

Open Access Repositories and Their Infrastructure

Principles, Practices and Challenges of Exposing
Scientific Publications

Friedrich Summann

Bielefeld University Library

4th NATIONAL OPEN ACCESS WORKSHOP

October, 19-21, 2015, TÜBİTAK, Ankara

Overview

- **Institutional Repositories - past and current situation**
 - The IR network
 - IRs in Turkey
 - Challenges and Future perspectives
-

Bielefeld University Library and the IR Background

- Pre-IR-System since 1999
 - IR Platform since 2004
 - BASE Search Engine since 2004
 - Open Access Declaration 2005
 - Publication Management System since 2010
 - EU DRIVER Project (2006 – 2009)
 - EU OpenAire Projects (2009 -)
 - DINI (German Network Initiative) Member
 - COAR (Confederation of Open Access Repositories)
Member
 - Research Data Declaration 2013
-

The IR – **past**, present, future

- Started late nineties
 - Contents starting with thesis
 - OAI-PMH protocol definition 2001
 - Open Access movement
 - Establishing a global repository network
 - Extending
 - Size
 - Quality
 - Services
-

The IR – past, **present**, future

- More than 3700 repositories, around 80 Mill. Objects
 - World-wide coverage
 - Strong Community, strong Network
 - But: Institutional Repositories are at a turning-point:
 - More and more overlapping systems
 - local (CRIS, Publishing Platforms, etc)
 - external (Subject Repositories, ResearchGate etc.)
 - Scholarly Communication process changes
-

BASE, OAI Service Provider and mirror of the IR landscape

- 3751 Repositories included (using OAI-PMH)
 - From 103 Countries world-wide
 - Ca. 79 Mill. Documents
 - Ca. 70 % Open Accessible
 - Dublin Core Format
 - Ca. 10.9 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
-

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
-

Contents:

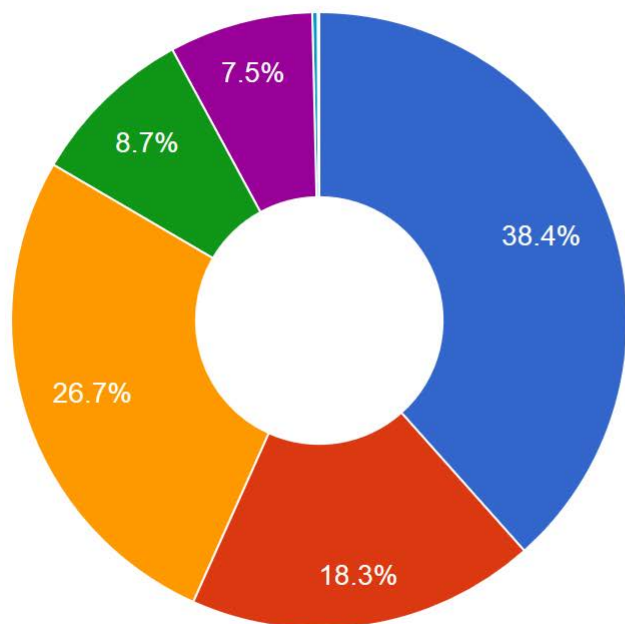
Global Scientific Metadata harvested from
Repositories

Technology:

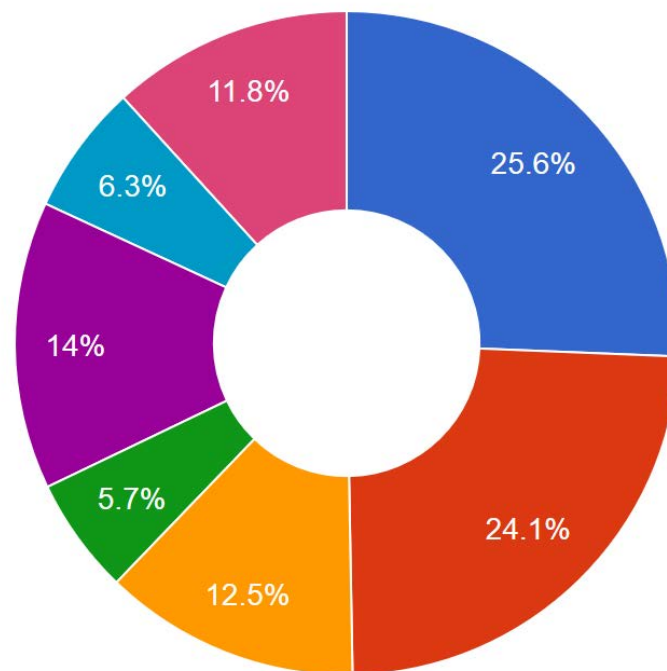
- Index (Solr)
- Search Interface (VuFind)
- APIs (HTTP, OAI)

BASE harvests, aggregates, enriches and exposes
OAI Metadata

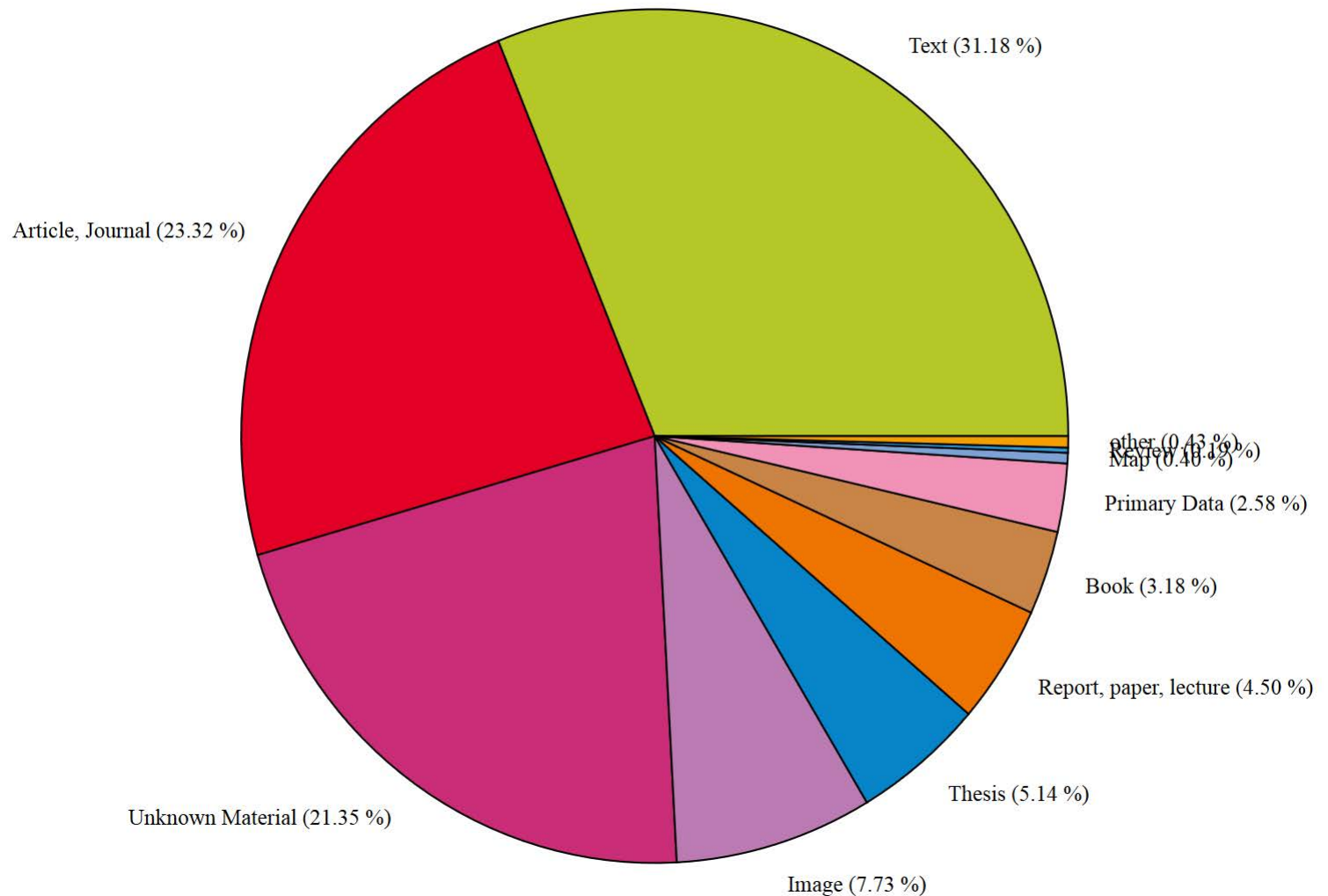
Repository Types covered in BASE



- Institutional Repositories
- Publication Server
- Electronic Journals
- Thesis/Dissertation Servers
- Digital Collections
- Research Data
- Sonstiges

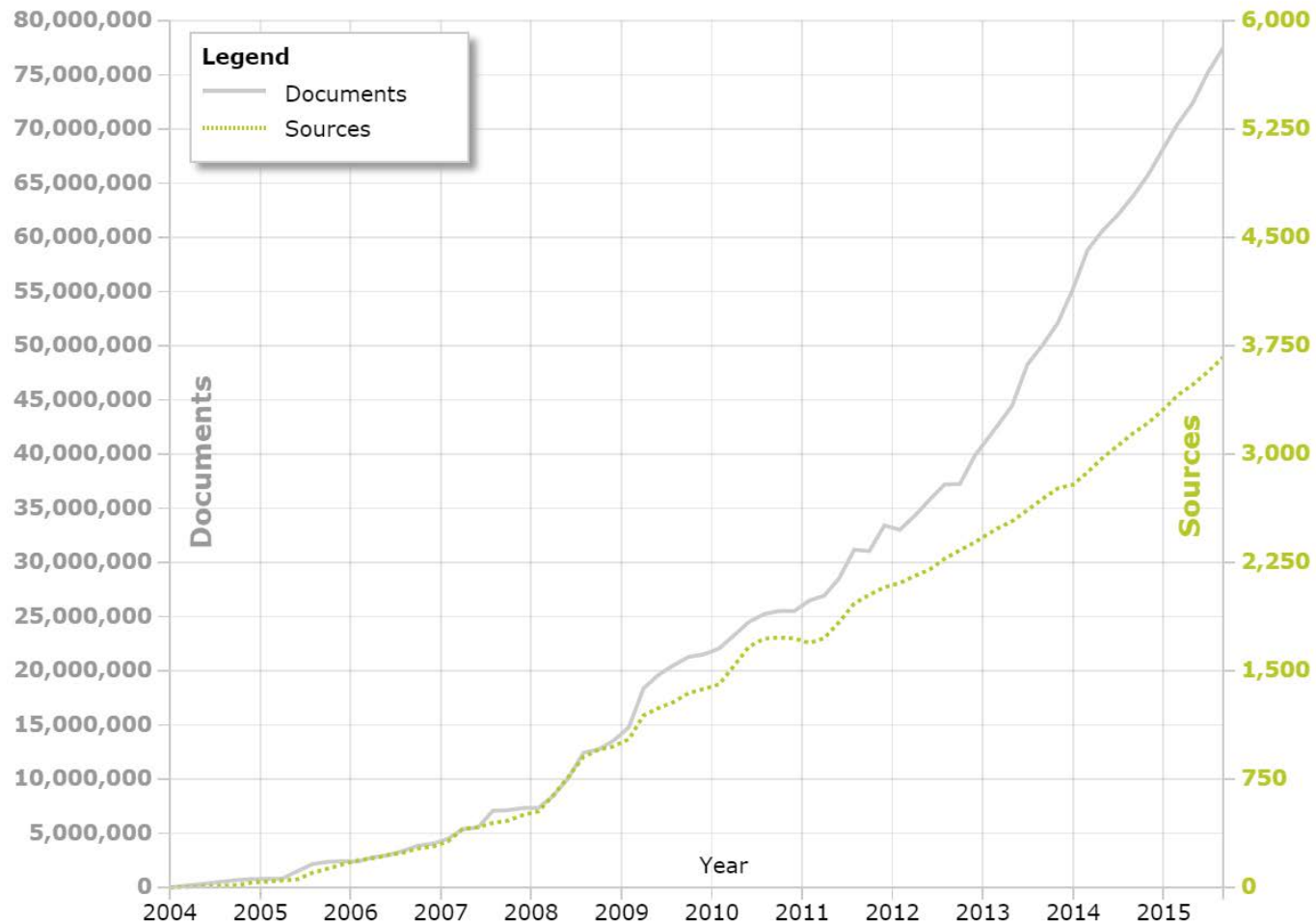


Distribution Publication Type

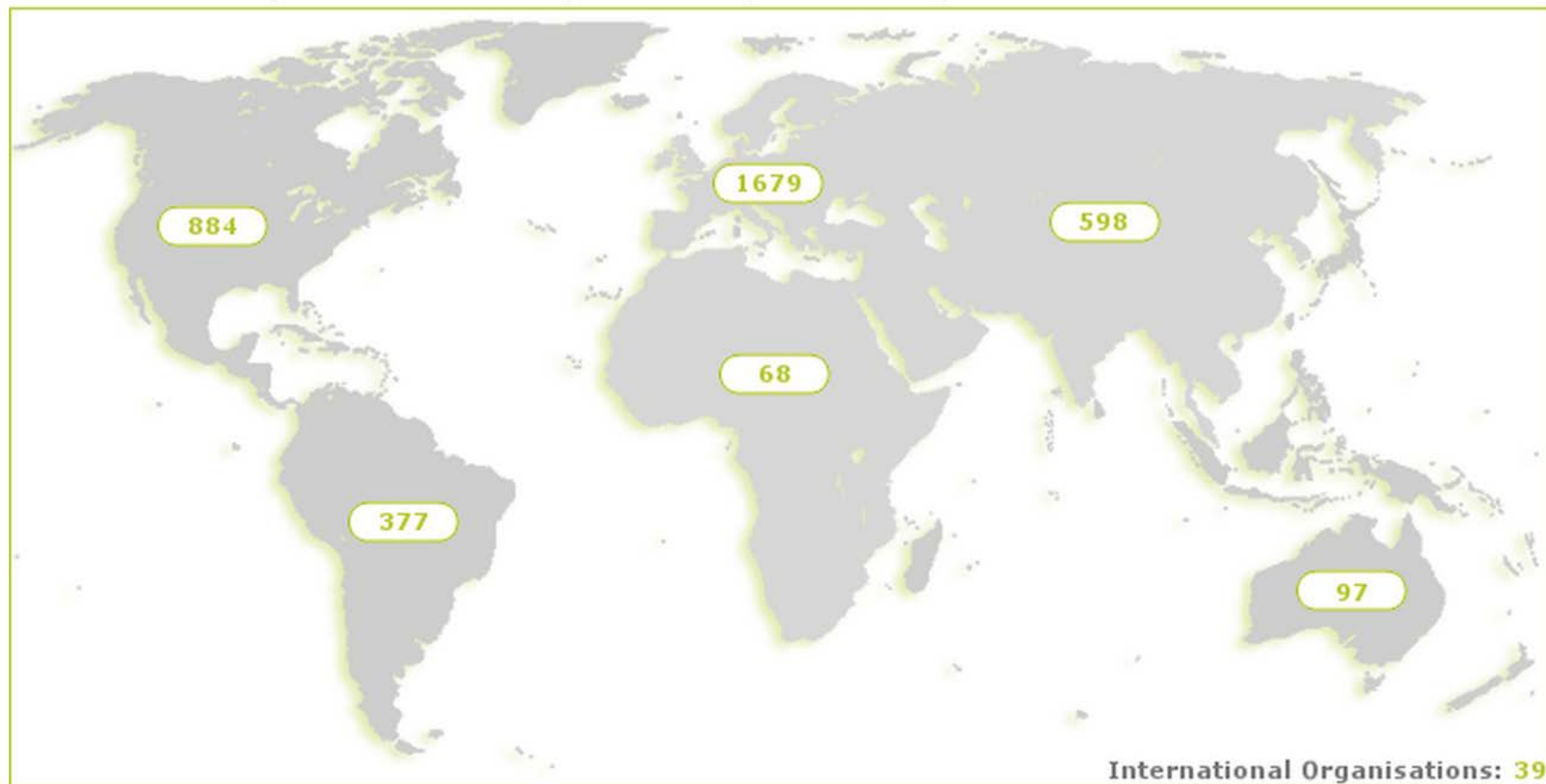


Development of the number of indexed sources and documents in BASE since September 2004.

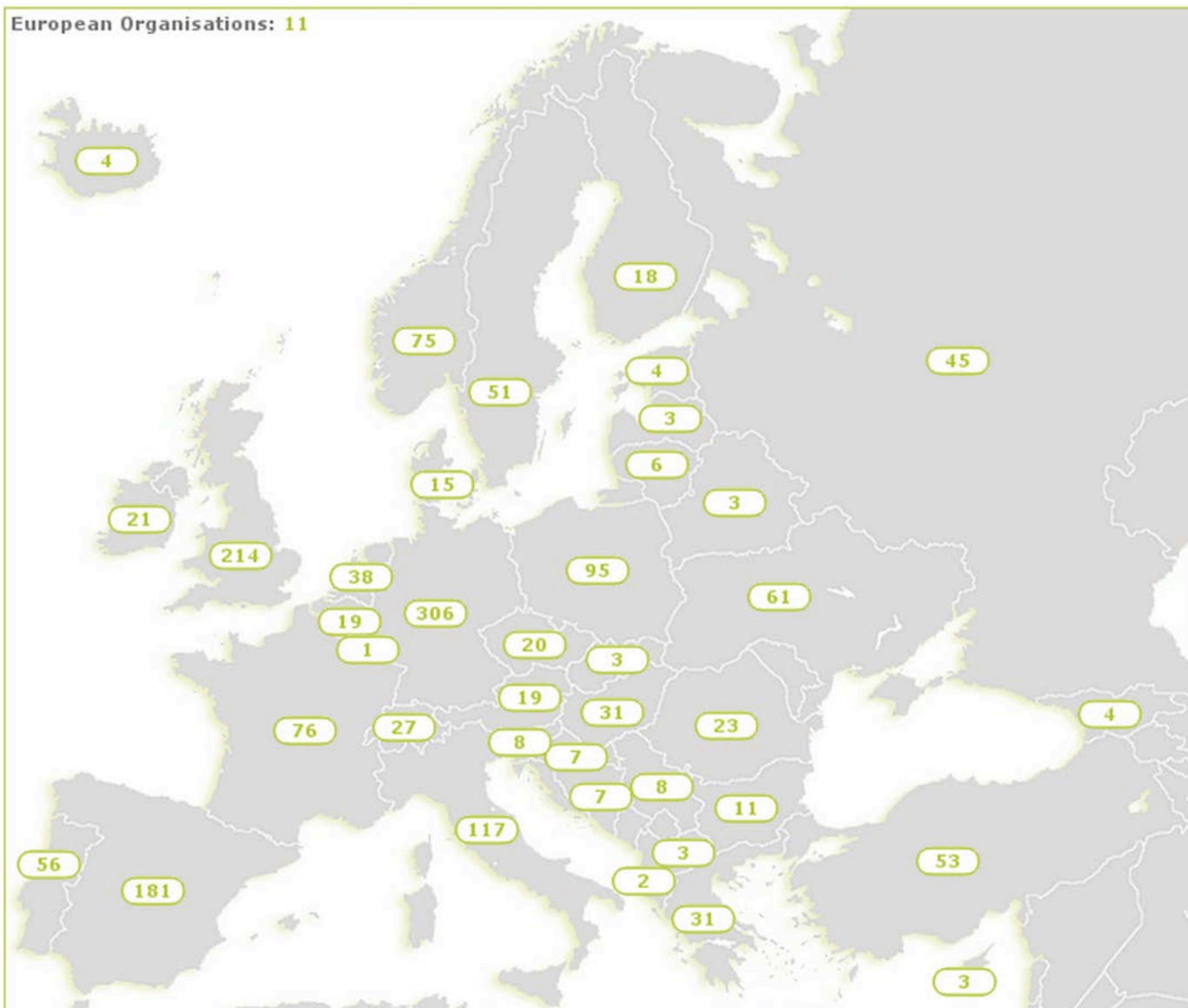
Chart: Number of indexed Documents and Sources



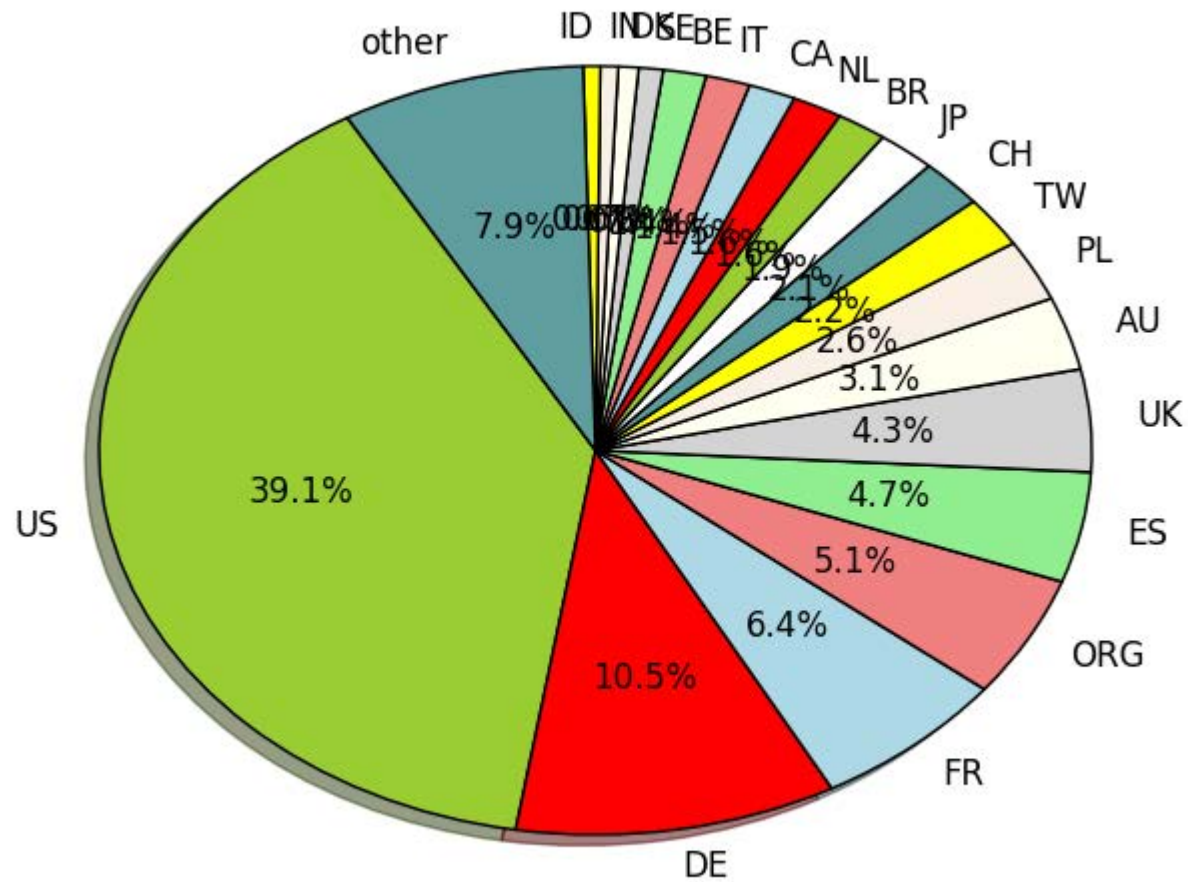
Sources indexed by BASE world wide by continent: (3742 in total)



European Organisations: 11



Distribution Country (nr of docs)




Overview

- Institutional Repositories - past and current situation
 - **The IR network**
 - IRs in Turkey
 - Challenges and Future perspectives
-

The IR Network – Participants and Stakeholders

- IR as Data Providers
 - Service Providers (OAlster, BASE, Google Scholar, ResearchGate etc.)
 - Registries (OpenDOAR, OpenArchives, ROAR ...)
 - Projects (national, international)
 - Organizations (national, international)
 - Funder Organizations
 - Publishers, Information Companies
 - Users, Authors, related Institutions
-

Country 🏳️	Content Sources 📄	Documents 📄	📡 Feed
United States of America	733	29,064,707	RSS ATOM
Germany	306	9,419,634	RSS ATOM
Brazil	219	1,347,371	RSS ATOM
United Kingdom	214	3,712,846	RSS ATOM
Spain	181	3,574,463	RSS ATOM
Japan	180	1,414,476	RSS ATOM
Indonesia	122	453,608	RSS ATOM
Italy	117	1,050,309	RSS ATOM
India	109	443,962	RSS ATOM
Canada	106	1,068,387	RSS ATOM
Poland	95	2,003,737	RSS ATOM
Australia	81	2,424,820	RSS ATOM
France	76	4,744,214	RSS ATOM
Norway	75	166,593	RSS ATOM
Ukraine	61	369,287	RSS ATOM
Portugal	56	304,326	RSS ATOM
Taiwan	54	1.433.782	RSS ATOM
Turkey	53	336,654	RSS ATOM
Sweden	51	632,819	RSS ATOM
Colombia	48	281,297	RSS ATOM
Russia	45	883,435	RSS ATOM
Argentina	39	212,326	RSS ATOM
International Organisations	39	3,988,097	RSS ATOM
Netherlands	38	1,068,207	RSS ATOM
Greece	31	337,016	RSS ATOM
Hungary	31	283,181	RSS ATOM
Switzerland	27	1,488,470	RSS ATOM

Country  	Content Sources  	Documents  	 Feed
United States of America	733	29,064,707	RSS ATOM
Germany	306	9,419,634	RSS ATOM
France	76	4,744,214	RSS ATOM
International Organisations	39	3,988,097	RSS ATOM
United Kingdom	214	3,712,846	RSS ATOM
Spain	181	3,574,463	RSS ATOM
Australia	81	2,424,820	RSS ATOM
Poland	95	2,003,737	RSS ATOM
Switzerland	27	1,488,470	RSS ATOM
Taiwan	54	1,433,782	RSS ATOM
Japan	180	1,414,476	RSS ATOM
Brazil	219	1,347,371	RSS ATOM
Belgium	19	1,078,884	RSS ATOM
Canada	106	1,068,387	RSS ATOM
Netherlands	38	1,068,207	RSS ATOM
Italy	117	1,050,309	RSS ATOM
Russia	45	883,435	RSS ATOM
Denmark	15	678,658	RSS ATOM
Sweden	51	632,819	RSS ATOM
Finland	18	482,409	RSS ATOM
Indonesia	122	453,608	RSS ATOM
India	109	443,962	RSS ATOM
Ukraine	61	369,287	RSS ATOM
Greece	31	337,016	RSS ATOM
Turkey	53	336,654	RSS ATOM
South Africa	25	326,411	RSS ATOM
Portugal	56	304,326	RSS ATOM

OpenDOAR: List of Turkish Repositories

UB Wiki - EDV/Op x No title x showbase.py (800 x Suchmaschine BA x Suchmaschine BA x OpenDOAR - Sum x DINE: DINI-Zertifik x Webmail Uni-Biel x

www.opendoar.org/find.php

Apps Erste Schritte http://www.revistag... Web Slice-Katalog Aus Firefox importiert

OpenDOAR

Directory of Open Access Repositories

Home | Find | Suggest | Tools | FAQ | About | Contact Us

Search or Browse for Repositories

Recent Additions RSS Feed

Any Subject Area Any Content Type Any Repository Type
China (39) Any Language Any Software Search
Summaries All per page. Sort by: Repository Name New Query

To search the contents of the repositories listed in OpenDOAR, please see our [Content Search](#) page.

Results 1 - 39 of 39. [RSS1 Feed for this Query](#)

CityU Institutional Repository

Organisation: [Run Run Shaw Library](#), [City University of Hong Kong](#) (香港城市大學), China

Description: This is an institutional repository preserving the scholarly output of the University community. This repository mainly contains theses and dissertations.

OAI-PMH: <http://dspace.cityu.edu.hk/oai/request>

Software: DSpace

Size: 3943 items (2014-10-16)

Subjects: Multidisciplinary

Content: Articles; Theses; Special

Languages: English

Policies: Metadata re-use policy explicitly undefined; Full data item policies explicitly undefined; Content policies explicitly undefined; Submission policies explicitly undefined; Preservation policies explicitly undefined

Remarks: Special items include student final year projects.

OpenDOAR ID: 1017, Last reviewed: 2012-09-11, [Suggest an update for this record](#), Missing data is needed for: [Policies](#)

Link to this record: <http://opendoar.org/id/1017/>

IR Network Activities



Establishment and development of the European open access repository infrastructure

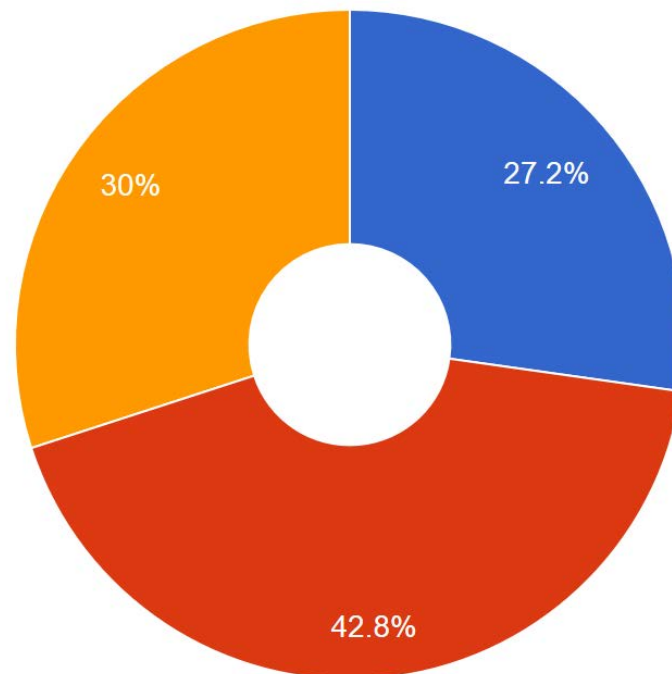
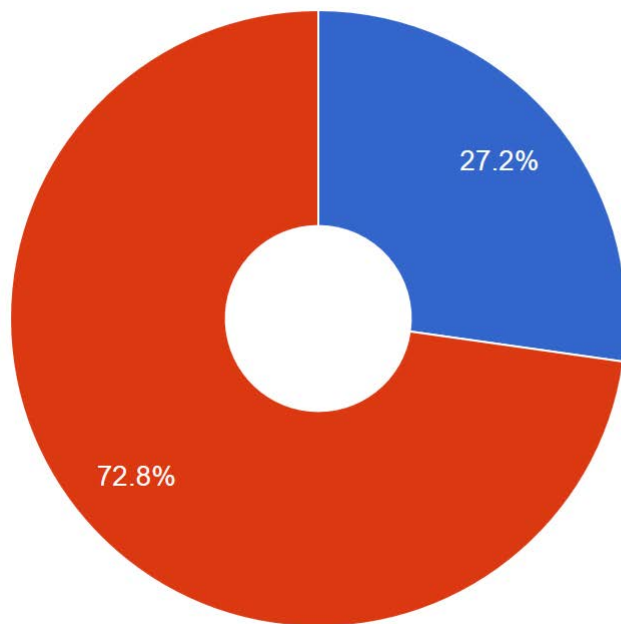


Building and Extending the Open Access Infrastructure for Research in Europe

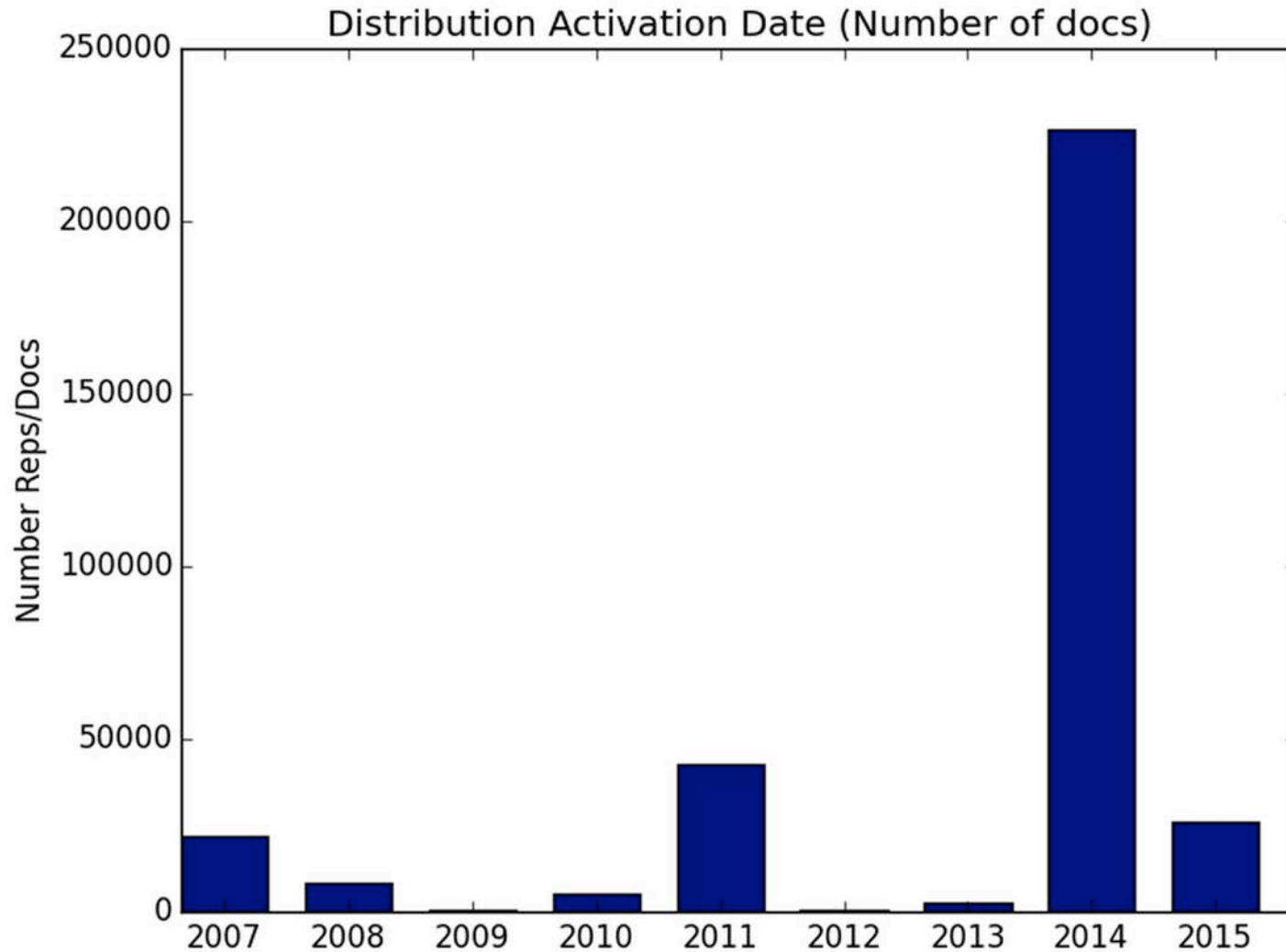


Greater visibility and application of research through global networks of Open Access repositories

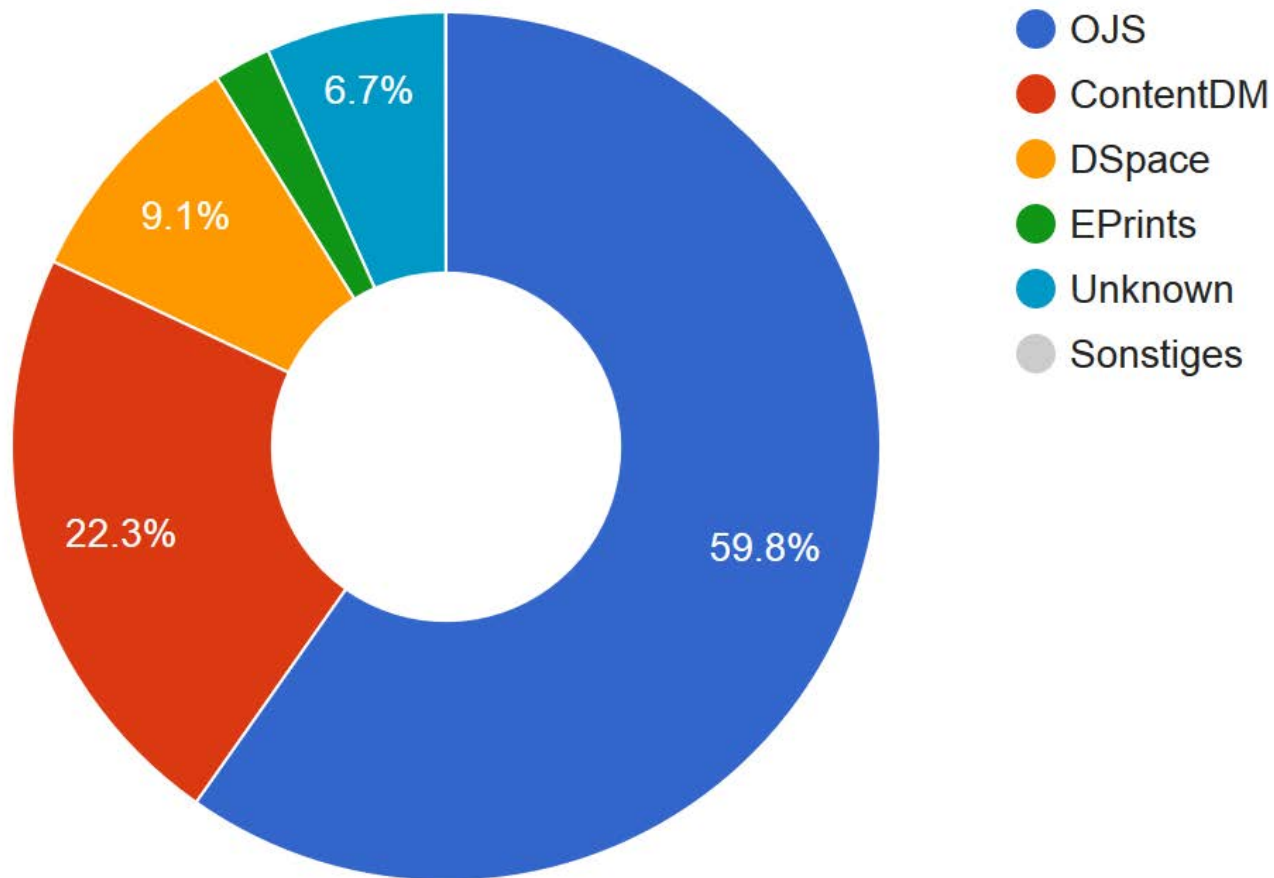
Open Access Status



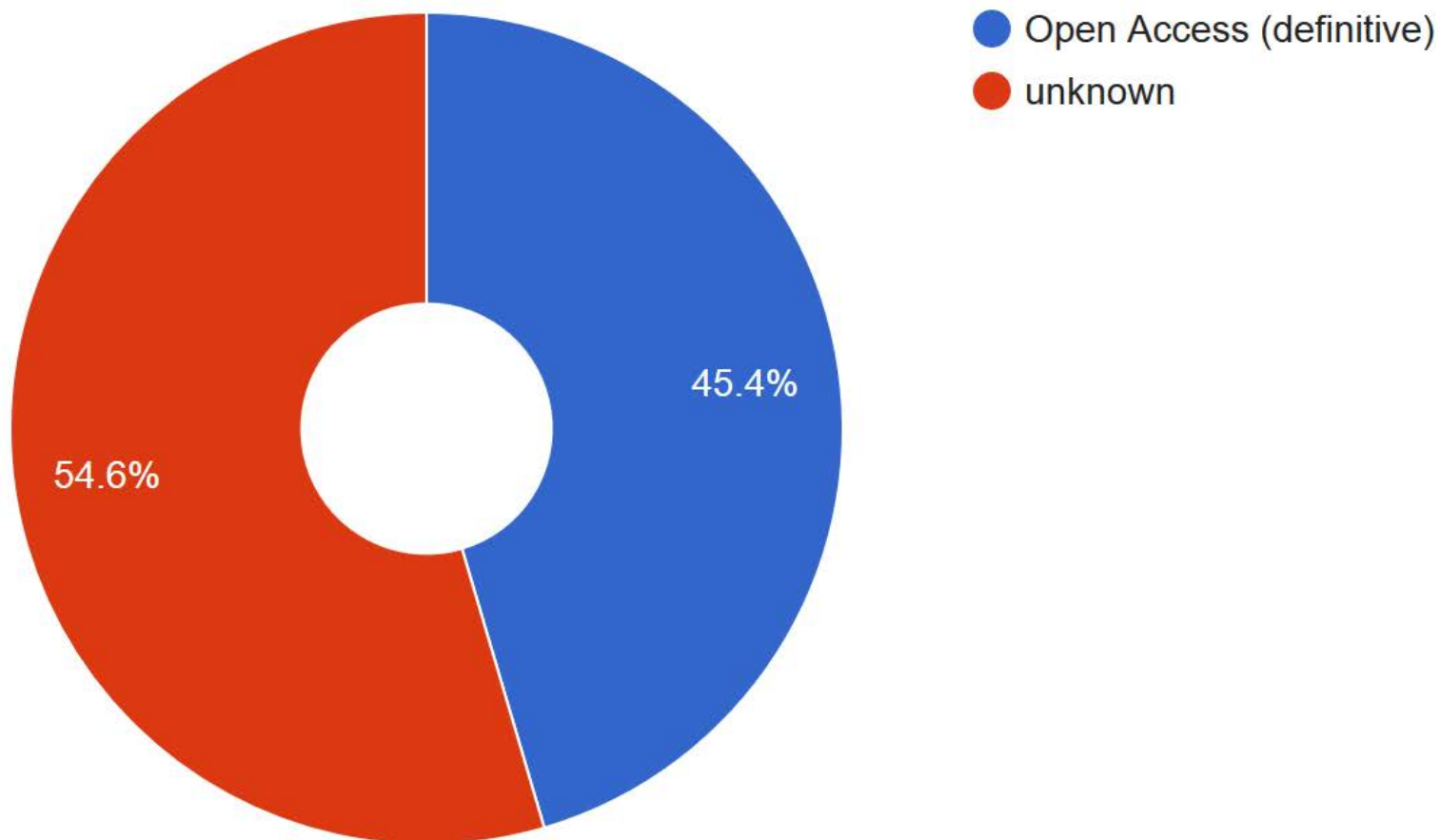
Timeline Activation of Turkish Repositories in BASE (number of documents)



Repository Platforms in Turkey - Number of Documents



Open Access Status - Turkey



Overview

- Institutional Repositories - past and current situation
 - The IR network
 - IRs in Turkey
 - **Challenges and Future perspectives**
-

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

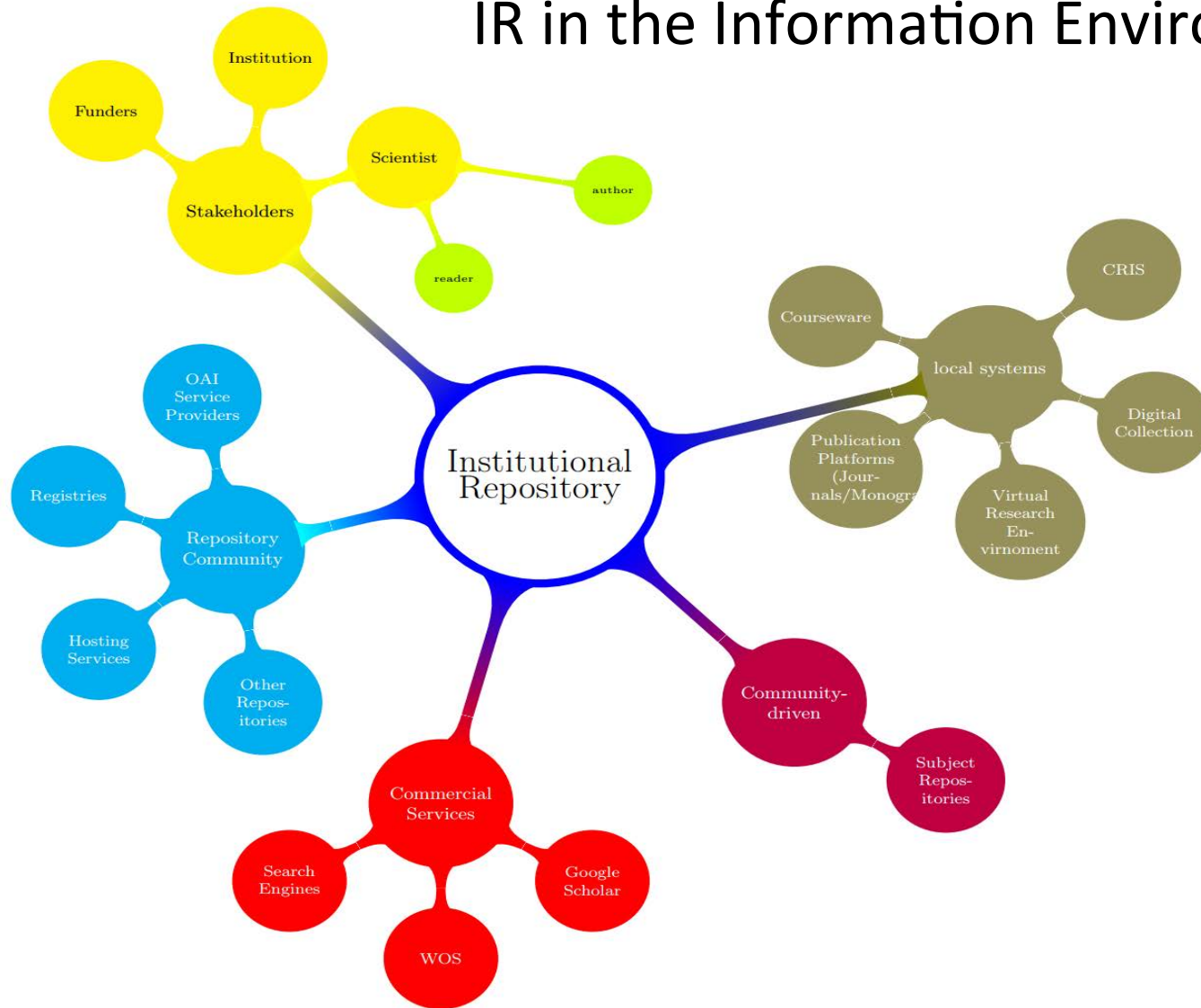
(a specific Dspace problem with non-configured links as
<dc:identifier>http://hdl.handle.net/123456789/87639</
dc:identifier>)

- 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 ...)
-

IR in the Information Environment



Impact and Visibility

- Supporting Search Engine Optimization (SEO)
- Supporting Repository Ranking Systems
- Exposing Usage Statistics
- Exposing Bibliometric Information
- Supporting Visibility in Repository Registries
- Improving Registry Infrastructure

Usability

- Supporting Authorization and Authentication
- Supporting Embedding Services
- Exposing Publication Lists
- Exposing Citation Formats
- Supporting Data Export Functions
- Integrating Availability Services
- Supporting Author Identification Systems
- Supporting Institutional Services
- Extending End-User Usability
- Extending Usage of Visualization Tools

Sustainability

- Improving Platform Stability
- Supporting Long-term Preservation and Archiving
- Exposing Persistent Identifiers
- Integrating different Persistent Identifiers

Data Issues

- Supporting additional Metadata Format(s)
- Improving Metadata Quality (Data Curation)
- Supporting Enhanced Publications
- Supporting Linked (Open) Data
- Publication of Research Data
- Handling of Complex/Compound/Nested Repository Objects
- Monitoring Open Access Mandate Compliance
- Exposing Versioning Information

Validation and Aggregation

- Validating Repository Metadata
- Processing Related Full-text
- De Duplication

Technical Issues

- Defining Architectural Recommendations for Repositories and their Interoperability
- Extending/Replacing Metadata Exposition Protocols
- Supporting OAI Service Provider Usage
- Supporting Deposit Protocols

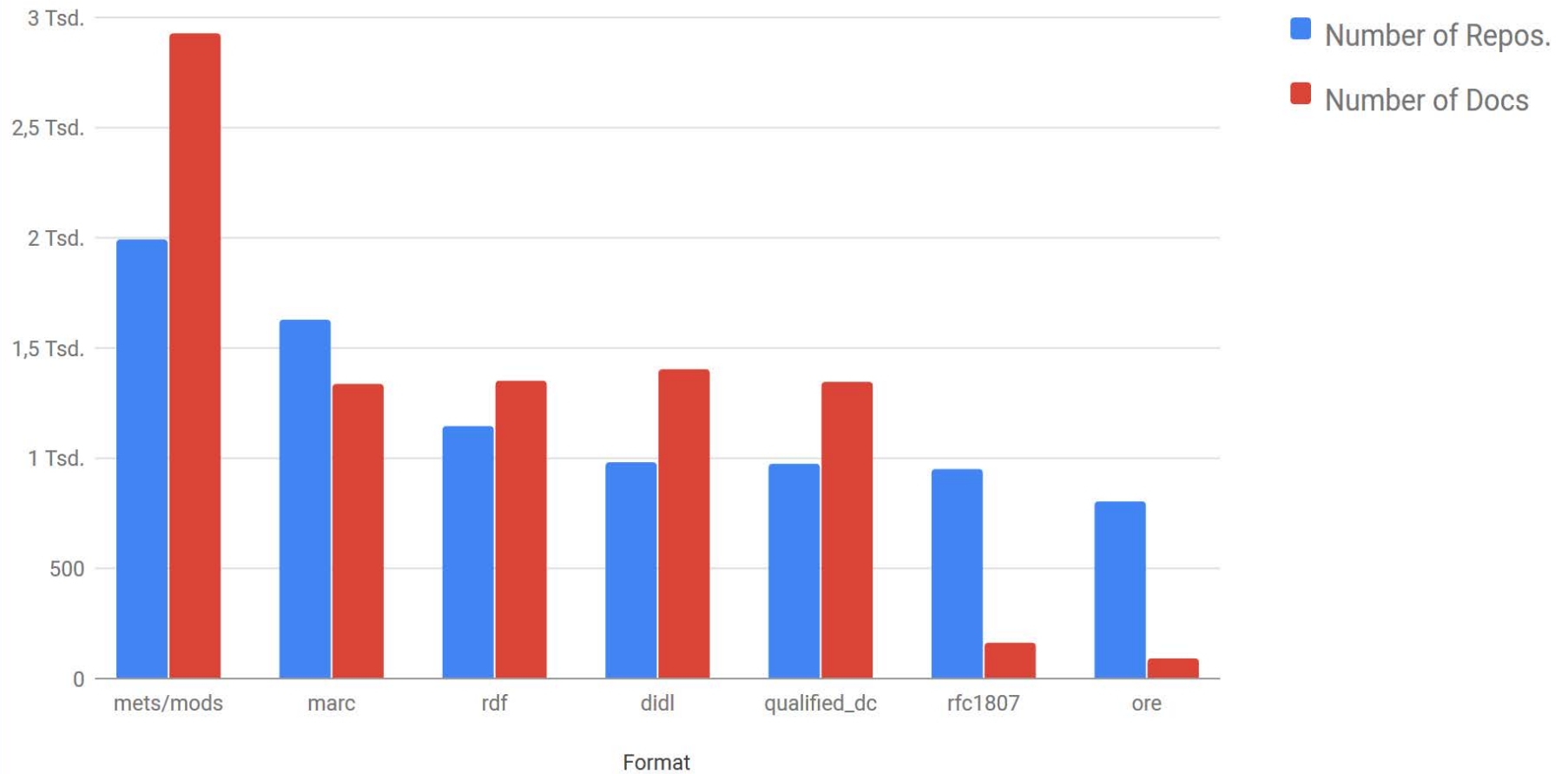
Issue: Data Enrichment (Linked Open Data Strategies)

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



Issue: Link Resolver Service

- OAI Identifier
 - DOI/Handle/ISSN/ISBN/URN/PMCID
 - Author/Organization/Funder IDs
 - Bibliographic Metadata
-

Issue: OA Status, Rights and Licences Normalization

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

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://creativecommons.org/licenses/by/2.0>

2x 504430

Issue: Big Data Activities

- Data Enrichment
(Linked Open Data Strategies)
 - Automatic Classification
 - De-Duplication/Version Detection
 - Fulltext Indexing
-

„Let us increase the number of academic OA documents and their metadata quality together and let us increase the visibility of Turkish academic output worldwide!“

Thank you for your attention!

- friedrich.summann@uni-bielefeld.de
-