Paving the road for Open Science strategies and services via the global Repository Network

Pragmatical approaches for a trans-national sharing of experiences, expertise and contents between China and Germany

Open Repositories Conference 2017
Brisbane, Australia
Outline

• The Partners
• Collaborative Approach
• The Results
• Conclusion
Outline

• The Partners
• Collaborative Approach
• The Results
• Conclusion
Bielefeld University, Germany

- Founded 1969
- 24000 Students,
- 4500 Staff

University Library

DINI, COAR member
DRIVER /OpenAire Project Member
DINI Certificate, DRIVER/OpenAire Guidelines
COAR Vocabulary Group
BASE Search Engine
Publication Management System incl.
RDM Service, E-Journals, Digital Collection
National Science Library, Chinese Academy of Sciences (NSL, CAS)
Services for 125 CAS institutes over 30 cities across China

- 500 staff
- 11.5 million items
- 80,000m² building floor capacity

NSL
- organized the 8th Berlin Open Access Conference
- organizes China Open Access Week each year
- supports 125 CAS institutional repositories
- leads China IR Implementation Group
- supports CAS authors to publish in OA journals
- participates in the Sponsoring Consortium for SCOAP3
- 11 members in various section standing committees of IFLA
- member in COAR and also a member in EIFL.
BASE (Bielefeld Academic Search Engine)

• Search Engine established in 2004
• OA Service Provider using mainly OAI-PMH
• Focused on Academic Content
• 5600 Repositories index with metadata for 112 Mill. Publications
• Providing Network Services (Retrieval API, OAI-PMH)
Outline

• The Partners
• Collaborative Approach
• The Results
• Conclusion
Germany

13532749 Documents from 354 Repositories
12% of Documents world-wide, 6% of Repositories world-wide

OA Status Distribution (Number of Documents)
- OA
- Unknown
- Non-OA

Repository Type Distribution (Number of Documents)
- IRs
- Journals
- DCols
- RData
- Other
The Academic Landscape in China

• Around 2000 Universities
• In total 2,553 National General Colleges and Universities (2015)
• The number of enrolled college students including undergraduate students, master and PhD students was 23.91 million in 2012.
• Many Institutional Repositories active, but only some visible
Joint activities Bielefeld UL – CAS NSL since 2009

2009 DRIVER Harvesting/Indexing Activity

2014 Sino-Germany Open Publishing Publishing Workshop
http://ir.las.ac.cn/handle/12502/6739

2014 China Institutional Repository Conference
http://ir.las.ac.cn/handle/12502/7388

2016 COAR GA Meeting Vienna

2016 China Institutional Repository Conference
http://ir.las.ac.cn/handle/12502/8749

2016 Dec Three Hour Project Meeting at Kassel Railway Station (Germany)

2017 BASE and CIRG are doing a monthly VC on Friday since 2017.

2017 COAR GA Meeting Venice
The Current BUL Activities – concrete in Detail

• Checking the Repository Infrastructure
• Vocabulary Extension
• Preparing Indexing
• Technical Support for Developing Network Services
• Sharing Technical Experience And API s
BASE shares expertise with CIRG about the OpenAIRE and DINI Guidelines as good practices

- **Standardised Usage Statistics** for Open Access Repositories and Publication Services
- Forschungsinformationssysteme in Hochschulen und Forschungseinrichtungen
- DINI-Zertifikat für Open-Access-Repositorien und -Publikationsdienste 2016
Checking and Preparing Repositories for Indexing

Changchun Institute of Applied Chemistry - Chinese Academy of Sciences
中国科学院长春应用化学研究所机构知识库
Xi'an Jiaotong University: XJTU Academic Hub / 西安交通大学机构知识库

221952 Documents, 0 Open Accessible

Repository Type: Hochschulpublikationen
System:
In BASE indexed since: 2017-05-23

OA Status Distribution (Number of Documents)
- Green: OA
- Gray: Unknown
- Red: Non-OA

Number of Documents:
- 0 in 1948
- 2,000 in 1950
- 4,000 in 1960
- 6,000 in 1970
- 8,000 in 1980
- 10,000 in 1990
- 12,000 in 2000
- 14,000 in 2010
- 16,000 in 2015
- 18,000 in 2020
- 20,000 in 2025

Years: 1948 to 2017
Metadata Quality Improvements
Extracting a list of publication types to prepare mapping in BASE

- 期刊论文 - 494567
  会议论文 - 113918
  学位论文 - 84768
  博士 - 43996
  专利 - 38955
  中文期刊论文 - 37380
  硕士 - 35567
  发明 - 23657
  专著 - 9331
  实用新型 - 8060
  专著章节/文集论文 - 4960
  成果 - 4394
- EI期刊论文 - 3916
- 其他 - 3469
- 发明专利 - 2102
- 发明授权 - 2030
- 图像 - 1513
- 工学博士 - 1339
- 文集 - 1198
- 博士后 - 1154
- 工学硕士 - 1007
- 演示报告 - 855
- 研究报告 - 787
- 传媒扫描 - 785
- 项目 - 746
BASE supports OpenSearch Engine (based on Retrieval API or Metadata Provision)
Improving the Repository Manager Information

Golden Rules for Repository Managers

We are indexing all kinds of academically relevant content. In case your source does not provide a

However, the best way to get your documents indexed is to make sure your source data is up-to-date.

You can check some of the following items using our:

- OAI interface working
  - Your OAI interface is stable and responds to requests.
  - Otherwise, it is not possible to index your oai.

- Comprehensive metadata
  - Each item exposed via your OAI interface provides metadata.
  - If important metadata is missing, documents are not indexed.

- Identifier (URLs) working
  - Each item provides a URL in the .id/.identifier that should lead to the front door.
  - Only documents whose identifiers start with "http" are indexed.

- Providing access information (Open Access)
  - Access information of the fullest extent should be provided by metadata record in the field "setSpec".
  - If correct access information is missing, these documents are not indexed.

- Providing information concerning reuse / li
  - Authors can publish their work under a Creative Commons license.
  - All content exposed via your OAI interface should be under a Creative Commons license.

- Character encoding
  - All content exposed via your OAI interface should be in UTF-8 encoding.

Knowledge Managers’ Golden Rules

We are indexing all kinds of academically relevant content. In case your source does not provide a

However, the best way to get your documents indexed is to make sure your source data is up-to-date.

You can check some of the following items using our:

- OAI interface working
  - Your OAI interface is stable and responds to requests.
  - Otherwise, it is not possible to index your oai.

- Comprehensive metadata
  - Each item exposed via your OAI interface provides metadata.
  - If important metadata is missing, documents are not indexed.

- Identifier (URLs) working
  - Each item provides a URL in the .id/.identifier that should lead to the front door.
  - Only documents whose identifiers start with "http" are indexed.

- Providing access information (Open Access)
  - Access information of the fullest extent should be provided by metadata record in the field "setSpec".
  - If correct access information is missing, these documents are not indexed.

- Providing information concerning reuse / li
  - Authors can publish their work under a Creative Commons license.
  - All content exposed via your OAI interface should be under a Creative Commons license.

- Character encoding
  - All content exposed via your OAI interface should be in UTF-8 encoding.

Knowledge Managers’ Golden Rules

We are indexing all kinds of academically relevant content. In case your source does not provide a

However, the best way to get your documents indexed is to make sure your source data is up-to-date.

You can check some of the following items using our:

- OAI interface working
  - Your OAI interface is stable and responds to requests.
  - Otherwise, it is not possible to index your oai.

- Comprehensive metadata
  - Each item exposed via your OAI interface provides metadata.
  - If important metadata is missing, documents are not indexed.

- Identifier (URLs) working
  - Each item provides a URL in the .id/.identifier that should lead to the front door.
  - Only documents whose identifiers start with "http" are indexed.

- Providing access information (Open Access)
  - Access information of the fullest extent should be provided by metadata record in the field "setSpec".
  - If correct access information is missing, these documents are not indexed.

- Providing information concerning reuse / li
  - Authors can publish their work under a Creative Commons license.
  - All content exposed via your OAI interface should be under a Creative Commons license.

- Character encoding
  - All content exposed via your OAI interface should be in UTF-8 encoding.
Further Joint Activities

• Data analysis and visualization of landscape of global repositories

• Infograph of IR Guidelines
  • CIRG wants to prepare Chinese summaries of the existing IR Guidelines

• Subject and Research Data Repository
  • CIRG will think about how to develop on this aspects
Outline

• The Partners
• Collaborative Activities
• The Results
• Conclusion
Including Chinese Repositories in BASE

New Repositories per year

Number (repositories)

- 2009: 1
- 2010: 1
- 2013: 1
- 2014: 10
- 2016: 2
- 2017: 18
Number of indexed Documente und Sources in BASE for China, People's Republic of
<table>
<thead>
<tr>
<th>Country</th>
<th>Content Sources</th>
<th>Documents</th>
<th>Feed</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States of America</td>
<td>733</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>306</td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>International Organisations</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russia</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>India</td>
<td>109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ukraine</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>25</td>
<td>326,411</td>
<td>RSS</td>
</tr>
<tr>
<td>Portugal</td>
<td>56</td>
<td>304,326</td>
<td>RSS</td>
</tr>
</tbody>
</table>

Nr 21
Outline

• The Partners
• Collaborative Activities
• The Results
• Conclusion
Conclusion

• Communication helps – the more the better
• Benefits for both partners
• We will go further
Benefits for Bielefeld UL

• Increasing Number of sources and documents
• Improving of Metadata Quality
• Improving of BASE Interface Quality and Global Usage
• Supporting Strategic Position in the Repository Network

Benefits for China Repository Infrastructure

• Improving Global Visibility of Chinese Scientific Output
• Improving Global Usage quality
• Improving Visibility of Global Research Output in China
• Improving Technical Infrastructure
References

10 years of „Bielefeld Academic Search Engine“ (BASE): Looking at the past and future of the world wide repository landscape from a service providers perspective
Pieper D, Summann F (2015)
Presented at the OR2015. 10th International Conference on Open Repositories, Indianapolis.
[https://pub.uni-bielefeld.de/publication/2766308]

Establishing Open Access Services in the Global Repository Network, Experiences and Challenges from the Service Provider Perspective (BASE)
Summann F (2016)
Presented at the 2016 Chinese Institutional Repository Conference, Chongqing.
[http://ir.las.ac.cn/handle/12502/8749]

Repository Network in China,
Li-Ping K （2017）
Presented at the 2017 COAR GA, Venice.
[http://ir.las.ac.cn/handle/12502/9099]
Thank you!

friedrich.summann@uni-bielefeld.de
gulp@mail.las.ac.cn
dirk.pieper@uni-bielefeld.de