2016 Chinese Institutional Repository Conference

Establishing Open Access Services in the Global Repository Network

Experiences and Challenges from the Service Provider Perspective (BASE)

Friedrich Summann
Bielefeld University Library, Germany
Overview

• Establishing a Service in the Repository Network: BASE

• The Repository Network - past and current situation

• Current and Future Activities
Bielefeld University, Germany

- Founded 1969
- 24,000 Students,
- 4,500 Staff

University Library

- 2.2 Million Printed Books, 500,000 Digital Media
- 22 Faculty Libraries
- Automation from the beginning in 1968
- Many innovative funded Digital Library Projects
Some BASE Milestones

- 2001 Starting point as a search engine follow-up for a metasearch system
- 2004 Official Start (FAST Data Search)
- 2006 starting participation in EU projects
- 2011 Switch to open source (Lucene/Solr, VuFind)
- 2012 OAI-PMH-Interface, data delivery of subject sections
- 2014 OA-boosting
- 2015 OA status and License information processing
BASE, OAI Service Provider and Mirror of the IR landscape

- 4622 Repositories included (using OAI-PMH)
- From 111 Countries world-wide
- Ca. 99 Mill. Documents
- Ca. 70 % Open Accessible
- Dublin Core Format
- Ca. 14.7 Mill. Documents enriched with DDC-Code (Dewey)
The BASE scope

- OA Repositories world-wide
- Academic-valuable Contents
- Focus on Institutional Repositories
- Aggregators (RePEc)
- Subject Repositories (arXiv, CiteSeerX etc)
- Electronic Journals
- Digital Collections
- Dataset Repositories
Contents:
Global Scientific Metadata harvested from Repositories via OAI-PMH

Technology:
• Index (Lucene/Solr)
• Search Interface (VuFind)
• APIs (HTTP, OAI)

BASE harvests, aggregates, enriches and exposes OAI Metadata
关于BASE：统计

BASE自2004年9月起发展为数众多的索引来源和文献。每两个月更新如下图表和表格。

图表

Chart: Number of indexed Documents and Sources
BASE Restrictions for including repositories

• OAI-PMH interface (available)
• OA is supported (at least some documents should be open accessible world-wide!)
• DC Metadata (basic set of field contents)
• Operable URL for the document landing page
Overview

• Establishing a Service in the Repository Network: BASE

• The Repository Network - past and current situation

• Current and Future Activities
The IR Network – Participants and Stakeholders

• IR as Data Providers
• Service Providers (OAIster, BASE, Google Scholar, ResearchGate etc.)
• Registries (OpenDOAR, OpenArchives, ROAR ...)
• Projects (national, international)
• Organizations (national, international)
• Funder Organizations
• Publishers, Information Companies
• Users, Authors, related Institutions
The Repository Network:

Documents per Country
The Repository Network:

OA Documents per Country (IRs only)
Overview

• Establishing a Service in the Repository Network: BASE
• The Repository Network - past and current situation
• Current and Future Activities
Concrete Current Activities

- Refinement Defining Publication Types
- CrossRef Integration (OA contents)
- ORCID Handling
  - Detecting ORCIDs in Metadata
  - Publication Claiming
  - Feeding to ORCID Publication List
- Developing Repository Network Tools
4. Validation indices for projective clustering

<table>
<thead>
<tr>
<th>Title:</th>
<th>Validation indices for projective clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author:</td>
<td>Chen, Lifei ; He, Shanjun ; Jiang, Qingshan ; 姜吉山</td>
</tr>
</tbody>
</table>
| Description:                | National Natural Science Foundation of China [10771176], National Science Foundation of Fujian Province of China [2009]01273] ; Cluster validation is a major issue in cluster analysis of data mining, which is the process of evaluating performance of clustering algorithms under varying input conditions. Many existing validity indices address clus...  

Show all

<table>
<thead>
<tr>
<th>Publisher:</th>
<th>FRONT COMPUT SCI CHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of Publication:</td>
<td>2009-12</td>
</tr>
<tr>
<td>Source:</td>
<td><a href="http://dx.doi.org/10.1007/s11704-009-0051-1">http://dx.doi.org/10.1007/s11704-009-0051-1</a></td>
</tr>
<tr>
<td>Document Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Language:</td>
<td>en_US</td>
</tr>
<tr>
<td>DDC:</td>
<td>006 Special computer methods (computed)</td>
</tr>
<tr>
<td>URL:</td>
<td><a href="http://dspace.xmu.edu.cn:8080/dspace/handle/2288/90340">http://dspace.xmu.edu.cn:8080/dspace/handle/2288/90340</a></td>
</tr>
<tr>
<td>Content Provider:</td>
<td>Xiamen University Institutional Repository</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>URL:</th>
<th><a href="http://dspace.xmu.edu.cn/">http://dspace.xmu.edu.cn/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Continent:</td>
<td>Asia</td>
</tr>
<tr>
<td>Country:</td>
<td>cn</td>
</tr>
<tr>
<td>Latitude / Longitude:</td>
<td>24.445500 / 118.070000 (Google Maps</td>
</tr>
<tr>
<td>Number of documents:</td>
<td>124,135</td>
</tr>
<tr>
<td>Open Access:</td>
<td>unknown</td>
</tr>
<tr>
<td>Type:</td>
<td>Elektronische Hochschulpublikationen</td>
</tr>
<tr>
<td>System:</td>
<td>DSpace</td>
</tr>
<tr>
<td>In BASE since:</td>
<td>2010-02-23</td>
</tr>
<tr>
<td>BASE URL:</td>
<td><a href="https://www.base-search.net/Search/Results?q=dccoll:ftxiamenuniv">https://www.base-search.net/Search/Results?q=dccoll:ftxiamenuniv</a></td>
</tr>
</tbody>
</table>
China, People's Republic of

276309 Documents from 12 Repositories

This means: 0.29 % of Documents world-wide, 0.28 % of Repositories world-wide
Chinese Repositories: Publication Type Distribution

Publication Types

- Unknown: 52.9%
- Article: 27.3%
- Thesis: 18.6%
- Report: 0%
- Video: 0%
- Other: 0%
Chinese Repositories: OA Status

Document OA Status

- 94.8% unknown
- 5.2% OA
Chinese Repositories: Language Distribution

Document language (Number of documents)

- chi: 69.9%
- unknown: 14.7%
- eng: 14.6%
- cha: 0%
- Other: 0%
IR in the Information Environment
Issue: OA Status, Rights and Licences Normalization

Nachnutzung/Lizenzen

- Alle
  - Creative Commons
    - CC-BY
    - CC-BY-SA
  - Public Domain
    - CC0
    - Public Domain Mark (PDM)

Zugang

- Open Access
- Kein Open Access
- Unbekannt
Issue: Data Enrichment (Linked Open Data Strategies)

Fundamental Aspect:

• More **Detailed Metadata Formats** have the potential to provide more detailed information

• But: It depends on the background quality
Metadata Formats in OAI-PMH Repositories

Sales, Expenses, and Profit: 2014-2017

- mets/mods
- marc
- rdf
- didl
- qualified_dc
- rfc1807
- ore

Number of Repos.
Number of Docs
Issue: Identifiers

- OAI Identifier
- DOI/Handle/ISSN/ISBN/URN/PMCID
- Author/Organization/Funder IDs
Future Issue: Big Data Activities

• Data Enrichment  
  (Linked Open Data Strategies)
• Automatic Classification
• De-Duplication/Version Detection
• Fulltext Indexing
Thank you for your attention!

• friedrich.summann@uni-bielefeld.de