identified
- A representation within the portal will contain different files to support the key features (i.e. transcription data for voice-over navigation, segmentation/temporal metadata for visual indexing, etc.)
- Two interesting findings regarding usablitiy:
  - *Community features in a scientific context were seen as potentially problematic by part of the focus group*
  - *The influence of portals such as YouTube/vimeo should not be underestimated*

Use Case 1: AV-portal

Technological:

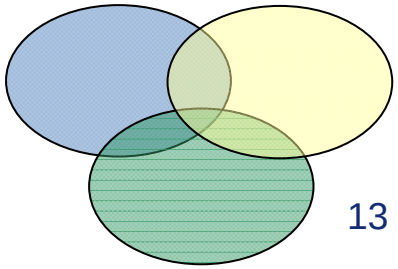
- The portal will be based on a combination of different tools and services
- A representation within the portal will be made up of different parts

Organizational:

- As a central service of our institution, we anticipate the portal to evolve over time (functionality and content wise)
- Our preservation intent is (currently) in the object on a content level

Usability:

- The key user expects familiar functionality, which will of course change over time.



Use Case 1: AV-portal

As the portal changes over time to facilitate the desired service to the users, the portal itself does not need to be preserved from an institutional viewpoint.

The preservation focus is on the object. As a complex/compound object, emulation can be a valid strategy for these object.

A concrete use case for this would be the integration of an emulator as a viewer for specific formats in the platform.

Metadata? Performance (over the web, in house)?

Use case 2: 3D objects

Starting point for this use case is the object.

TIB currently focuses on 3D objects from the architectural domain.
The single domain contains a wide variety of highly proprietary formats. The development/versioning cycle of CAD software is medium to high.

PDF/E is a common viewing representation, IFC a common data exchange representation.

Extension	Format	Number of Files
3dm	Quicktime 3D Metafile	1
3dm	Rhinoceros 3D Model	1
3dmf	Quicktime 3D Metafile	91
3ds	Autodesk 3D Studio	19871
c4d	CINEMA 4D	80
dwg	AutoCAD Drawing Object	70
fbx	Autodesk FBX	70
gsm	GDL Object	15109
lwo	LightWave 3D and Binary Object	70
max	3DS Max	1003
mb	Maya Binary	70
obj	Wavefront Object	87
skp	Google SketchUp Document	989
step	STEP Files	87
stl	Stereolithography File	3
wrl	VRML Worlds	293

Use case 2: 3D objects

Technological:

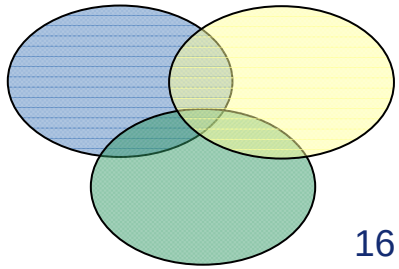
- heterogeneous, proprietary, complex file formats with little to no tool support
- Shortcomings of common export formats (e.g. lack of common metadata in ifc)

Organizational:

- 3D CAD software is highly proprietary
- As a library, TIB neither has the expert knowledge, nor the ressources to maintain all original CAD software

Usability:

- Data producers voiced need for a digital preservation service of the original files
- Comparability and browsing in a portal can best be achieved through a common export format



Use case 2: 3D objects

1. Library Viewer:

As the 3D formats evolve, it is unclear how authentic PDF/E representations of an object will be for older/newer formats. Rendering the PDF/E in an appropriate emulated environment is a valid usecase.

2. Data producer service:

The data producer voiced a need for digital preservation services. Emulation is a likely strategy for such a service. Best authenticity would most likely be achieved by emulating the producers machine. This would call for the need of an generation and ingest workflow in which the producer could easily generate a virtualization of their machine.

Metadata? PREMIS extension (BnF) to capture reading room environments? When is a good point to „freeze“ the reading room environment for emulation?
Is a data producer service something that could be realized? What would be the maintanence implications beyond bitstream preservation?

Questions, questions, questions ….

- Does anyone currently use emulation for AV materials or 3D objects?
- What kind of metadata should we capture now to have all bases covered for future strategies?

  As we are usually not the producer of the objects and have little influence on the information given to us with an object:

  What of the metadata needed can be / cannot be extracted from the object alone?

Image: http://iconbug.com/detail/icon/1128/blue-question-mark/

Answers, comments ….

….. or questions

Image: http://iconbug.com/detail/icon/3301/exclamation-point-key/


Save the Date
GOPORTIS
Leibniz Library Network for Research Information
Goportis Conference 2013 on Non-Textual Information
Strategy and Innovation Beyond Text
Hannover, Germany
18./19. March 2013
The Topics:
• Strategy and Policies
• Best Practices
• Innovation and Research
• Digital Preservation
www.goportis.de

TIB | German National Library of Science and Technology

Thank you for your attention!