Deutsches DSpace-User-Group-Meeting 2016

DSpace im Repository-Netwerk

Schnittstellen, Zertifikate, Guidelines, Metadatenqualität

Friedrich Summann

Universitätsbibliothek Bielefeld

BASE = Bielefeld Academic Search Engine (www.base-search.net)

- Wissenschaftliche Suchmaschine
- Wiss. Inhalte bereitgestellt von Repositorien via OAI-PMH (OAI Service Provider)
- Konzeption und Entwicklung ab 2001

Heute:

- Mehr als 4680 Quellen
- Mehr als 99 Mill. Dokumente/Objekte

BASE heute: OAI Service Provider und Spiegelbild der Repository Landschaft

- 4688 Repositorien (via OAI-PMH)
- Aus 111 Ländern weltweit
- Ca. 99 Mill. Dokumente/Objekte
- Ca. 70 % Open Access
- Dublin Core Format
- Ca. 14.7 Mill. Dokumente mit angereichertem DDC-Code (Dewey)

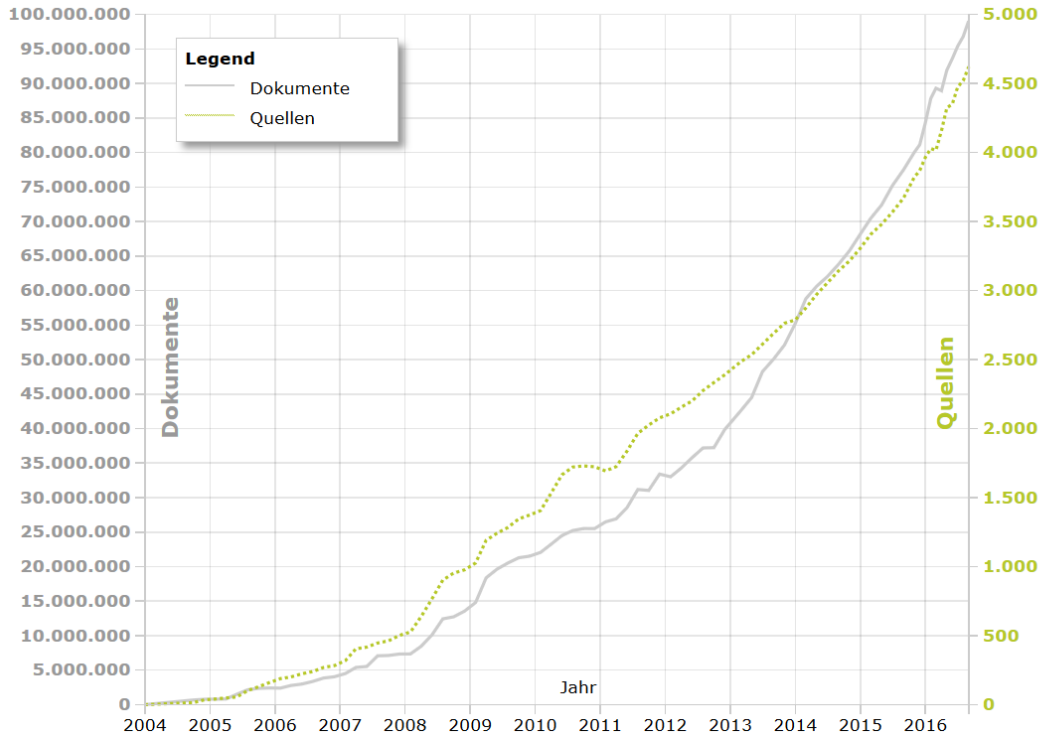
Some BASE Milestones

- 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, HTTP Interface
- 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

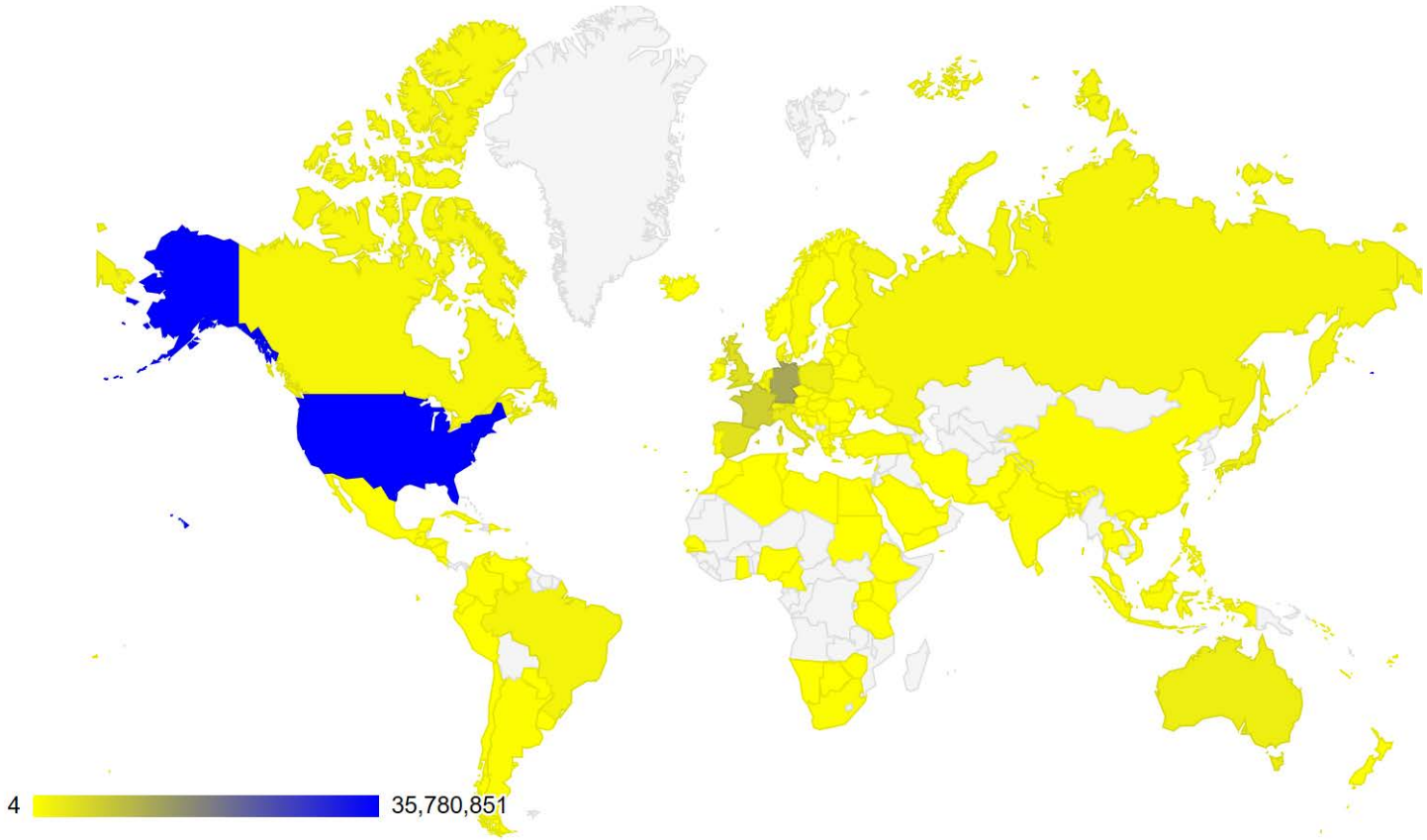
Uber BASE: Statistik

Entwicklung der Zahl der von BASE indexierten Quellen und Dokumente seit September 2004.

Diagramm: Zahl der indexierten Dokumente und Quellen in BASE



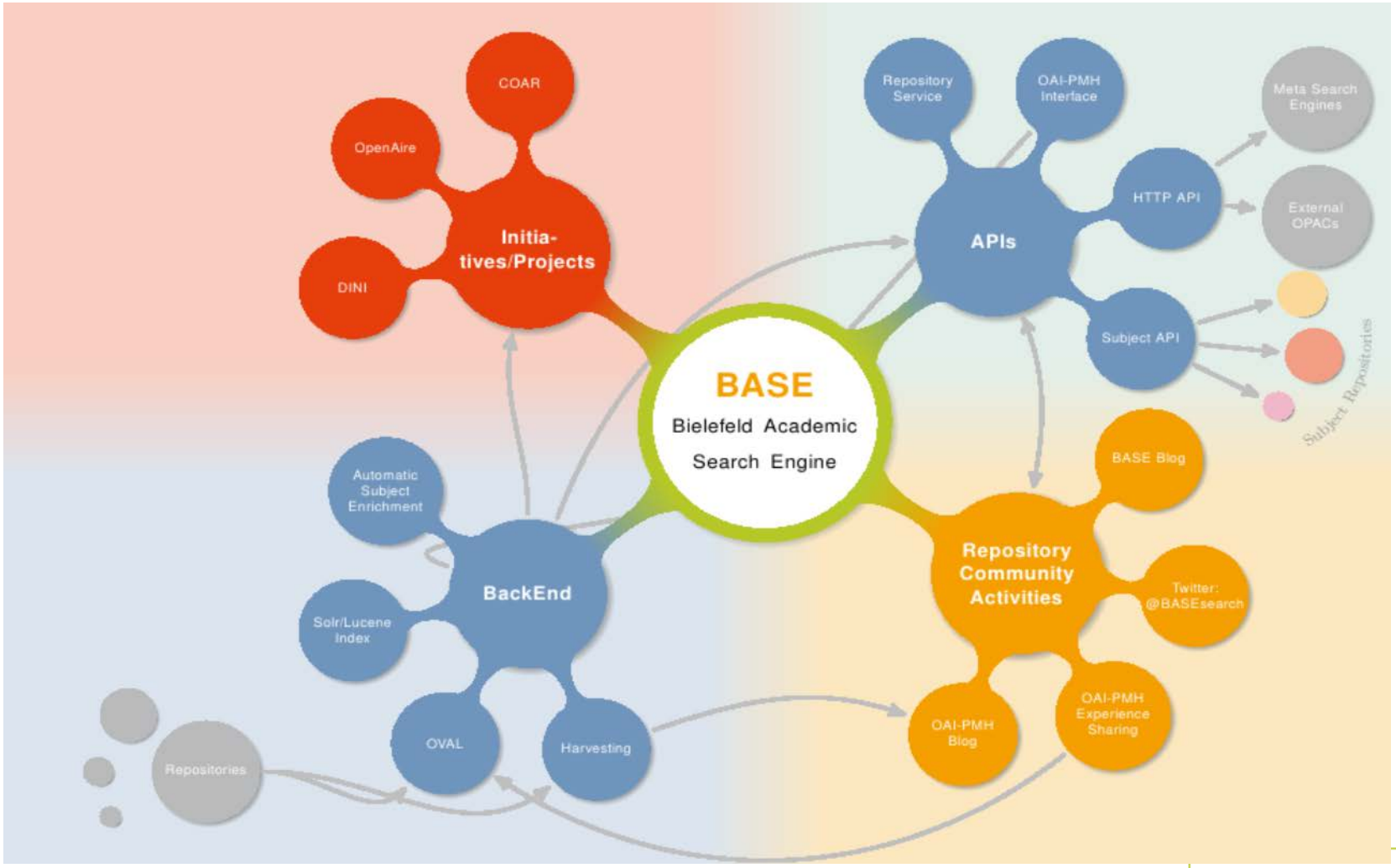
Anzahl der Dokumente in BASE nach Ländern



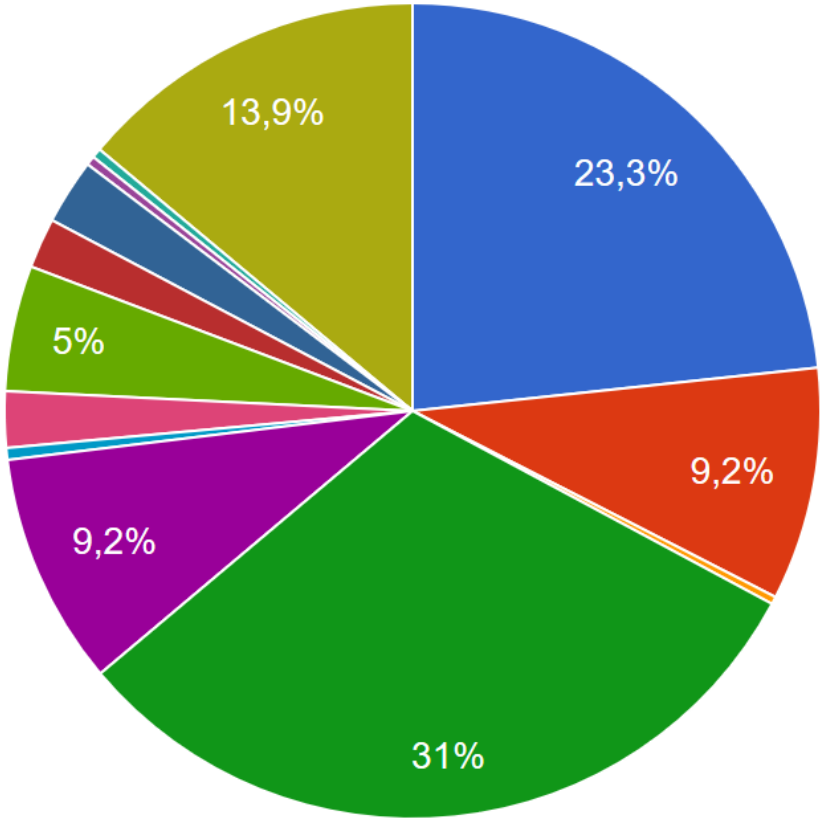
Der BASE Focus

- OA Repositorien (global)
- Wissenschaftliche Inhalte
- Schwerpunkt Institutionelle Repositorien
- Aggregatoren (RePEc etc)
- Subject Repositories (arXiv, CiteSeerX etc)
- Electronic Journals (DOAJ, OJS-Installationen)
- Digitale Sammlungen
- Forschungsdaten-Repositorien

BASE im Wissenschaftlichen Netzwerk



Verteilung DSpace weltweit - Anzahl der Installationen



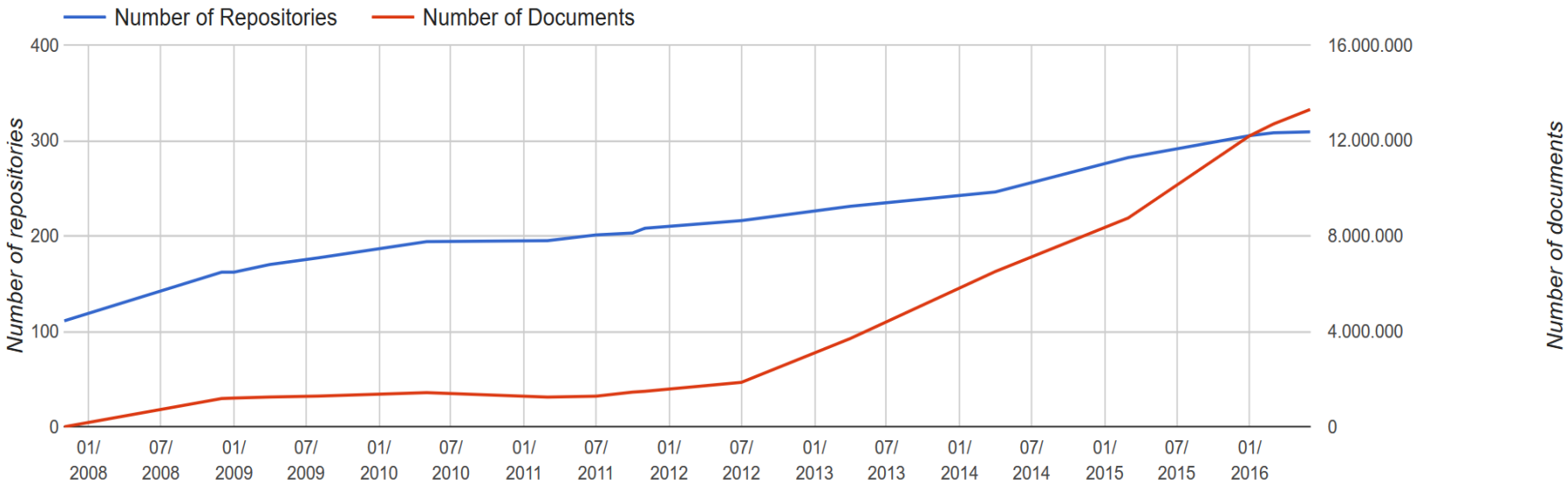
Eprints 432
Dspace 1090
DigitalCommons 431
ContentDM 121
Fedora 22
Opus 104
Digitool 11
WEKO 233
HAL 93
OJS 1451
Pure 18
Invenio 14
Greenstone 17
Andere 651

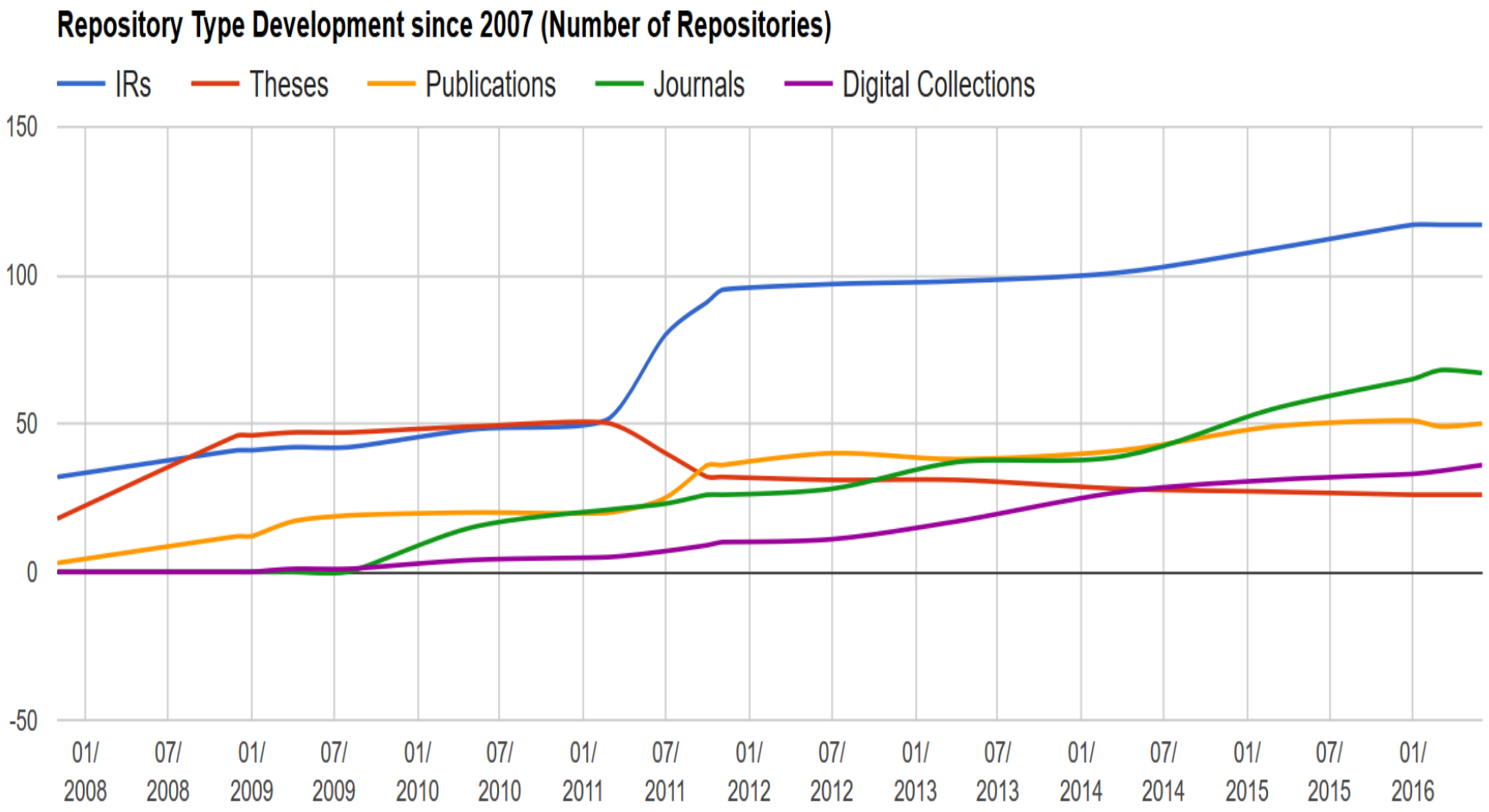
Germany

11853051 Documents from **309** Repositories

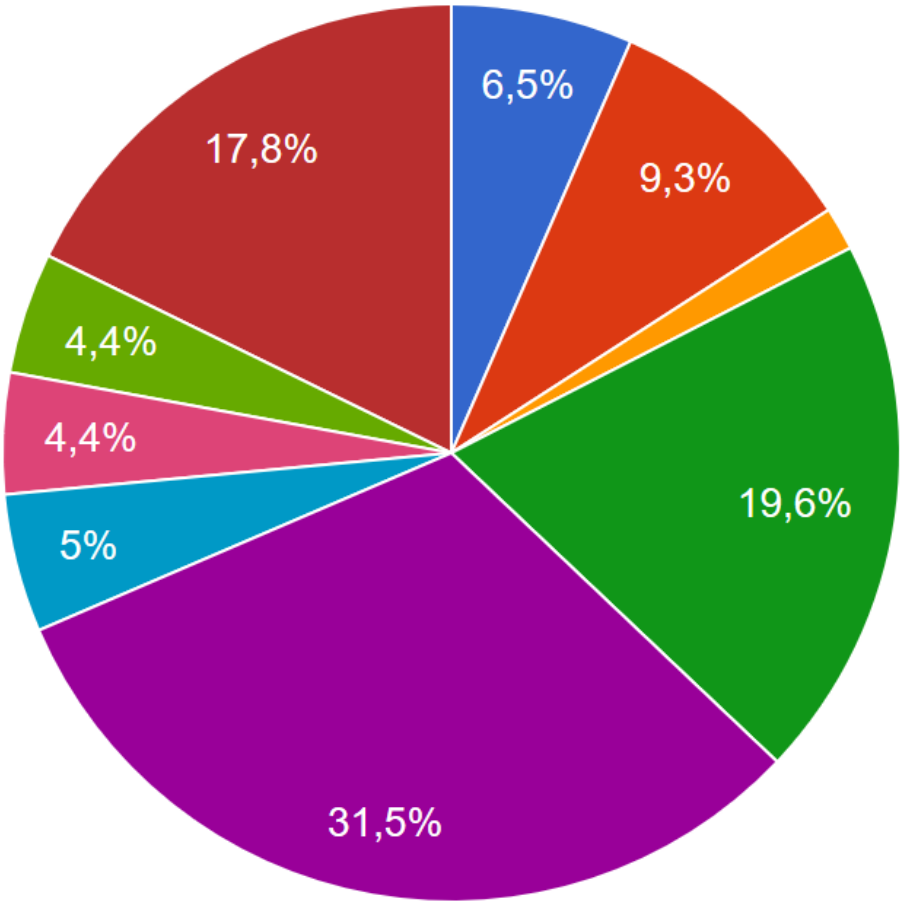
This means: **12.59 %** of Documents world-wide, **7.10 %** of Repositories world-wide

Repository Development since 2007




Repository Type Development since 2007 (Number of Repositories)
IRs
Theses
Publications
Journals
Digital Collections
150
100
50
0
-50
01/ 2008
07/ 2008
01/ 2009
07/ 2009
01/ 2010
07/ 2010
01/ 2011
07/ 2011
01/ 2012
07/ 2012
01/ 2013
07/ 2013
01/ 2014
07/ 2014
01/ 2015
07/ 2015
01/ 2016

Verteilung DSpace in DE - Anzahl der Installationen



DSpace 21
EPrints 30
Invenio 5
OJS 63
Opus 101
Goobi 16
VisualLibrary 14
MyCoRe 14
Andere 57

Universität Bielefeld

DSP

About D

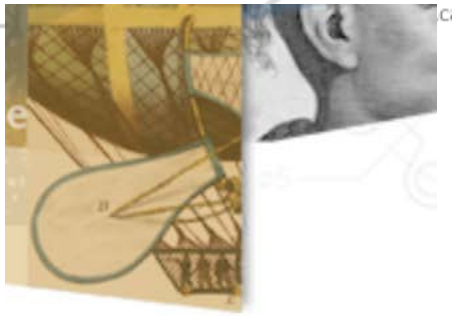
DSpa

Check out
repository

Search th

BASE

cademic Search Engine



Filter By Country:

- [ ] united states (294)
- [ ] india (157)
- [ ] spain (105)
- [ ] japan (101)
- [ ] brazil (91)
- [ ] turkey (86)
- [ ] ukraine (67)
- [ ] united kingdom (65)
- [ ] taiwan (63)
- [ ] italy (62)
- [ ] norway (54)
- [ ] portugal (50)
- [ ] colombia (49)
- [ ] canada (47)
- [ ] germany (32)
- [ ] peru (29)
- [ ] australia (28)
- [ ] mexico (28)

Filter By Dspace Version:

- [ ] Unknown (334)
- [ ] 5.x (145)
- [ ] 4.x (328)
- [ ] 3.x (251)
- [ ] 1.8.x (181)
- [ ] 1.7.x (214)
- [ ] 1.6.x (153)
- [ ] 1.5.x (117)
- [ ] 1.4.x (89)
- [ ] 1.3.x and earlier (40)

The **10** Biggest Misunderstandings around OAI-PMH

- OAI-PMH means ‚Everything is Open Access‘
- Persistent Identifiers are persistent
- Link to the Document page is not necessary
- Configuration is not needed
- Checking the Service is needless
- DublinCore is simple but sufficient
- OpenAccess Status Information is needless
- End-User Interface is not necessary
- Personal Sys Admin Email Address is not needed
- Incremental Harvesting is sufficient

Main Issues to Avoid

- Wrong Document URLs

(a specific Dspace problem with non-configured links as <dc:identifier>http://hdl.handle.net/123456789/87639</dc:identifier>)

- Empty Records
- Invalid XML delivered
- Crashing Harvest Processes
- Changing OAI-PMH basicurls without dissemination/redirect

Main Issues Preferred

- OA Status delivered (on repository or document level)
- Metadata Guidelines compatible
  - Vocabularies used (for type, language, date, classification etc.)
- English end-user interface (in parallel)
- Citation/Abstract information delivered
- Available repository contact information
- Visible in Registries (OpenDOAR, openarchives …)

DSpace aus BASE-Sicht

Positives

- Stabil, robust, valide, weit verbreitet
- Hohe Metadatenqualität
- CRIS-Erweiterung

Verbesserungsoptionen

- Kontakt zur Entwickler-Community
- OAI-PMH (korrekt) konfigurieren
- ORCID-Unterstützung
- Vokabular OA-Status
- Vokabular für Lizenzen
- Internationale Sichtbarkeit
- Volltextlink

DSpace OAI-PMH issues (1)

*** In many cases the persistent identifiers are not configured at all.**

<dc:identifier>http://hdl.handle.net/123456789/3</dc:identifier>

*** a certain local identifier is configured but it is not registered at the handle system**

*** Identifier links are mixed (some wrong, some ok)**

DSpace OAI-PMH issues (2)

* **local url format is not correct** (containing some additional strings as xmlui for example)
* **Admin email address often unconfigured**.
* **No response from the feedback form**
* **Harvesting crashes**
* **A really often occurring problem is that the basicurl has lost the port number (mostly 8080) in the OAI basicurl**
* **Instead of UTF-8 characters the OAI-PMH interface delivers question marks**

Lyncode XOAI problems

*** *No records* message while the repository has contents**

*** Links in the Lyncode navigational bar are different with the starting basicurl and do not work!**

*** Incremental harvesting fails when using the day granularity in from parameter**

Aktuelle BASE-Umsetzungen

Verbesserung Normalisierung OA-Status

Verbesserung Normalisierung Lizenzbedingungen

Verbesserung Normalisierung Publikationstypen

Issue: OA Status, Rights and Licences Normalization

SUBGoettinger	DE	ceu	181	Scan	**5x** http://creativecommons.org/licenses/by-sa/3.0 **5x** http://creativecommons.org/licenses/by-nc/2.5/ **5x** http://creativecommons.org/licenses/by-nc/2.0/ **4x** http://goedoc.uni-goettingen.de **4x** info:eu-repo/semantics/http://creativecommons.org/licenses/by-nd/3.0/de/ **4x** https://creativecommons.org/licenses/by/4.0 **4x** http://creativecommons.org/licenses/by/2.0/uk/legalcode **4x** http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de

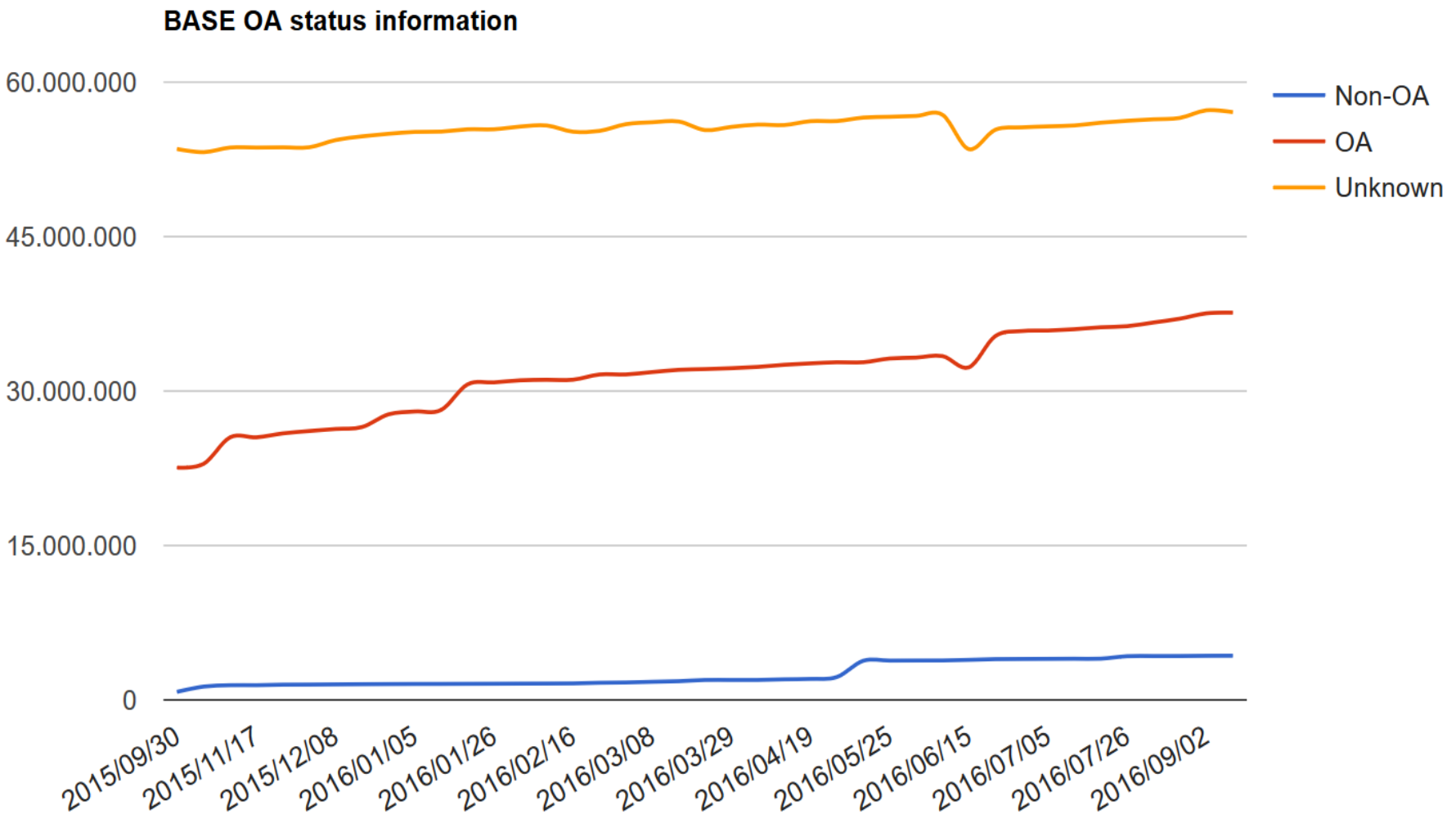
Nachnutzung/Lizenzen

- ☑ **Alle**
  - ☑ **Creative Commons**
    - ☑ CC-BY
    - ☑ CC-BY-SA
    - ☑ CC-BY-ND
    - ☑ CC-BY-NC
    - ☑ CC-BY-NC-SA
    - ☑ CC-BY-NC-ND
  - ☑ **Public Domain**
    - ☑ CC0
    - ☑ Public Domain Mark (PDM)

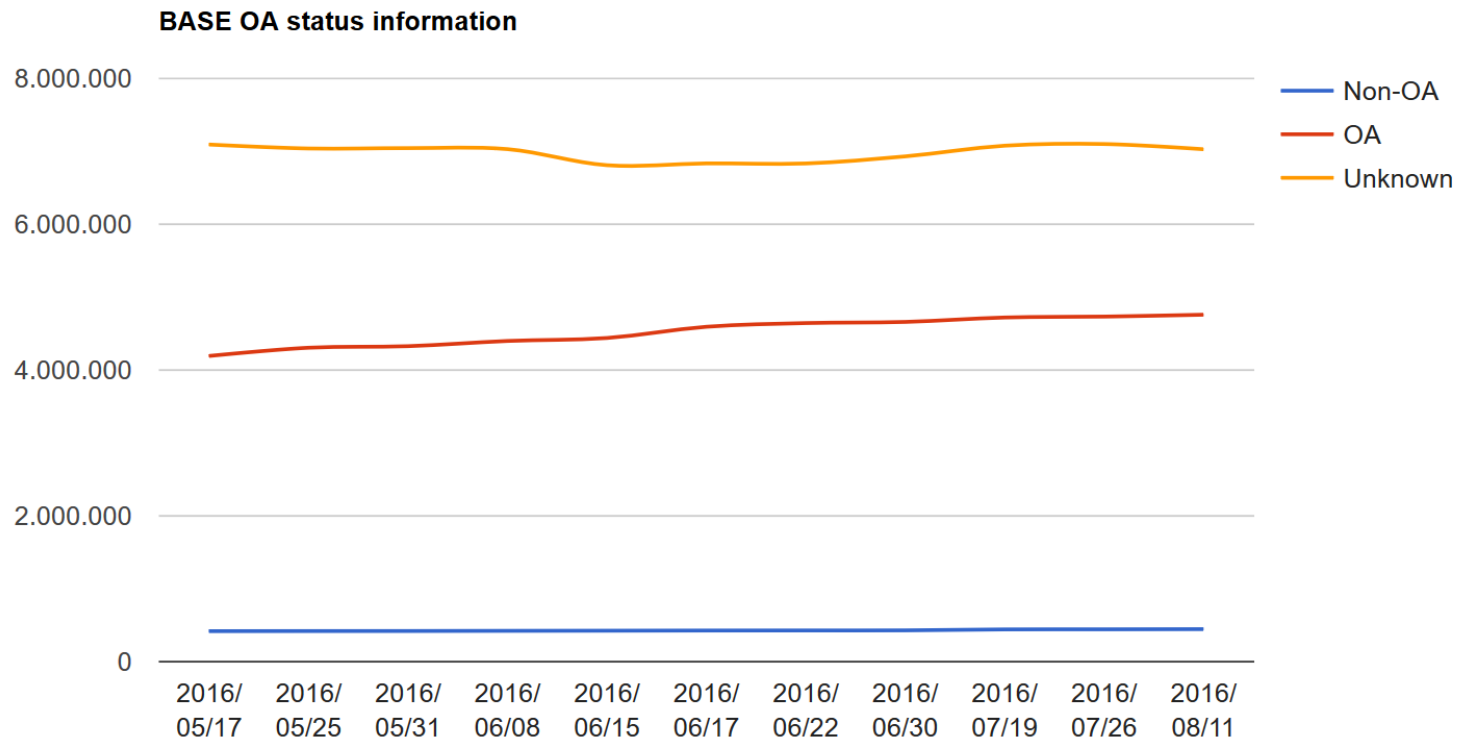
Zugang

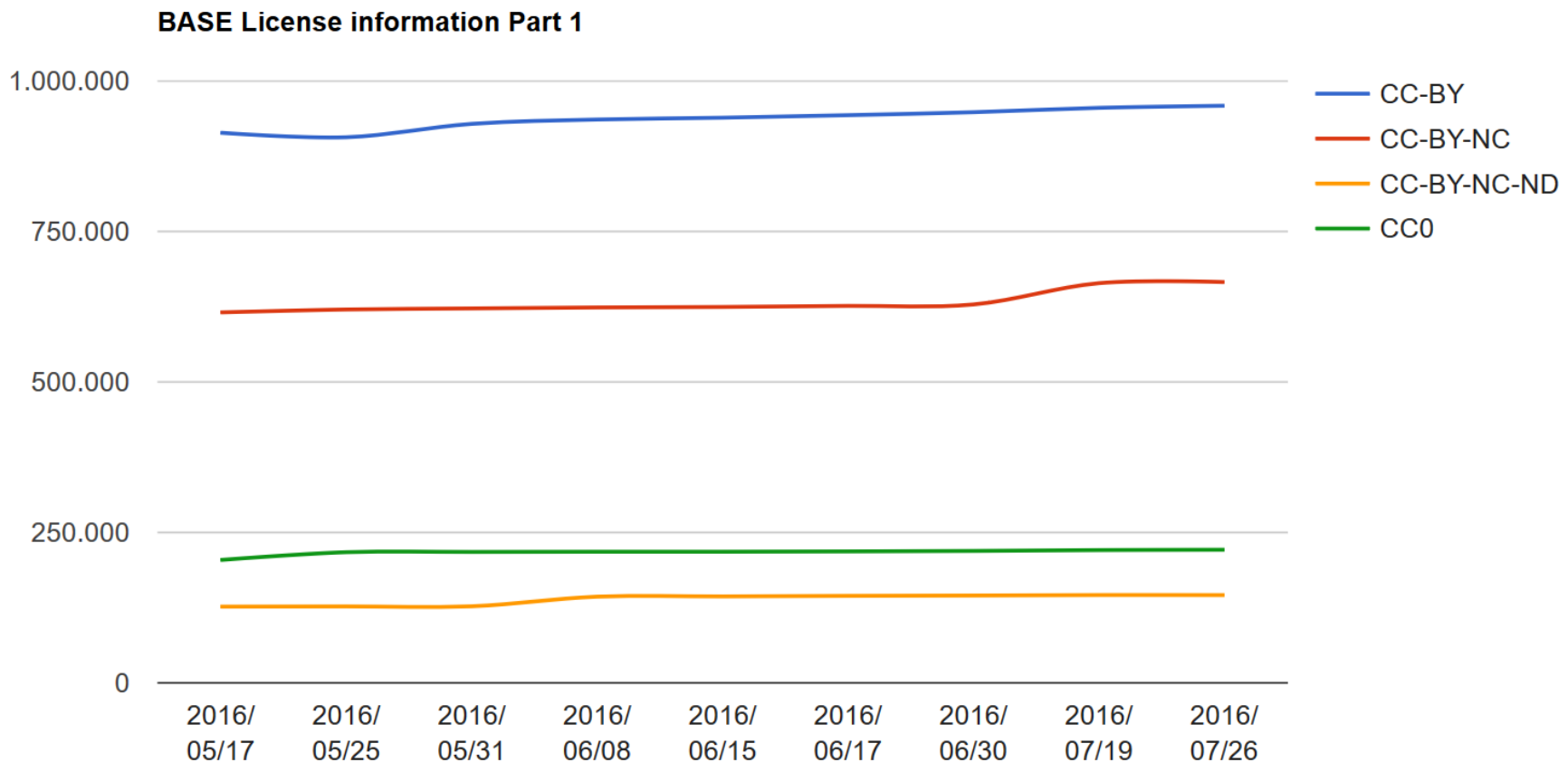
- ☑ Open Access
- ☑ Kein Open Access
- ☑ Unbekannt

					2x http://creativecommons.org/licenses/by-nc-nd/3.0/at/ **2x** http:// http://creativecommons.org/licenses/by/2.0 **2x** 594430


BASE OA status information
60.000.000
45.000.000
30.000.000
15.000.000
0
Non-OA
OA
Unknown
2015/09/30
2015/11/17
2015/12/08
2016/01/05
2016/01/26
2016/02/16
2016/03/08
2016/03/29
2016/04/19
2016/05/25
2016/06/15
2016/07/05
2016/07/26
2016/09/02

OA status and licence Information Germany




BASE License information Part 1
1.000.000
750.000
500.000
250.000
0
2016/ 05/17
2016/ 05/25
2016/ 05/31
2016/ 06/08
2016/ 06/15
2016/ 06/17
2016/ 06/30
2016/ 07/19
2016/ 07/26
CC-BY
CC-BY-NC
CC-BY-NC-ND
CC0

Dokumentart

- ☑ ***Alle Dokumentarten***
  - ☑ **Text**
    - ☑ Buch
      - ☑ Teil eines Buches
    - ☑ Zeitschrift/Zeitung
      - ☑ Artikel in einer Zeitschrift/Zeitung
      - ☑ Anderer Beitrag in einer Zeitschrift/Zeitung
    - ☑ Konferenzveröffentlichung
    - ☑ Bericht
    - ☑ Review
    - ☑ Lehrmaterial
    - ☑ Vortrag
    - ☑ Manuskript
    - ☑ Patent
    - ☑ Abschlussarbeit
      - ☑ Bachelorarbeit
      - ☑ Masterarbeit
      - ☑ Dissertation und postgraduale Arbeiten
  - ☑ **Noten (Musik)**
  - ☑ **Karte**
  - ☑ **Audio**
  - ☑ **Bild/Video**
    - ☑ Einzelbild
    - ☑ Animation/Video
  - ☑ **Software**
  - ☑ **Forschungsdaten**
  - ☑ **Unbekannt**

Aktuelle Entwicklungen

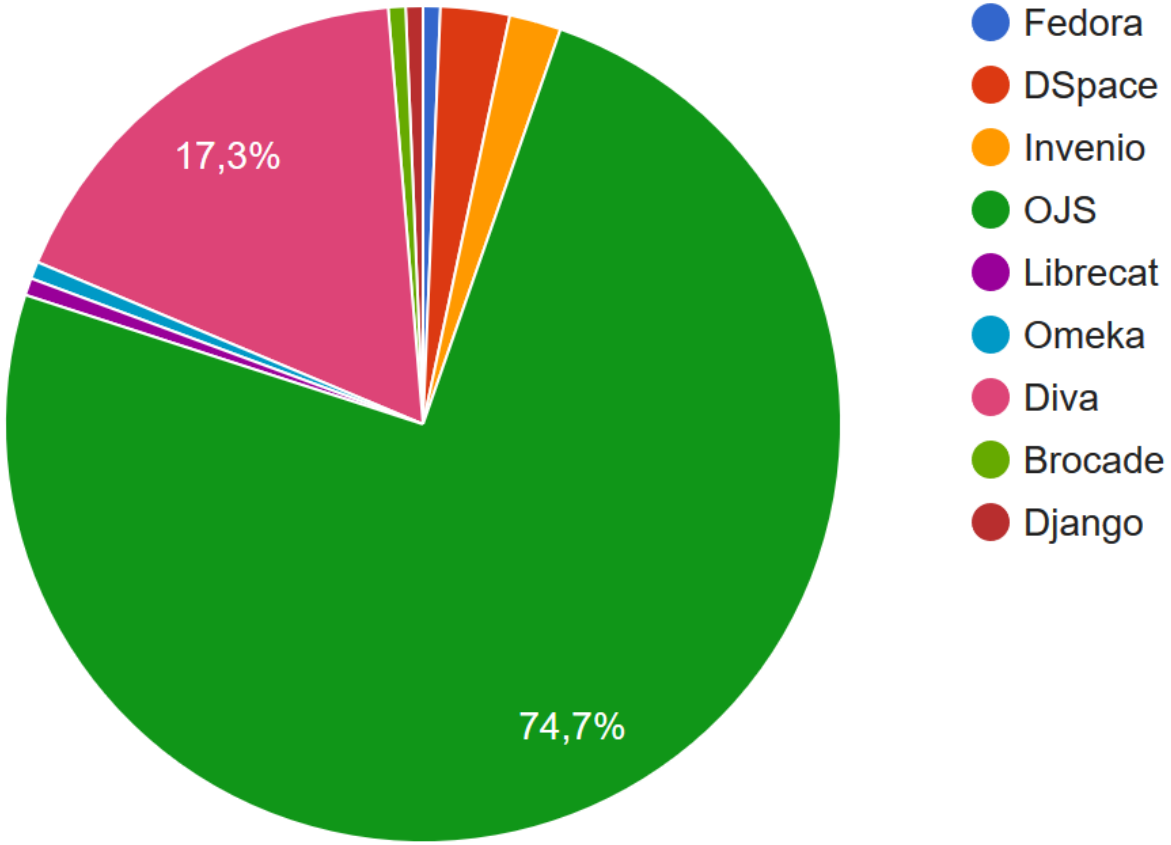
Implementierung eines ORCID-Claiming-Services im Rahmen von ORCID.DE

Auswertung und Analyse von alternativen Metadaten-Formaten jenseits von oai_dc

Integration von (Open-Access-)Inhalten aus CrossRef

Verbesserung der Registry-Informationen (Auswertungen zu Repositories und Ländern)

Autorenidentifikatoren - Verteilung Repository-Systeme mit ORCID-IDs

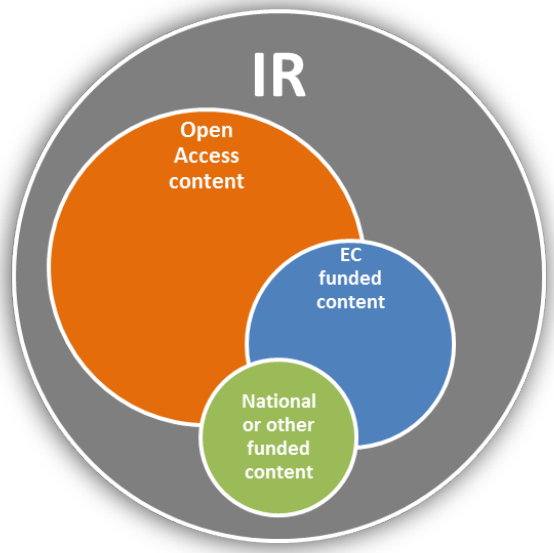


Es geht um Metadatenqualität!

- DRIVER Guidelines
- DINI Zertifikat
- OpenAire Guidelines
- COAR Vocabularies

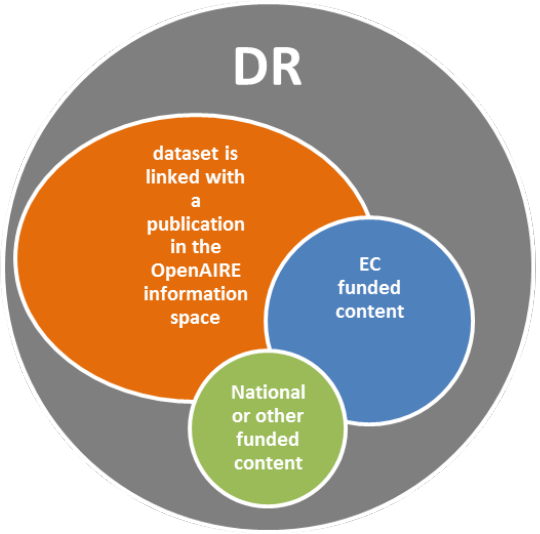
OpenAIRE Content Acquisition

Publications



- Repository registriert in OpenDOAR
- OpenAIRE Literature Guidelines v3

Research Data



- Repository registriert in re3data
- OpenAIRE Data Archive Guidelines v2

Schnittstellen in OpenAIRE

- Repository-Validierung und Registrierung über
  - https://www.openaire.eu/validator/welcome
- Aggregation und Indexierung i.d.R. wöchentlich
- Auffindbarkeit der aggregierten Quelle über
  - https://www.openaire.eu/search/data-providers
- OpenAIRE exponiert (ggf. angereicherte) Metadaten über
  - http://api.openaire.eu/
  - Unter Verwendung von OAI-PMH bzw. REST-API
  - EU-Projektinformation bereitgestellt für u.a. DSPace
    - http://api.openaire.eu/#cha_projects_http

Unterstützung für DSpace Publikationen

- Allg. OpenAIRE Literature Guidelines v3
  - https://guidelines.openaire.eu/en/latest/literature/index.html
  - Metadata-Format “oai_dc”
  - Empfohlenes OAI-Set “openaire”
- Implementiert / unterstützt in DSpace 5.x
  - Wichtig: separater XOAI-Context für OpenAIRE:
    - http://www.example.com/oai/openaire
    - Exponiert OA + funded Records
- Allg. OpenAIRE Data Archive Manager Guidelines v2
  - Metadata-Format “oai_datacite”
  - Empfohlenes OAI-Set “openaire_data”

Unterstützung für DSpace Forschungsdaten

- Allg. OpenAIRE Data Archive Manager Guidelines v2
  - https://guidelines.openaire.eu/en/latest/data/index.html
  - Metadata-Format “oai_datacite”
  - Empfohlenes OAI-Set “openaire_data”
- Implementierungsbeispiel Lindat-CLARIN DSpace Repository
  - https://github.com/ufal/lindat-dspace/wiki/OpenAIRE

Deutsche Dspace Repositories in OpenAIRE

- OPARU (Universität Ulm)
- GoeScholar (Georg-August-Universität Göttingen)
- TOBIAS-lib (Universität Tübingen)
- Social Science Open Access Repository (SSOAR)
- RepOSitorium (Universität Osnabrück)
- KOPS (Konstanzer Online-Publikations-System)
- Open Access Repository der TUHH (TUBdok)
- res doctae (Akademie der Wissenschaften zu Göttingen)
- DSpace an der Universität Kassel (KOBRA)
- EconStor (ZBW)
- Eldorado (Dortmund)
- Helmholtz Zentrum für Infektionsforschung Repository
- DepositOnce (TU-Berlin)

Vielen Dank für Ihre Aufmerksamkeit!

friedrich.summann@uni-bielefeld.de

base-search.net
follow us: **@BASEsearch**