Mecklenburg Yearbooks - a Digital Library Project

Ilvio Bruder

Database Research Group
Department of Computer Science
University of Rostock

What are Mecklenburg Yearbooks?

- Collection of books about Mecklenburg classical antiquity and history between 1836 and 1940
- Edited by different persons from the association of classical antiquity and the association of history, e.g., Friedrich Stuhr
- Published by different publishers, e.g., Bärensprungsche Hofbuchdruckerei
- One book = a collection of papers, reviews and an annual report
**Project: Mecklenburg Yearbooks - Architecture**

- Internet
  - Search, Navigation
  - Web Information Portal
  - Distributed Search
  - Yearbooks

**Digital Library (CM7, Miless)**

- Yearbook data model
- Digitalized raw data import
- GVK

**Creation of a DublinCore-based Metadata Structure**

- Creation of a list of requirements, descriptions, modeling ideas
- Analysis of the books and their structure
- Specification of DC-fields with bibliographic experts
- Examination of models and schemes: XHTML, DC, ISO 12083 DocBook
- Building an integrated yearbook model from the models above
- Implementation possibilities in Miless/MyCore
Properties of the Yearbook Document I

- Different volumes consist of different parts,
- Mini-collections successive volumes: e.g. till YB1930 two different book parts for history and classical antiquity
- Bound yearbooks and 2 additional books
- Possible book parts:
  * Table of contents of one/more?? parts
  * YB for history
  * YB for classical antiquity
  * Annual report/quarterly report
  * Advertising
  * Pictures (appended)
- Yearbooks and reports consists of:
  * articles (possibly continued at next book)

Properties of the Yearbook Document II

- Essay/paper/article is subdivided in:
  * Title/author (front matter)
  * Pictures - within text
    - at own page but without page number
    - unfoldable sketches
    - related pictures at the end of the yearbook
  * Tables - not normalized -> to digitalize as picture and text
  * Footnotes - many possibilities
    (half page, one footnote at two pages)
  * Special chars - to mark + own font and include as XML-entities)
  * Annotations - handwritten extensions
Jahrbücher
des
Vereins für mecklenburgische Geschichte
und Alterthumsfunde.

gegründet von Friedrich Cohn,
Fortgeführt von Friedrich Wagner und Hermann Blankenhorn.

Neugegründet und herausgegeben
von der
Staatsarchivverwaltung der Provinz Mecklenburg.

Mit angehängtem Jahresbericht.

Schwerin. 1926.
Druck und Vertrieb der Mecklenburgischen Stadtschriften-Druckerei.

Die Übergangung vorliegen war, dass die rechtliche Macht im direkten
Gefolge der Verfügungen über das Besitzrecht besteht und dass die
Verwaltung der Stadtschriften nicht nur in der Regel, sondern
auch in der Praxis solche Verfügungsgewalt besitzt.

Die Bedeutung der Verwaltung der Stadtschriften geht aus den
vorliegenden Berichten hervor, die die Entwicklung der Stadtschriften
erleichtern. Die Entwicklung der Stadtschriften wird in den
Verwaltungsberichten der Stadtschriften in den Jahren 1870 und 1871
nachzuweisen sein. 

Further Demands on the Data Model

- Images for each page: What should happen with empty pages?
- Necessary marks in the fulltext: annotations, fonts, pagebreaks, tables (text with link to the image), images, ...
- What bibliographic data (metadata) does the library already register and what could be automatically recognized by OCR or by additional data analysis?
- Which picture size, how to store?
  - 400dpi (OCR) or 100dpi (display friendly) or suitable for web
- Note the structure of the new, current yearbook (not free)
- Special web representation for the yearbook collection
Scheme and Model of Yearbooks I

- Yearbook as a whole XML document or separated?
  - Store the whole yearbook as structured text document (XML) and also as parts internal
  - References by id/idref relations (XML)
  - Separation of content, structure and layout
- YB scheme is combined from different schemes:
  - Collection scheme consists of collection definition and/or the assignment of bibliographic objects like books (e.g. Mecklenburgika/Meekl. YBs/YB01)
  - Book scheme consists of metadata (Dublin core) and structure to a specific degree (above level of the docbook scheme of the DocBook Initiative)
  - actual content is defined in the fulltext scheme

Scheme and Model of Yearbooks II

- DC scheme (Dublin Core without restrictions)
  principally, all fields in DC are optional and as many as you like, follow restrictions are internally silent defined:
  title+ (1 bis n), creator+ (1 bis n), subject* (0 bis n), description* (0 bis n), identifier (1), publisher* (0 bis n), contributor* (0 bis n), type? (0 bis 1), format+ (1 bis n), date+ (1 bis n), source? (0 bis 1), lang* (0 bis n), relation* (0 bis n), coverage? (0 bis 1), rights+ (1 bis n)
- Further metadata schemes (not yet specified, e.g. physical properties)
- Fulltext scheme consists of a subset of XHTML and extensions for pagebreaks, footnotes and tables as pictures + keywords
Tree Diagram of the Book Scheme

Dublin Core of a Mecklenburg Yearbooks Paper

- **title**: Title of the article, subtitle
- **alt title**: Title of the yearbook
- **creator**: Author of this article
- **subject**: Mecklenburgica collection, keywords of the article
- **description**: Text, coarse description of the content
- **identifier**: LBMV-33-1234 (signature)
- **publisher**: Bärensprung'sche Holzbuchdruckerei (e.g.)
- **contributor**: Friedrich Stuhr (e.g.)
- **type**: Article
- **format**: XML, TXT, PNG
- **date**: xxxx
- **source**: Collection Mecklenburg Yearbooks, LBMV
- **language**: DE
- **relation**: Coverage
- **coverage**: Temporal and local coverage of the article
- **rights**: State library Mecklenburg Western Pomerania (LBMV)

*green: metadata + possible specific data from the catalog*
Possibilities of the Miless Data Model

- Possibilities: classifications, Dublin core related metadata, Derivate/DerivateParts
- Digitalized documents are one XML file per yearbook + images
- XML file is separated in bibliographic-relevant parts, with redundancy
  → per YB file: whole yearbook,
    yearbook of history,
    yearbook of classical antiquity (possibly?),
    articles 1..n
  1 annual report or 4 quarterly reports

Classifications in Miless

- existing classifications are: FORMAT - text, audio, video
  TYPE - article, book, dissertation
  ORIGIN - university, institution, library

- new classification for Mecklenburg Yearbooks:
  book - yearbook - article
  - report
  - volume

```
  Mecklenburg Yearbooks
     /\          /\          /\          /\          /\          /\          /\
       \         \         \         \         \         \         \         \
       yearbook report    vol001    ...    vol104

```
### Example of a Problematic Inquiry

- Inquiry: Show all yearbooks including articles with title "..."!
- Problem: to recognize relation between yearbook and article
- Possible solutions:
  - secret implementation, misuse DC field
  - inherit all DC-fields from articles to the yearbook
  - create category volume, determine intersection of yearbook and volume of the found articles

### Mapping of Dublin Core to Miless Metadata

<table>
<thead>
<tr>
<th>Title</th>
<th>Mecklenburg Yearbooks, Volume xx, ...</th>
<th>titles.maintitle</th>
<th>titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>yearbook of the association of history</td>
<td></td>
<td>creators</td>
</tr>
<tr>
<td>Creator</td>
<td>association of Mecklenburg history</td>
<td>creator</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Mecklenburgica, history, association, annual report</td>
<td>subjects</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>text, coarse description of the content</td>
<td>descriptions</td>
<td></td>
</tr>
<tr>
<td>Identifier</td>
<td>LBMV-33-1234 (signature)</td>
<td>idescription</td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td>Bärensprungsche Hofbuchdruckerei</td>
<td>publishers</td>
<td></td>
</tr>
<tr>
<td>Contributor</td>
<td>Friedrich Stühr (e.g.)</td>
<td>contributors</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>bound yearbook, journal series</td>
<td>types (category)</td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>XML, TXT, PNG</td>
<td>formats (category)</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>xx.xx.xx.xx</td>
<td>dates (different types)</td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>collection of Mecklenburg yearbooks, LBMV</td>
<td>sources</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>DE</td>
<td>origin (category)</td>
<td></td>
</tr>
<tr>
<td>Relation</td>
<td></td>
<td>languages</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>18xx, Mecklenburg, Germany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rights</td>
<td>Landesbibliothek Mecklenburg Western Pomerania</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Derivate and DerivateParts

- per yearbook: 1 XML file + images, x pages as PNG image
- per DC described object from the yearbook:
  - 1 XML document
    - 1 XML file
    - n belonging pictures
  - 1 image document
    - m pages PNG images
  - 1 TXT document (conform to XML)
    - 1 TXT-File
- possible further generated derivates (e.g. PDF, HTML)

Further Information and Future Work

- Project homepage (in german):
  www.uni-rostock.de/ub/UB_PRO_DIGB_MJ.HTM
- Test server not public available yet
- Mecklenburg yearbooks will be part of the digital library of the university of Rostock
- Another collection the gazette „Ratzeburger Anzeiger“ are digitalized and will be included in the digital library