Long term preservation within the digital library „UrMEL“ and the cooperation with the KOPAL team and IBM DIAS

Andreas Trappe
Scientist of Information and Media Technologie

Library of the Friedrich-Schiller-University Jena and State Library of Thuringia, Germany

Agenda

1. The library – ThULB
2. The digital library – UrMEL
3. Long term preservation
    1. General
    2. Current strategy in UrMEL
    3. Aimed solution

Agenda

1. **The library – ThULB**
2. The digital library – UrMEL
3. Long term preservation
    1. General
    2. Current strategy in UrMEL
    3. Aimed solution

The library - ThULB

- Library of the Friedrich-Schiller-University Jena and State library of Thuringia
- Size: about 4 Mio. physical print units
- Staff: about 180 employees

The library - ThULB

The library - ThULB

The library - ThULB

ThULB provides:

- ✓ own source material in great diversity
- ✓ materials from partner institutions
- ✓ project management
- ✓ restoration management
- ✓ digitalisation department
- ✓ University Multimedia Electronic Library (UrMEL)
- ✓ rooms for workshops / conferences

Agenda

1. The library – ThULB
2. **The digital library – UrMEL**
3. Long term preservation
    1. General
    2. Current strategy in UrMEL
    3. Aimed solution

Digital library UrMEL - Goals

UrMEL realizes the integration of diverse cultural heritage materials in one system. These materials can derive from our own collections or from partner institutions (libraries, archives, museums). Research projects benefit through

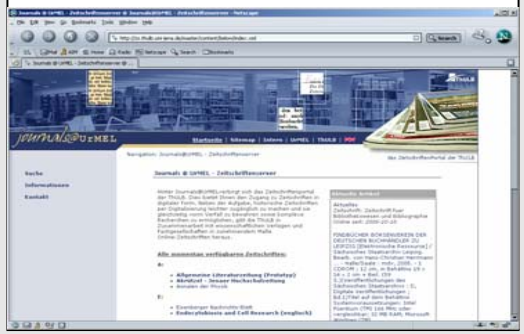
- digitization of huge collections of relevant sources
- virtual reconstruction and contextualization

Structure

http://www.urmel-dl.de

University @ UrMEL

Journals @ UrMEL



Collections @ UrMEL

Partners

DFG-Forschergruppe 530
„Selbstzeugnisse in transkultureller Perspektive“
Göttingen

Institut für Kirchen- und Dogmengeschichte
Hamburg

Biblioteka Jagiellonska
Kraków

Max-Planck-Institut zur Erforschung multireligiöser & multiethnischer Gesellschaften
Göttingen

Universität
Leipzig

Inst. für Dt. Presseforschung
Bremen

Vereinigte Domstifter zu Merseburg & Naumburg & des Kollegiatstifts
Zeitz (VDS)

Universität
Halle

Friedrich-Schiller-Uni
Jena
Institute & Einrichtungen

Herzogin Anna Amalia Bibliothek (HAAB)
Weimar

U r M E L

University Multimedia Electronic Library

Network

Thüringisches Hauptstaatsarchiv (ThHStAW)
Weimar

Stadtarchiv
Jena

Universität
Erfurt

Fachhochschule
Jena

Stadt
Bad Langensalza

Stadtbücherei & Stadtarchiv
Suhl

Universität
Ilmenau

Evangelisch-Lutherische Superintendentur
Jena

Vittorio Klostermann Verlag
Frankfurt/Main

Thüringisches Staatsarchiv
Rudolstadt

Stadtarchiv & Stadtverwaltung
Eisenberg

Types of documents

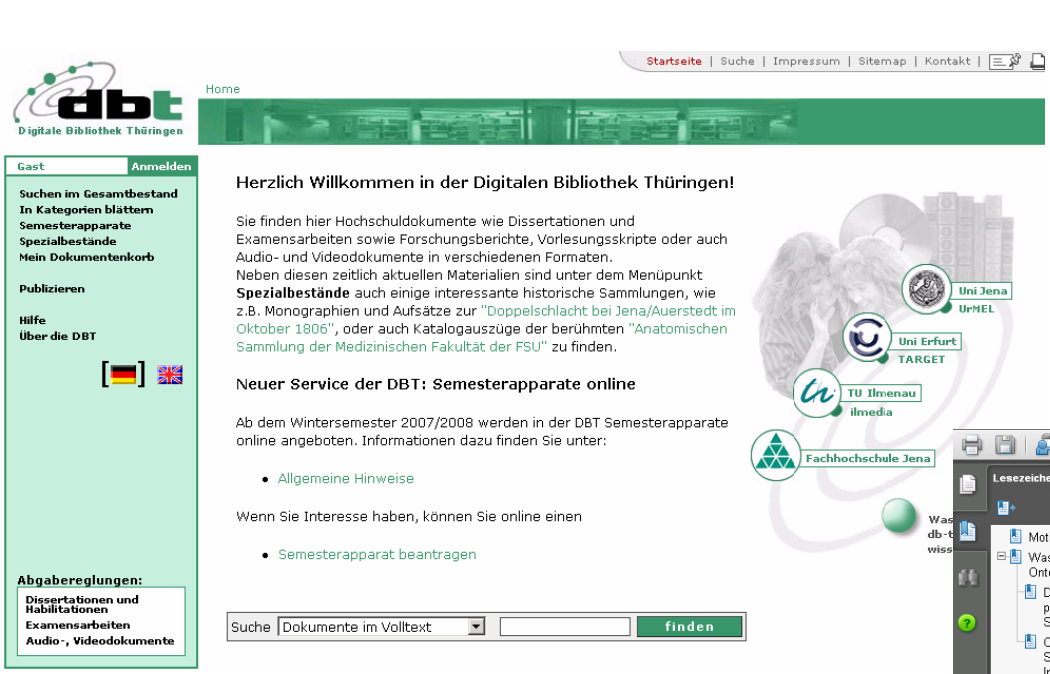
University@UrMEL

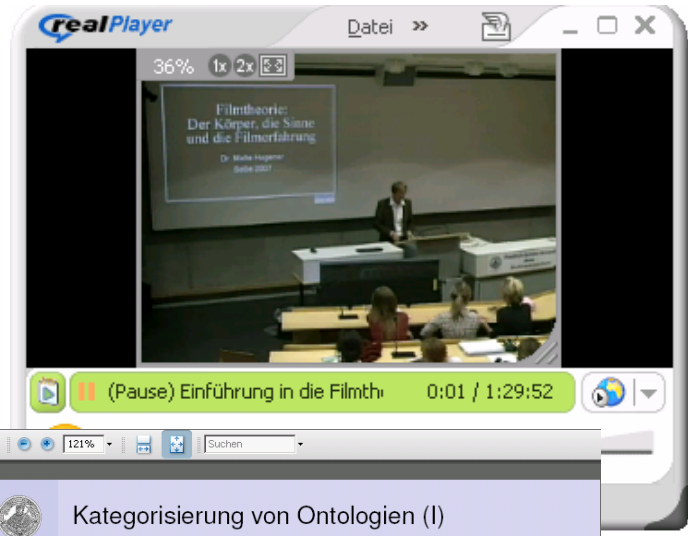
University publishing portal for multimedia teaching and research documents

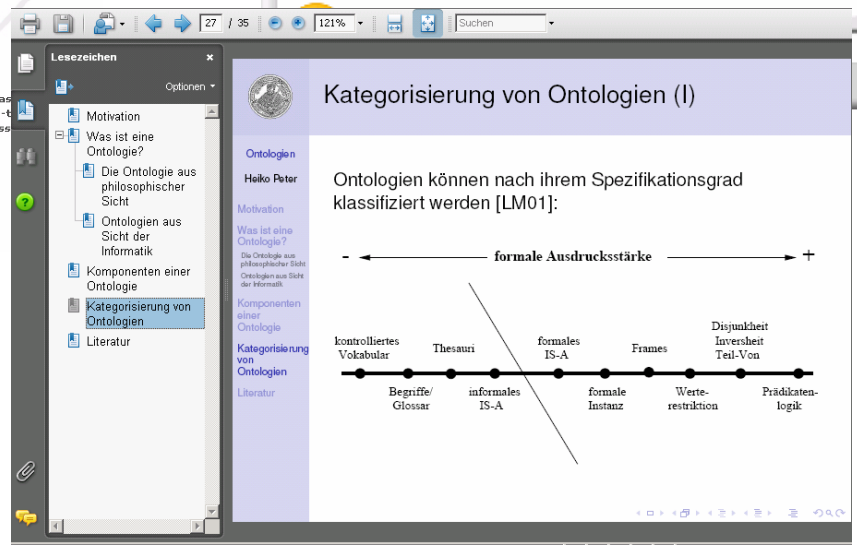
- like dissertations, diploma thesis, research reports and lecture notes
- Like recordings of lectures, congresses, educational videos as well as other multimedia documents
- Thematic collections
- Semester collections

Types of documents

University@UrMEL – How does it look like







Types of documents

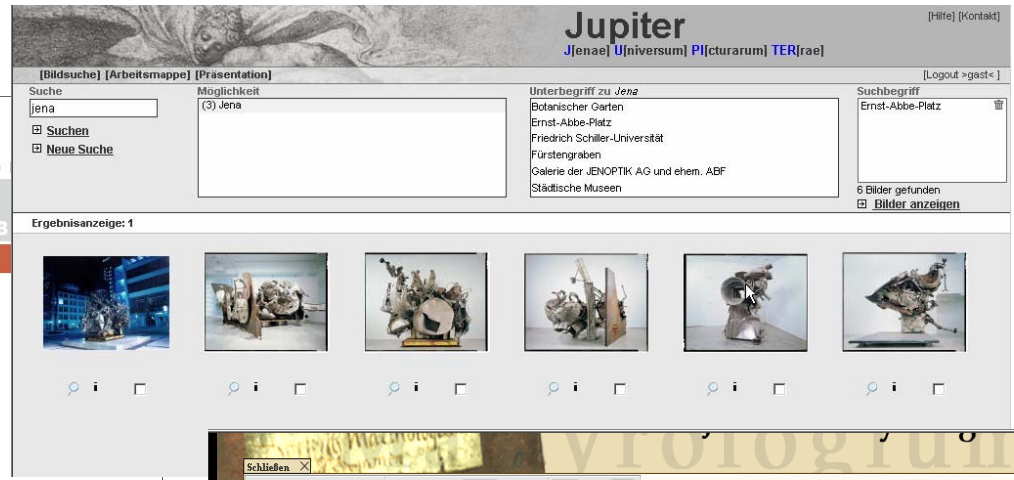
Collections@UrMEL

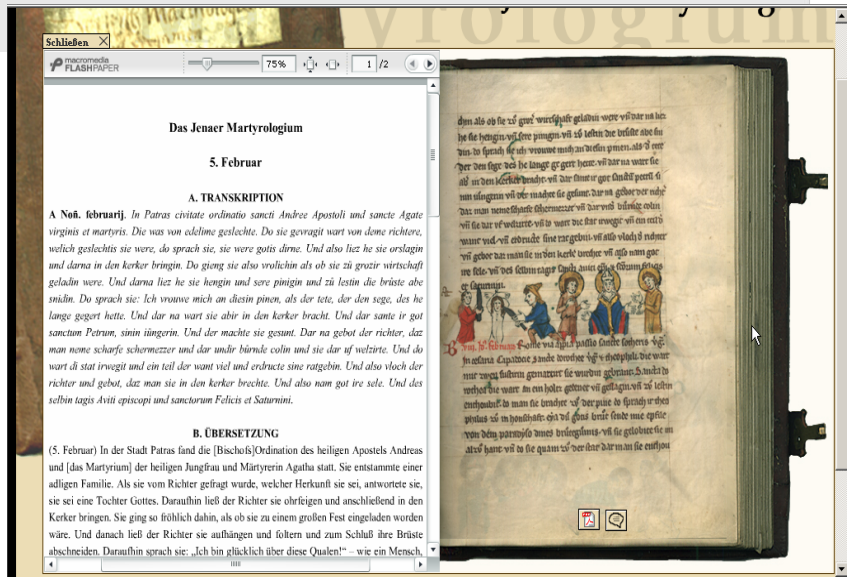
A portal for historic sources and collections

- Archive stocks
- Estates
- Manuscripts and –fragments
- Thematic collections (Papyri, Pictures ….)

Types of documents

Collections@UrMEL – How does it look like





Types of documents

Journals@UrMEL

E-Journals, will be covered by the next presentation…

Technical infrastructure

Hardware:

- Applications – 1 x SunFire 4200, 16 GB RAM, 2 x DualCore 2,6 GHz CPU‘s
- Database – IBM OpenPower 720, 16 GB RAM, 1,6 GHz Power 5 CPU
- Cent OS – Linux distribution
- Storage: currently 8 Tera Byte connected via Fibre Channel to Data Centre of the University of Jena

Technical infrastructure

Technical infrastructure

Software:

- **Java** as a platform independent programming language
- **Internet** applications ensure usage indepenent from geographical region and time
- **MyCoRe** as a common used content repository framework to build up digital library software solutions

Agenda

1. The library – ThULB
2. The digital library – UrMEL
3. **Long term preservation**
    1. **General**
    2. Current strategy in UrMEL
    3. Aimed solution

Long term preservation – Why ?

- Preservation of information and cultural heritage indepentent from time

- „Analogue“ information in terms of e.g. books are ageing because of environmental influences

- Digital information are much more sensitive and more difficult to achive

Long term preservation – Why ?

- different media and file formats
- different versions between same formats
- Operating systems change
- Hardware architectures change
- Data of medias might be corrupted by failure of storages
- Storage might be destroyed by some accident occurrences (natural influences or desasters)
- Et cetera

Methods of resolution - General

1. Migration

   Data migration is the process of transferring data between storage types, formats, or computer systems. To achieve an effective data migration procedure, data on the old system is mapped to the new system providing a design for data extraction and data loading. The design relates old data formats to the new system's formats and requirements.

2. Emulation

   An emulor duplicates (provides an emulation of) the functions of one system using a different system, so that the second system behaves like (and appears to be) the first system. This focus on exact reproduction of external behavior is in contrast to some other forms of computer simulation, which can concern an abstract model of the system being simulated.

Methods of resolution - Detailed

- Transformation of formats, if possible automatically
- Migration between different versions
- Ensure storage solutions run fail-safely
- Reduce risk of data loss by storing it on different and independent geographical located systems
- Use common standards for metadata formats

Agenda

1. The library – ThULB
2. The digital library – UrMEL
3. **Long term preservation**
    1. General
    2. **Current strategy in UrMEL**
    3. Aimed solution

Current strategy in UrMEL

- Data are mirrored periodically
- Differentation between data of:
  - Operating system
  - Software of applications (MyCoRe)
  - Data base (IBM DB2)
  - Content itself (journals, collections, image scans, etc.)
- Data where backed up for 3 weeks and might be recovered within this time frame

Problems with this strategy

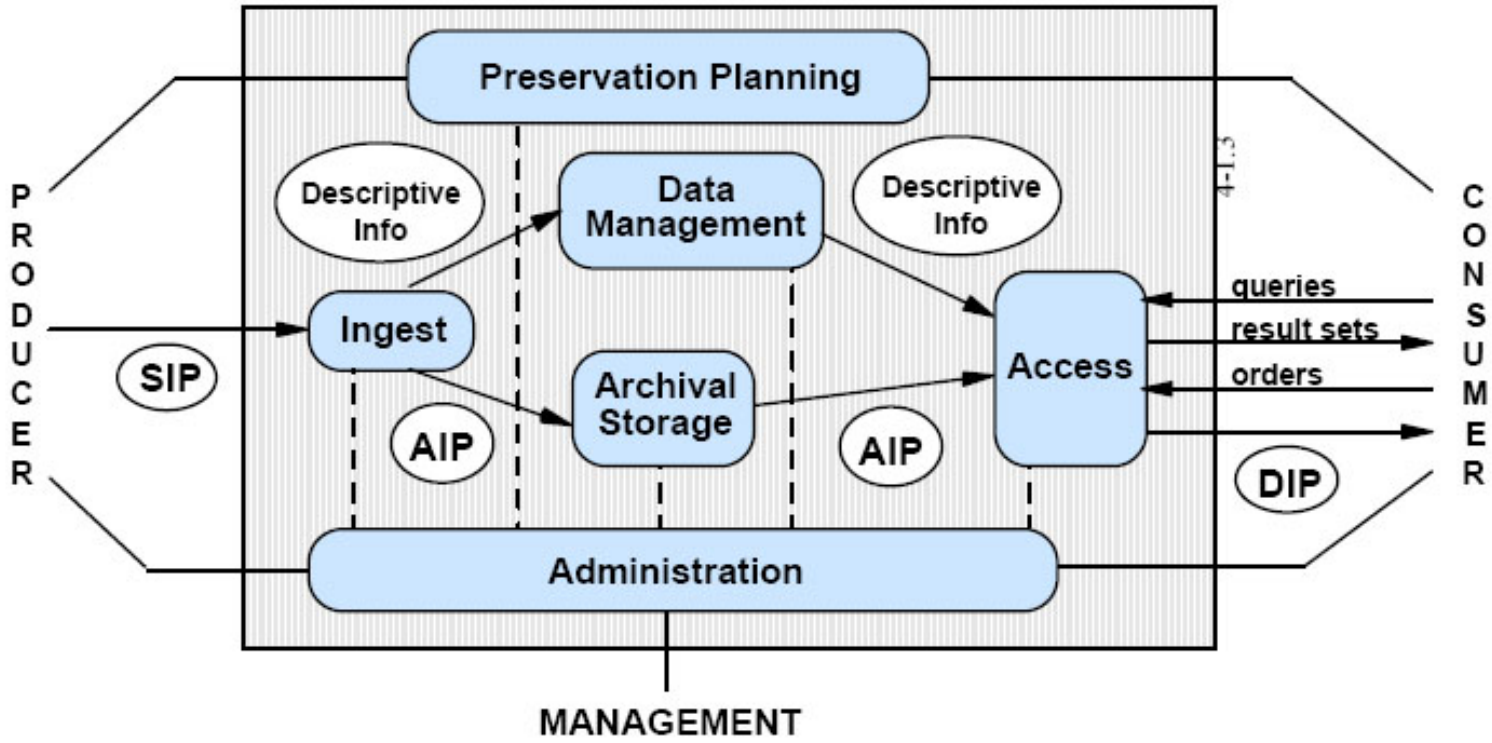
- Loss of data after 3 weeks
- Different media formats and versions are not migrated → possibly unreadable for newer players or clients
- Data are not stored in form of logical objects → increased expense to recover repositories
- Use of proprietary data models → transformation to other formats might be difficult
- No redundant storage → increases risk of data loss

Agenda

1. The library – ThULB
2. The digital library – UrMEL
3. **Long term preservation**
    1. General
    2. Current strategy in UrMEL
    3. **Aimed solution**

Aimed solution for UrMEL

Implementation of OAIS-Model :



Aimed solution for UrMEL

Implementation of OAIS-Model by:

- Joint use and customization of system „KOPAL“
- Cooperation between German National Library, Lower Saxonian State and University Library Göttingen, Association For Scientificaly Data Processing and IBM Germany
- KOPAL consists of a software kernel „DIAS“, developed by IBM and extensions according to the OAIS-Reference-Model
- Applications will be integrated using „KoLibRi“ software components
- 3 models of use are:
    - „Participant“ – provisional archiving of data by a KOPAL-Tenant
    - „Tenant“ – autonomous hosting of a KOPAL-Archiving-System, without administration of servers itself
    - „Own operation“ – complete autonomous hosting of a KOPAL-Archiving-System and servers, too

Aimed solution for UrMEL

- Seperation of UrMEL into presentation- and archive system
- Archive system build by KOPAL, presentation system will keep MyCoRe
- Data for long term preservation will be migrated into KOPAL
- Creation of logical data packages (1 journal, 1 collection, etc.)
- Usage of METS-Format „Metadata Encoding and Transmission Standard“
    - The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium.

Aimed solution for UrMEL

Organisational realisation:

• GRF-Application of Lower Saxonian State - and University Library Göttingen and State and University Library Jena to extend KOPAL as tenant supporting system and support in licence charges
• Another GRF-Application to develop a connector in KoLibRo

Advantages for UrMEL:

• Real long term data preservation
• Data can be read and understood in the remote future
• Ensurance of preservation of cultural heritance
• Library acts as provider of scientific information indepentent from time and region