The Digital Knowledge Store
of the Berlin-Brandenburg Academy of Sciences and Humanities

Possibilities and Challenges of linking heterogenous Humanities Research Data and Metadata

Marco Jürgens & Sascha Grabsch
Introduction

The Digital Knowledge Store is:

- a data research infrastructure developed by TELOTA
- initialized by the BBAW in 2009
- funded by the Deutsche Forschungsgemeinschaft (DFG) in 2011
Introduction

The Digital Knowledge Store is:

- a data research infrastructure developed by TELOTA
- initialized by the BBAW in 2009
- funded by the Deutsche Forschungsgemeinschaft (DFG) in 2011
Presentation Overview

1. Motivation for the development of a Digital Knowledge Store

2. Current project status

3. Future development
The Motivation for a Knowledge Store

- the BBAW hosts various humanities long-term projects

- research results are published digitally and include a variety of different resource types e.g. medieval charters, ancient inscriptions, manuscripts, etc.

- the necessity of a centralized access in order to unify and connects the various resources with regard to metadata and content
Goals

Aims to:

• offer a centralized access to all digital resources of the BBAW
• structure, organize and present knowledge online
• make implicit information explicit
• mediate knowledge online
• realize the promise of Open Access
Current Status of Development

- A prototype was created and first presented at DH2012 (Hamburg)
- Creation of a central index (Apache Lucene)
- Fulltext retrieval of multilingual resources and various formats (e.g. German, Ancient Greek, Arabian; PDF, HTML, XML/TEI)
- Basic semantic knowledge extraction
- DONATUS Morphology
Advancements

- Improved index fields like *content abstract*, *DDC*

- 9000 Ressources integrated from 9 Projects

- A first set up for a Metadata System
Metadata System

- is set up in two ways: from automatically and 'human-' created Metadata

- modelled in a flexible Schema: OAI-ORE

- BBAW-projects identifiable by the use of named graphs
Metadata System

- embedded in the *Default Graph* the resources exist/are available as a semantic net

- a base for the representation of linked and (in regards to content) overlapping resources

- Software components like Sparql-Adapter or Concept-Identifier allow powerful querying
Modelling Metadata in OAI-ORE

• Why OAI-ORE?

• Problems in modelling with OAI-ORE

```xml
<rdf:Description rdf:about="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/suche.xml">
    <dcterms:hasVersion rdf:resource="http://pom.bbaw.de/avh/index.php#suche.xml"/>
</rdf:Description>

<rdf:Description rdf:about="http://pom.bbaw.de/avh/index.php#suche.xml">
    <dcterms:isVersionOf rdf:resource="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/suche.xml"/>
</rdf:Description>

<rdf:Description rdf:about="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/literaturverzeichnis.xml">
    <dcterms:hasVersion rdf:resource="http://pom.bbaw.de/avh/index.php#literaturverzeichnis.xml"/>
</rdf:Description>

<rdf:Description rdf:about="http://pom.bbaw.de/avh/index.php#literaturverzeichnis.xml">
    <dcterms:isVersionOf rdf:resource="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/literaturverzeichnis.xml"/>
</rdf:Description>

<rdf:Description rdf:about="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/ortsverzeichnis.xml">
    <dcterms:hasVersion rdf:resource="http://pom.bbaw.de/avh/index.php#ortsverzeichnis.xml"/>
</rdf:Description>
```
Modelling Metadata in OAI-ORE

- Why OAI-ORE?
- Problems in modelling with OAI-ORE

```xml
<rdf:Description rdf:about="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/suche.xml">
  <dcterms:hasVersion rdf:resource="http://pom.bbaw.de/avh/index.php#suche.xml" />
</rdf:Description>

<rdf:Description rdf:about="http://pom.bbaw.de/avh/index.php#suche.xml">
  <dcterms:isVersionOf rdf:resource="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/suche.xml" />
</rdf:Description>

<rdf:Description rdf:about="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/literaturverzeichnis.xml">
  <dcterms:hasVersion rdf:resource="http://pom.bbaw.de/avh/index.php#literaturverzeichnis.xml" />
</rdf:Description>

<rdf:Description rdf:about="http://pom.bbaw.de/avh/index.php#literaturverzeichnis.xml">
  <dcterms:isVersionOf rdf:resource="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/literaturverzeichnis.xml" />
</rdf:Description>

<rdf:Description rdf:about="http://telota.bbaw.de:8087/exist/servlet/db/AvHBriefedition/Briefe/ortsverzeichnis.xml">
  <dcterms:hasVersion rdf:resource="http://pom.bbaw.de/avh/index.php#ortsverzeichnis.xml" />
</rdf:Description>
```
Current areas of development

- Knowledge Browsing
- Modelling in OAI-ORE
- Combining Metadata and Full Text
Future development

• “Knowledge Browsing” extends the research results by offering semantic recommendations

• the process of searching as an interaction between contents and user
Future Development

- Creation of knowledge domains by:
  
  Automated semantic interpretation

  Evaluation of previous search results

  Feedback and statistical evaluation