

Aktuelle Entwicklungen im globalen Repository- und Open-Access-Netzwerk

Relevante Themenbereiche aus der technischen Sicht eines Service Providers

Friedrich Summann

ORCID ID

 orcid.org/0000-0002-6297-3348

Universitätsbibliothek Bielefeld

Zusammenfassung

- Protokollfragen
- Metadatenqualität
- Weitere Repository-Formen
- Repository-Tools

Der Vortrag stellt aus der Sicht eines Repository Service Providers aktuelle Themen vor, die in der Diskussion sind.



Zusammenfassung

- **Protokollfragen**
 - OAI-PMH
 - ResourceSync
 - Weitere APIs (Crossref, PLOS etc.)
- Metadatenqualität
- Weitere Repository-Formen
- Repository-Tools

OAI-PMH

- Einfach, bewährt, weit verbreitet und stabil
- Probleme
 - Technische Umsetzung
 - Falsche Konfiguration
 - Schwierigkeiten bei grossen Datenmengen
 - Nicht REST-basiert
 - Zuwenig Transparenz in Bezug auf Verfügbarkeit

Ein konkretes Beispiel für ein Harvesting-Problem:

```
<OAI-  
PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.  
w3.org/2001/XMLSchema-  
instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/  
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">  
<responseDate>2017-09-10T12:54:30Z</responseDate>  
<request verb="ListRecords" metadataPrefix="oai_dc">xxx</request>  
<error code="entityTooLarge">
```

**entityTooLarge: Too many records, allowed are
only 10000**

```
</error>  
</OAI-PMH>
```

**In der Konsequenz läßt sich das Repository nicht mehr harvesten,
da es mehr als 10000 Sätze/Objekte enthält.**

ResourceSync als Folgeprotokoll?



Open Archives Initiative
ResourceSync Framework Specification



ResourceSync Framework Specification - Table of Contents

22 February 2017

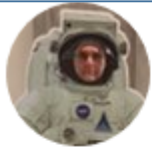
Start Here with the ResourceSync core specification.

- [ResourceSync Framework Specification](#) (ANSI/NISO Z39.99-2017)

Additional Specifications provide extensions to the ResourceSync core. These include a specification to support archives of synchronization information-based change communication.

- [ResourceSync Change Notification](#)
- [ResourceSync Framework Notification](#)
- [ResourceSync Archives](#) (beta draft)
- [Relation Types used in the ResourceSync Framework](#)

Die OAI-PMH-Gründungsväter sind sich einig, dass ein eventueller Übergang nicht zu einer Lücke führen darf!



Herbert @hvdsomp · 29. Juni

Great presentation on why repositories should implement **#ResourceSync** but not yet give up on **#oaipmh**



Herbert @hvdsomp · 29. Juni

Great presentation on why repositories should implement **#ResourceSync** but not yet give up on **#oaipmh**



Amanda Curnow @Mand1can3 · 29. Juni

goodbye OAI-PMH, hello **ResourceSync**

zimeon @zimeon

ResourceSync should replace OAI-PMH; but don't break harvesting by creating a gap. 24x7 slides: slideshare.net/simeonwarner/m... **#OR2017**



Crossref API

- Technisch ähnlich wie OAI-PMH
- DOI, Lizenzinformation, ORCID enthalten
- Aktueller Stand
 - 8,793 Member
 - 82,703,000 Publications
 - In BASE harvested: 896 (Stand: Sept. 2017)
 - In BASE indexed: 16 (18)

Crossref in BASE

Liste der Quellen

[Neuer Eintrag](#) |
 [20 Einträge pro Seite](#) |
 [Alle Einträge](#) |
 [Listenfelder](#) |
 [Übernahme ins Live-System](#)

Nr.	Int. ID	Datum	Harvestername	System	Kont.	Land	Bundesst.	Breitengr.	Längengr.	OAI Basic UI
				crossref						
1.	6640	30.08.2017	OpenBookPubl	CrossRef	ceu	uk				
2.	6635	29.08.2017	ISTEOpenScience	CrossRef	ceu	fr				
3.	6634	29.08.2017	OpenventioPubl	CrossRef	cas	in				
4.	6633	29.08.2017	DeGruyterOpen	CrossRef	ceu	pl				
5.	6615	18.08.2017	TranscriptVerlag	CrossRef	ceu	de	denw			
6.	6548	11.07.2017	Winnower	CrossRef	cna	us				
7.	6547	11.07.2017	F1000	CrossRef	ceu	uk				
8.	6343	05.05.2017	OpenAccessPub	CrossRef	cna	us				
9.	6342	05.05.2017	OAPublLondon	CrossRef	ceu	uk				
10.	6331	03.05.2017	PLOS	CrossRef	cna	us				
11.	6133	08.03.2017	OpenLibHumanities	CrossRef	cww	org				
12.	6116	06.03.2017	OpenAccessText	CrossRef	ceu	uk				
13.	5953	08.02.2017	FrontiersMedia	CrossRef	ceu	ch	chvd			
14.	5903	25.01.2017	Wiley	CrossRef	cna	us				
15.	5870	16.01.2017	ElsevierBV	CrossRef	cna	us				
16.	5564	14.10.2016	JLMCS	CrossRef	ceu	de	dehe			

Crossref Member List

☰ Crossref Lister

Crossref Wachter

Config Entries

Number of Repository Entries: 8281
 Number of Records: 13064
 Number of activated Members: 896
 Cover : 13064
 Last Update: 08.09.

Name	
Elsevier BV	CROSSREF-ElsevierBV Activate U Edit Profil
Springer Nature	CROSSREF-SpringerN Activate U Edit Profil
Wiley-Blackwell	CROSSREF-Wiley Activate U Edit Profil
Informa UK Limited	CROSSREF-InformaU Activate U Edit Profil
Institute of Electrical and Electronics Engineers (IEEE)	CROSSREF-IEEE-CR Activate U Edit Profil
Oxford University Press (OUP)	CROSSREF-OxfordUni

Crossref Wachter

Config Entries

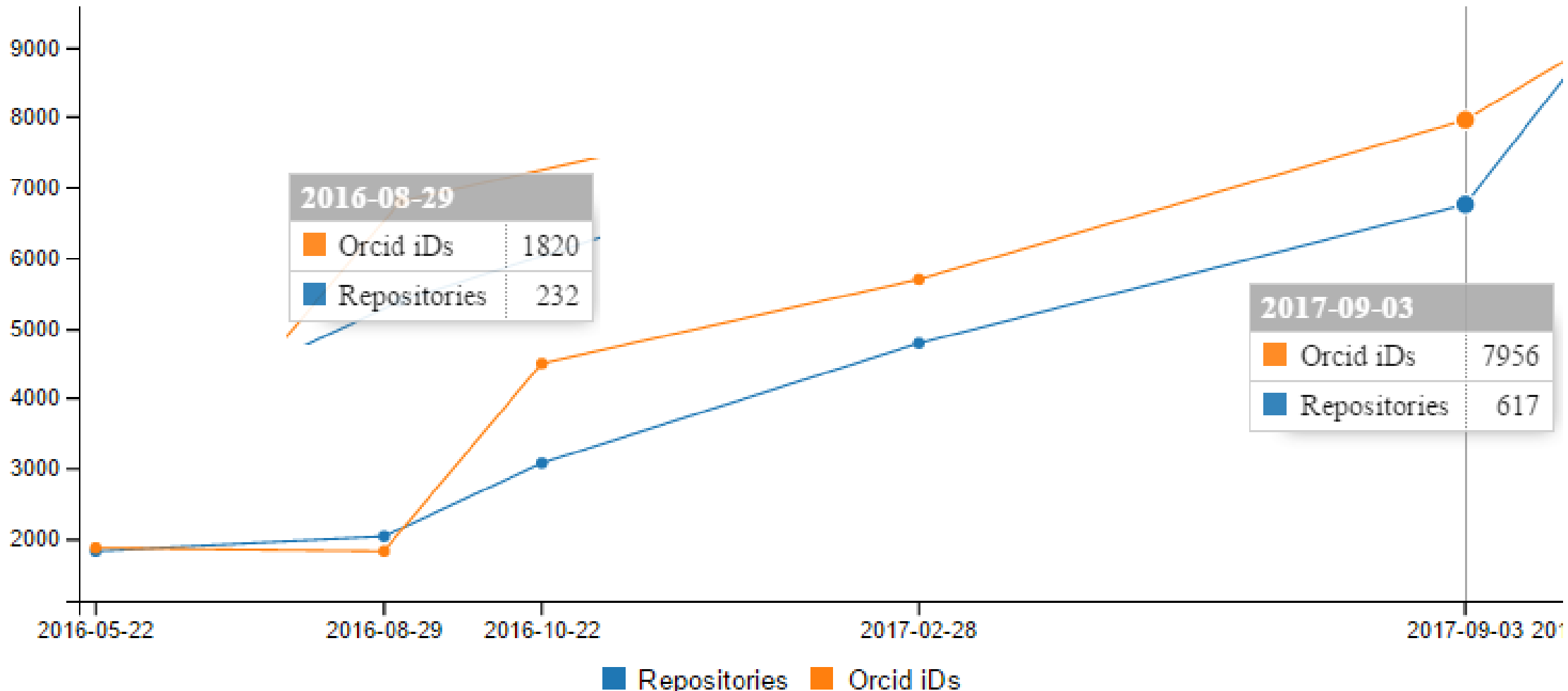
Number of Repository Entries: 8793
 Official number of Records: 82703036
 Number of Harvested Records: 130646343
 ---- Unique Records: 50672688
 Number of activated Members: 896
 Members with License Information: 1010
 Members with ORCIDs: 1250
 Last Update: 10.09.2017 - 18:22

Zusammenfassung

- Protokollfragen
- **Metadatenqualität**
 - Identifier (DOI, ORCID, Fundref ...)
 - Vokabulare
- Weitere Repository-Formen
- Repository-Tools

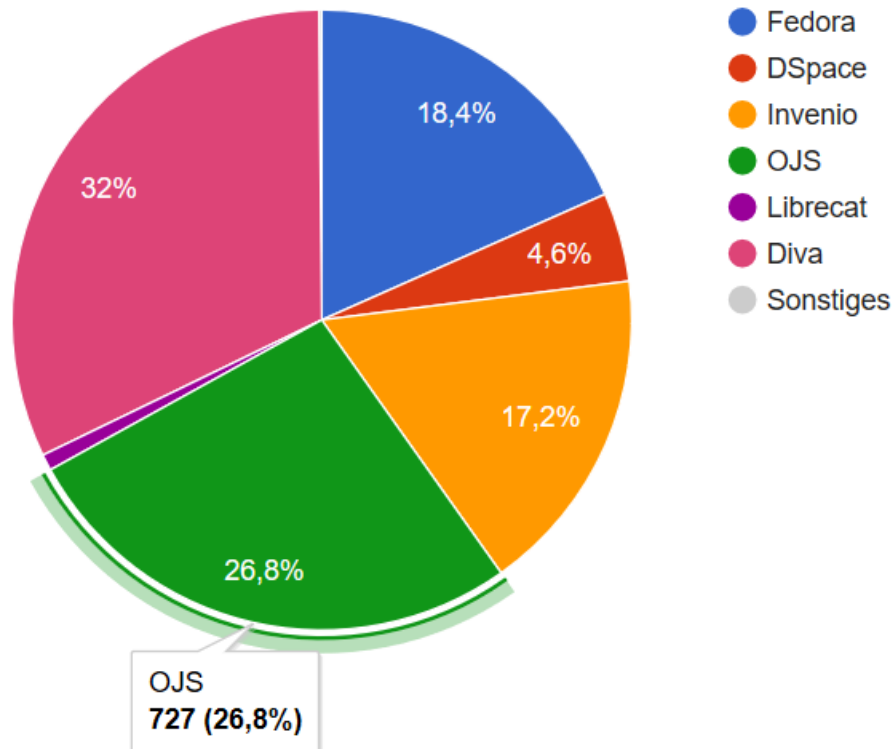
Globale Entwicklung ORCID iDs in Repositorien

(Zahlen aus dem ORCID-DE-Projekt)

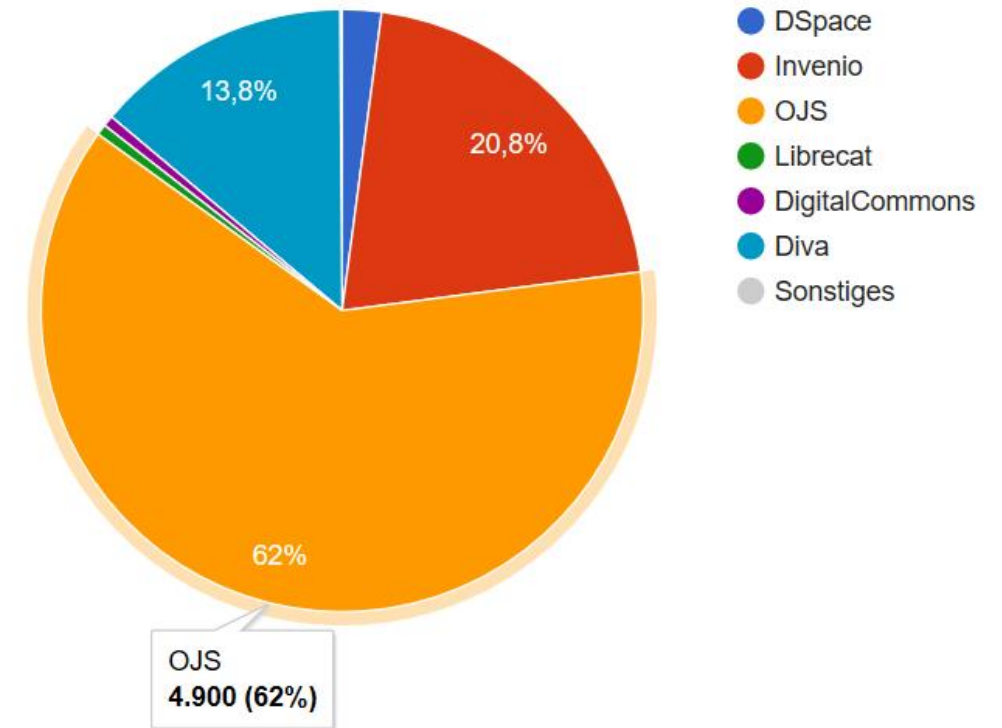


Globale Verteilung ORCIDs in Repository-Systemen

Autorenidentifikatoren - Verteilung Anzahl ORCID-IDs in Repository-Systemen (Juni 2016)



Autorenidentifikatoren - Verteilung Anzahl ORCID-IDs in Repository-Systemen Sept. 2017



Erhöhung der Metadatenqualität: Vokabularnutzung

COAR
Confederation of Open Access Repositories

Towards a global knowledge commons

Log In Search

Activities Community News & Publications About

Home » Activities » Repository Interoperability » COAR Vocabularies » Resource Type Vocabulary

Advocacy & Leadership +

Repository Interoperability -

COAR Vocabularies -

Resource Type Vocabulary

Access Rights Vocabulary

FAQs for Controlled Vocabularies

Related Resources

About the Editorial Board

Resource Type Vocabulary

Introduction

The Resource Type vocabulary defines concepts to identify the genre of a resource. Such resources, like publications, research data, audio and video objects, are typically deposited in institutional and thematic repositories or published in e-journals. This vocabulary supports a hierarchical model that relates narrower and broader concepts. Multilingual labels regard regional distinctions in language and term. Concepts of this vocabulary are mapped with terms and concepts of similar vocabularies and dictionaries.

The Resource Type vocabulary builds on and extends the controlled list of publication types defined in info:eu-repo/semantics. Backward compatibility is ensured by a suggested mapping.

The Resource Type vocabulary is part of the [COAR Controlled Vocabularies](#).

Available Versions

Current Version 1.1

What has changed: [Changelog \(github\)](#)

Online: [Resource Type Vocabulary v.1.1 \(October 2016\)](#)

SKOS-XL: [SKOS-XL RDF file \(github\)](#).

Zusammenfassung

- Protokollfragen
- Metadatenqualität
- **Weitere Repository-Formen**
 - Digitale Sammlungen
 - Forschungsdaten-Repositories
- Repository-Tools

Forschungsdaten als Bestandteil von Repositorien : In Großbritannien weniger als 1 % aller Publikationen



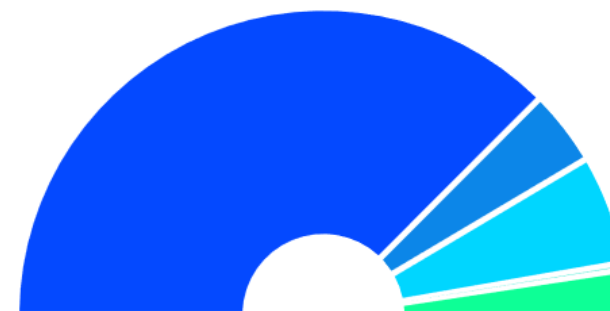
United Kingdom

4719998 Documents from 288 Repositories

4 % of Documents world-wide, 5 % of Repositories world-wide



OA Status Distribution (Number of Documents)



Repository Type Distribution (Number Documents)





United Kingdom

4719998 Documents from 288 Repositories

4 % of Documents world-wide, 5 % of Repositories world-wide

```
var reptyes = [  
  { 'mean': 'IR' , 'value': 3529908}, // },  
  { 'mean': 'Journals' , 'value': 348195} , // },  
  { 'mean': 'Digital Collections' , 'value': 5175591} // },  
  { 'mean': 'Research Data', 'value' : 3890},  
  { 'mean': 'other', 'value' : 196434},  
];
```

Other

Forschungsdaten-Repositories

- Neue Plattformen mit Startproblemen
- Etabliert: Eprints, Dataverse, Fedora, Dspace, Invenio
- Cloud-Nutzung via Figshare
- Bisher eher geringe Objektzahlen

Forschungsdaten-Repositories in BASE

Forschung

- Neue Plattfo
- Etabliert: Ep
- Cloud-Nutzu
- Bisher meist

Nr.	Int. ID	Datum	Harvestername	System	Kont.	Land	Bundesst.	Breitengr.	Längengr.	OAI Basic URL	OA	OASet	Typ
													Forschungsdaten
1.	6602	15.08.2017	Caltech-Data	Invenio	cna	us				http://data.caltech.edu/oai2d	1		Forschungsdaten
2.	6458	16.06.2017	REPOD		ceu	pl				https://repod.pon.edu.pl/oai	2		Forschungsdaten
3.	6146	14.03.2017	UnivAston-DE	Eprints 3	ceu	uk				http://researchdata.aston.ac.uk/cgi/oai2	2		Forschungsdaten
4.	6105	03.03.2017	SheffieldHallamUniv-RDA	Eprints 3	ceu	uk		53.379106	-1.466066	http://shurda.shu.ac.uk/cgi/oai2	2		Forschungsdaten
5.	6104	03.03.2017	LeedsUniv-RD	Eprints 3	ceu	uk		53.808005	-1.553456	http://archive.researchdata.leeds.ac.uk/cgi/oai2	1		Forschungsdaten
6.	6040	17.02.2017	UnivLiverpool-RDC	Eprints 3	ceu	uk				http://datacat.liverpool.ac.uk/cgi/oai2	1		Forschungsdaten
7.	6031	16.02.2017	UnivBath-RDA	Eprints 3	ceu	uk				http://researchdata.bath.ac.uk/cgi/oai2	1		Forschungsdaten
8.	6030	16.02.2017	PurdueUniv-PURR		cna	us				https://purr.purdue.edu/oaipmh	1		Forschungsdaten
9.	6029	15.02.2017	UnivReading-RDA	Eprints 3	ceu	uk				http://researchdata.reading.ac.uk/cgi/oai2	1		Forschungsdaten
10.	6016	14.02.2017	UnivHeidelberg-Heidata	Dataverse	ceu	de	debw			https://fdmsrv1.urz.uni-heidelberg.de/dvn/OAIHandler	1		Forschungsdaten
11.	6007	13.02.2017	UGlasgow-ERD	Eprints 3	ceu	uk				http://researchdata.gla.ac.uk/cgi/oai2	2		Forschungsdaten
12.	6006	13.02.2017	LMUMuenchen-OpenData	Eprints 3	ceu	de	deby			https://data.ub.uni-muenchen.de/cgi/oai2	1		Forschungsdaten
13.	5621	08.11.2016	Dados	DSpace XOAI	ceu	pt				http://dados.rcaap.pt/oaextended/request	2		Forschungsdaten
14.	5441	08.09.2016	UnivTartu-Datadoi	DSpace XOAI	ceu	ee		58.381163	26.721089	http://datadoi.ut.ee/oai/request	2		Forschungsdaten
15.	5404	25.05.2016	LSHTM Data	Eprints 3	ceu	uk				http://datacompass.lshtm.ac.uk	2		Forschungsdaten

Zusammenfassung

- Protokollfragen
- Metadatenqualität
- Weitere Repository-Formen
- **Repository-Tools**
 - Usage Statistics
 - OpenAPC-Auswertung
 - Repository-Analysen



PARTICIPATE

SEARCH

OpenAIRE Guide for Usage Statistics

Updated on 06 October 2015



HARVESTING USAGE DATA

Aggregation requires standardization on the recording of user events, exclusion of robot accesses and data exchange mechanisms. The [OpenAIRE Guidelines for Usage Statistics \(v1.0\)](#), based on OAI-PMH, specify a common description of usage events for a straightforward adoption by data providers.

LEGAL ISSUES - PRIVACY POLICY

In alignment with the [European Act of personal data protection](#), the IP address, session-id and in some cases also the C-class Subnet must be pseudonymised before transferring the usage data to the aggregator service.



Find more

Guides for

[Repository Managers](#)

[OA Journals](#)

[Repository Managers - Usage Statistics](#)

[Register & Validate](#)

[Enrich metadata](#)

[Intro for Data Providers](#)

The Open APC initiative releases datasets on fees paid for Open Access journal articles by universities and research institutions under an Open Database License.

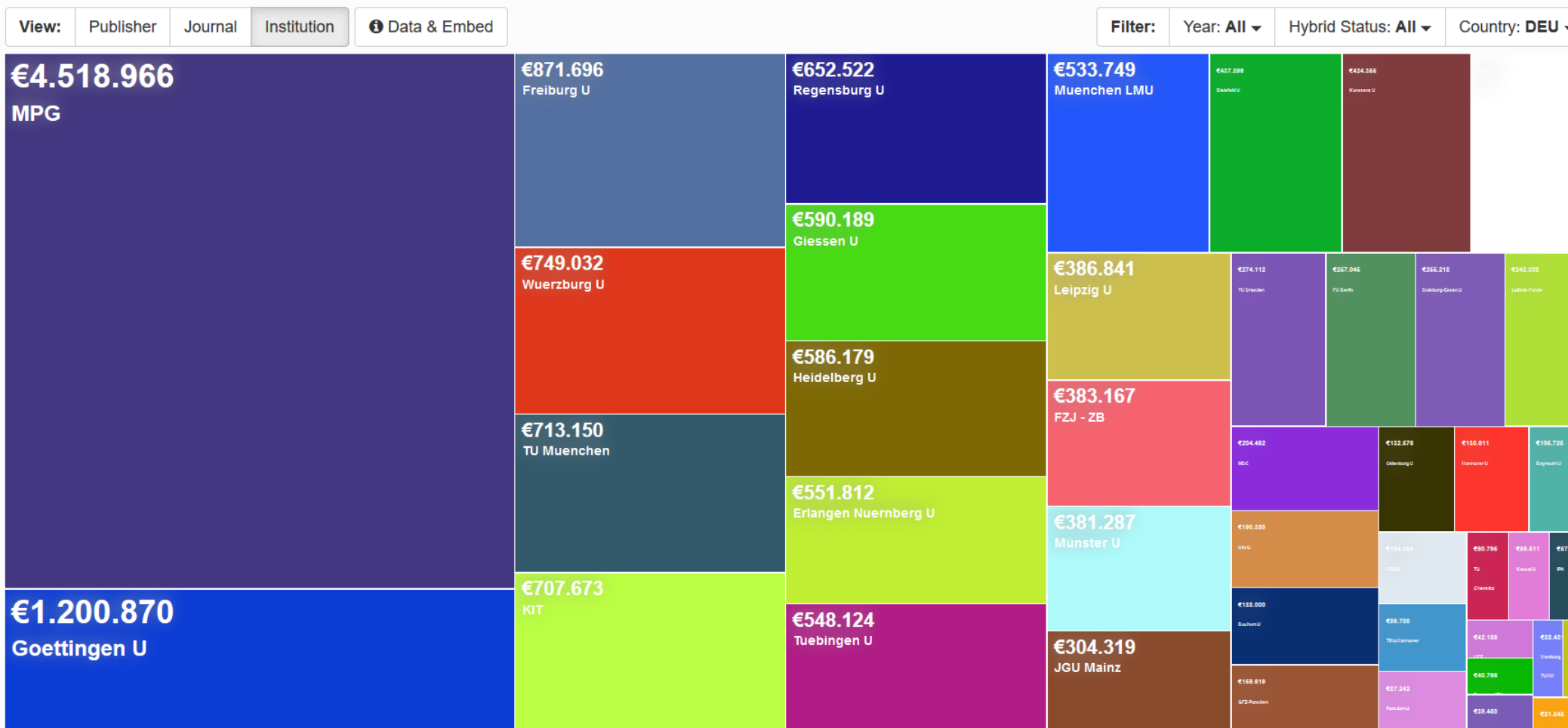
Select a region from the map or use the list below to view APC data for individual institutions.

You may also browse the following aggregated collections:

- The [OpenAPC](#) dataset is an aggregation of all institutional APC data.
- The [Offsetting](#) dataset contains articles published under offsetting deals such as the Springer Compact pilot agreements.



OPENAPC



Repository-Analysen: Beispiel Repository-Netzwerk Deutschland



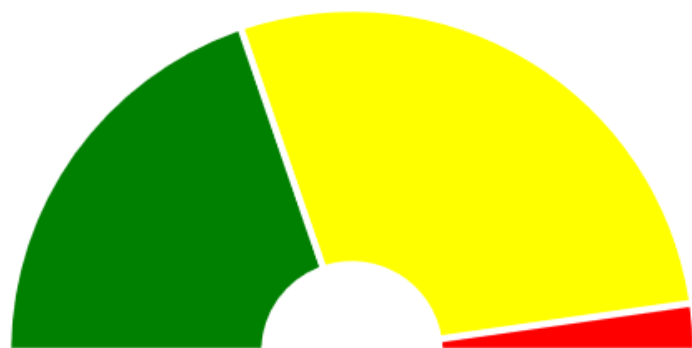
BASE Statistics: - Content Sources



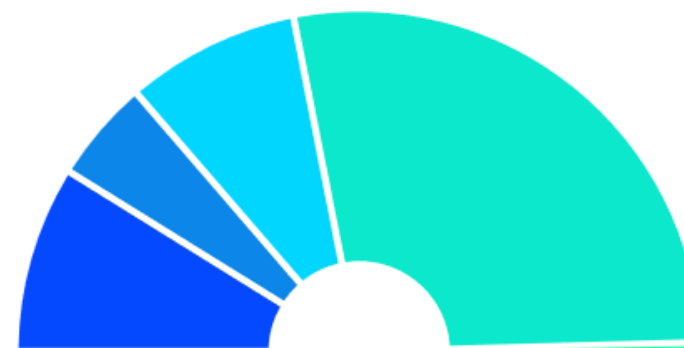
Germany

13532749 Documents from **354** Repositories

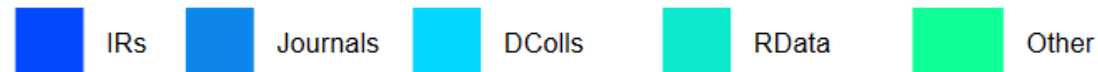
12 % of Documents world-wide, 6 % of Repositories world-wide



OA Status Distribution (Number of Documents)

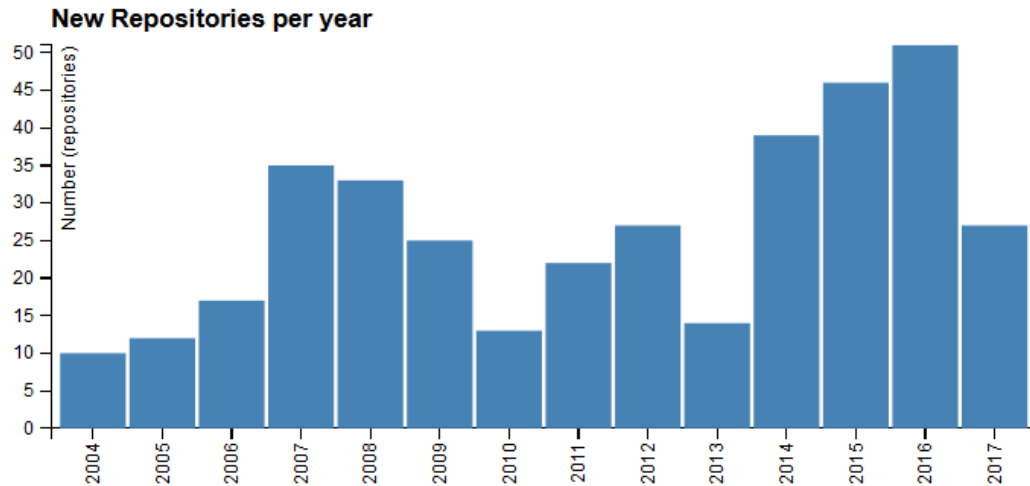


Repository Type Distribution (Number Documents)

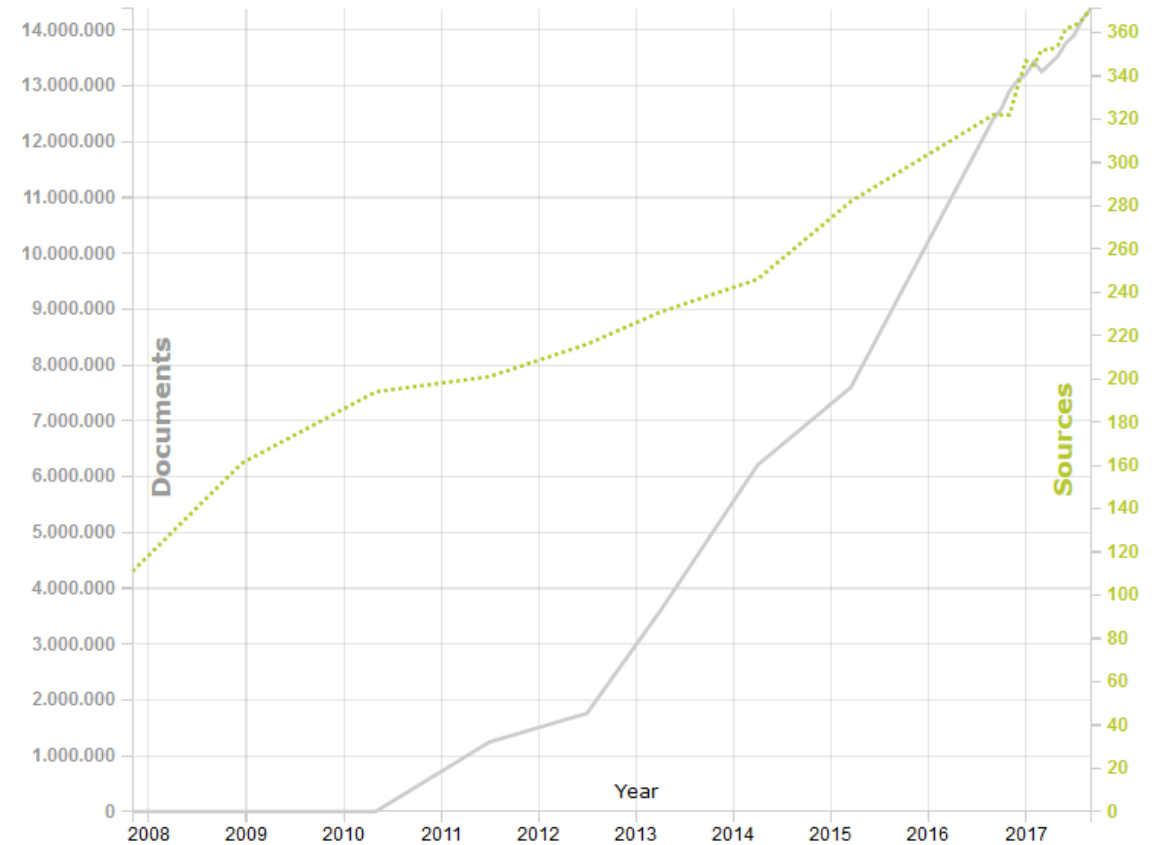


Repository-Analysen: Beispiel Repository-Netzwerk Deutschland (2)

Repository Development



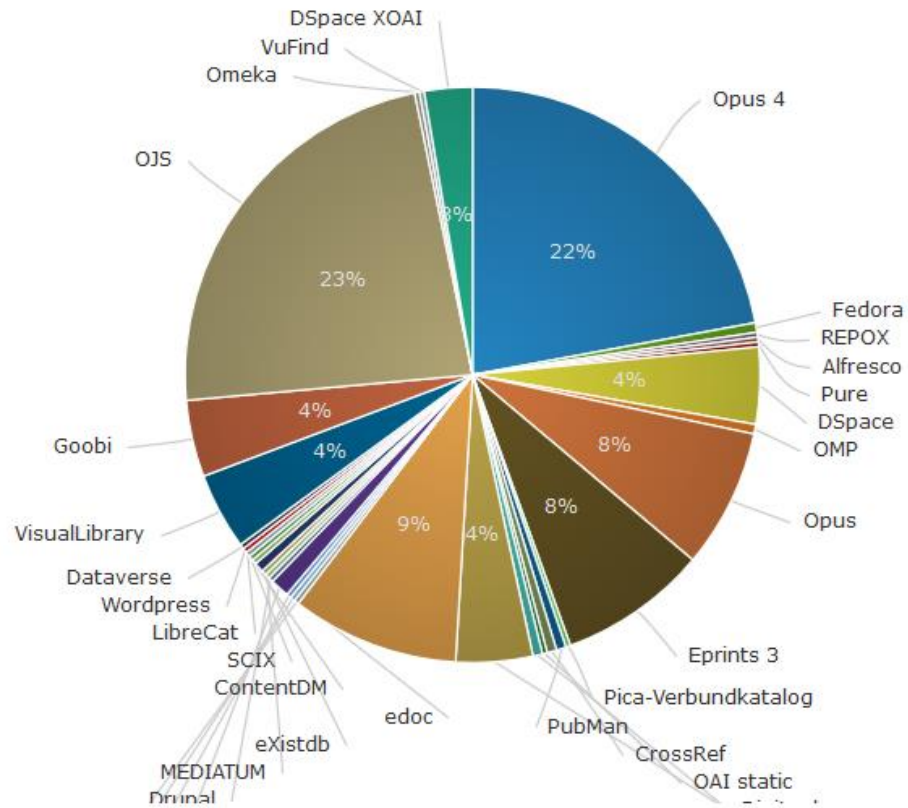
Number of indexed Documente und Sources in BASE for Germany



Repository-Analysen: Beispiel Repository-Netzwerk Deutschland (3)

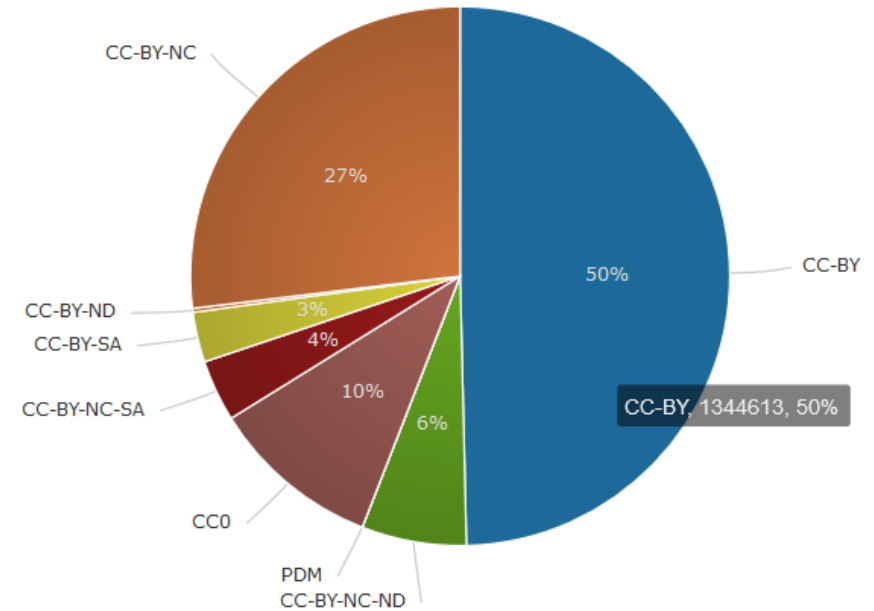
Platform System Distribution

(number of repositories)



License Type Distribution

(number of documents)



Percentage Records with License Information = 18.84 %



Repository-Analysen: Beispiel Einzelnes Repository



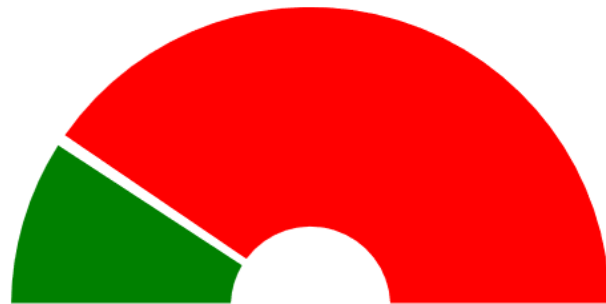
PUB - Publikationen an der Universität Bielefeld (PUB - Publications at Bielefeld University)

54823 Documents, 9935 Open Accessible

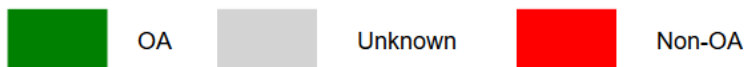
Repository Type: Hochschulpublikationen

System: LibreCat

In BASE indexed since: 2011-10-07



OA Status Distribution (Number of Documents)





Bielefeld Academic Search Engine

BASE Statistics: - Content Sources



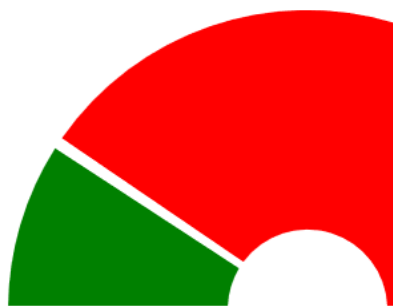
PUB - Publikation
(PUB - Publications

54823 Documents, **9935** Open Access

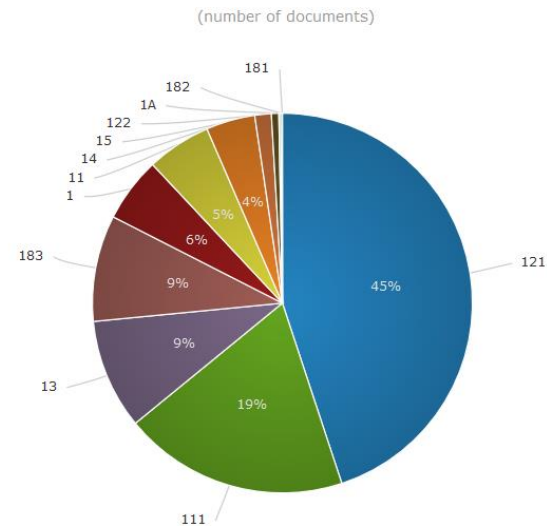
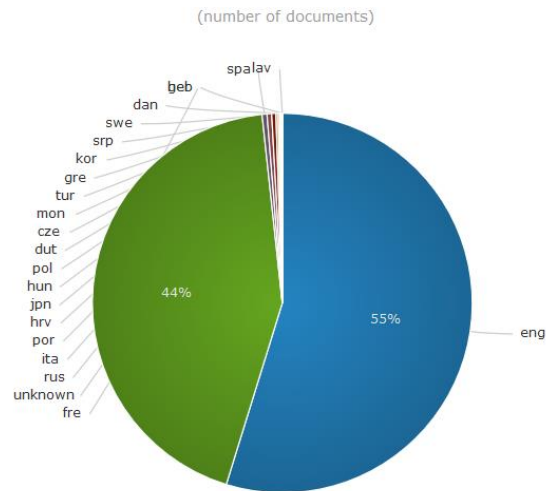
Repository Type: Hochschulpublikationen

System: LibreCat

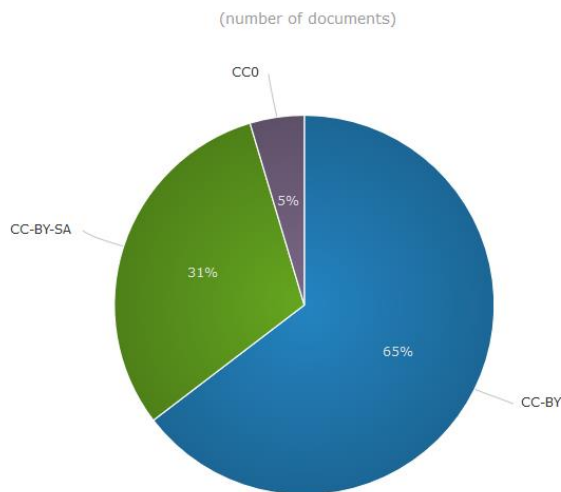
In BASE indexed since: 2011-10-07



OA Status Distribution (Number of Documents)



Licence Type Distribution



Repository-Analysen: Mögliche Aspekte

Scope:

- Global
- National
- Repository

Attribute

- Repository-Type
- Repository-Plattform
- OA-Status
- Lizenzinformationen
- Language
- Publikationstyp
- Date
- ...

Möglich ist eine flexible Auswertungsschnittstelle, die die obigen Aspekte in beliebiger Weise kombinieren lässt und die bei BASE vorliegenden umfangreichen Daten (inkl. historischer Informationen) bereitstellt.

Danke für Ihre Aufmerksamkeit!

<https://www.base-search.net>

friedrich.summann@uni-bielefeld.de