

OpenAire and BASE

Services supporting the Interoperability of the
European Open Science Network

Lyon, 20.11.2018

Friedrich Summann

ORCID ID

 orcid.org/0000-0002-6297-3348

▼ **Employment (1)**

Bielefeld University: Bielefeld, Northrhine-Westphalia, Germany
Head LibTec (Library)

Source: Friedrich Summann

Created: 2015-06-01

Overview

- Introduction
- BASE
- OpenAire
- Comparing OpenAire and BASE

Se connecter

Français

Recherche simple

Recherche avancée

Navigation

Historique de recherche



Search 136 280 107 documents from 6 642 sources

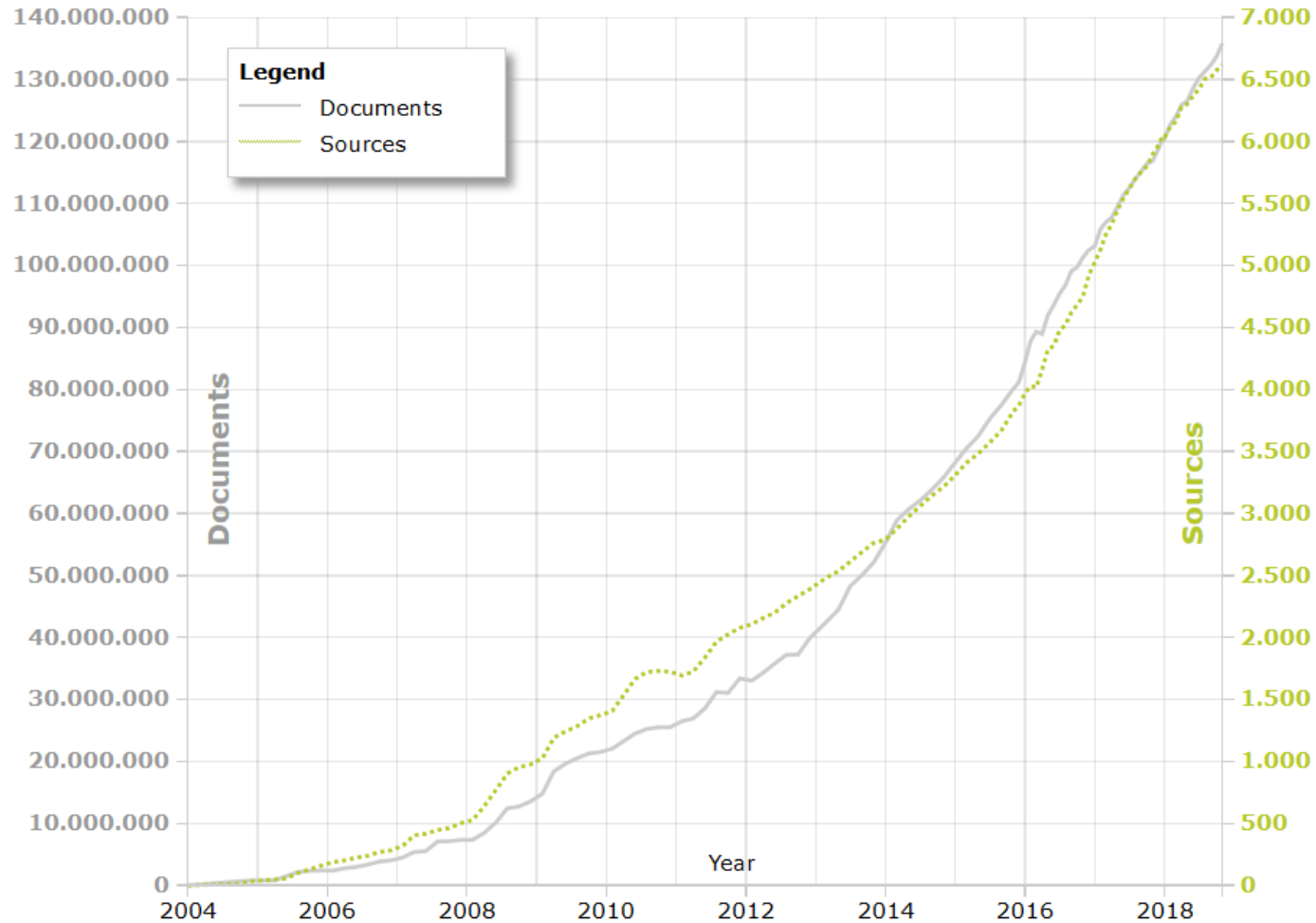


- 2001 Starting point as a search engine follow-up for a metasearch system
- 2004 Official Start (FAST Data Search)
- 2006 Starting participation in EU projects
- 2011 Switch to open source (Lucene/Solr, VuFind)
- 2012 OAI-PMH-Interface, data delivery of subject sections
- 2014 OA Boosting
- 2015 OA status and License information processing
- 2016 CrossRef Integration (OA publications)
- 2017 ORCID Integration, DOI Processing, VR Prototype
- 2018 Migration to Multinode-System

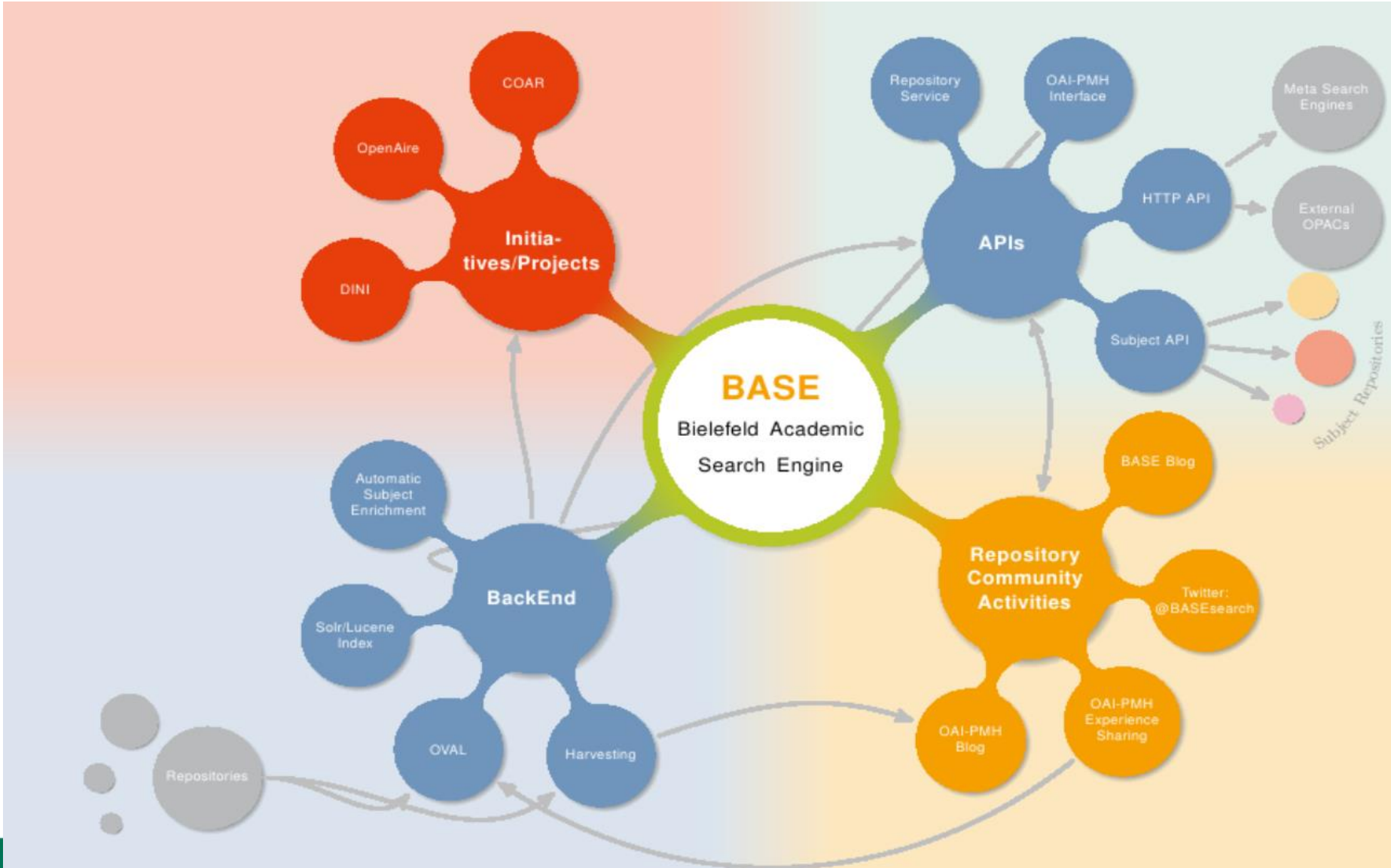
- 6642 Repositories included (mostly via OAI-PMH, some few via Crossref)
- From 124 Countries world-wide
- Ca. 136 Mill. Documents/Objects
- Ca. 70 % Open Accessible
- Focus on Repository Level
- Internal Scheme: extended Dublin Core Format
- Ca. 15.7 Mill. Documents enriched with DDC-Code (Dewey)

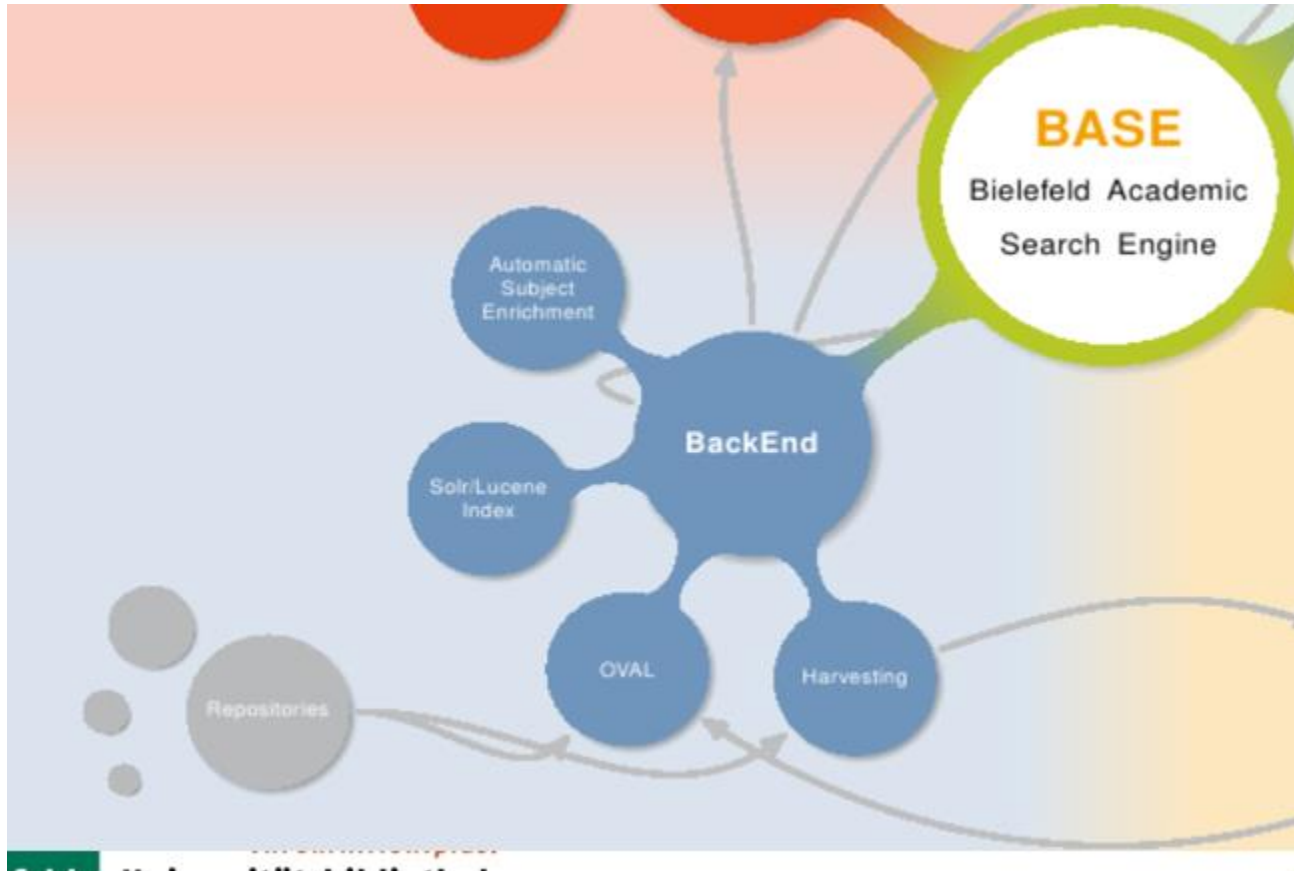
Statistics

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

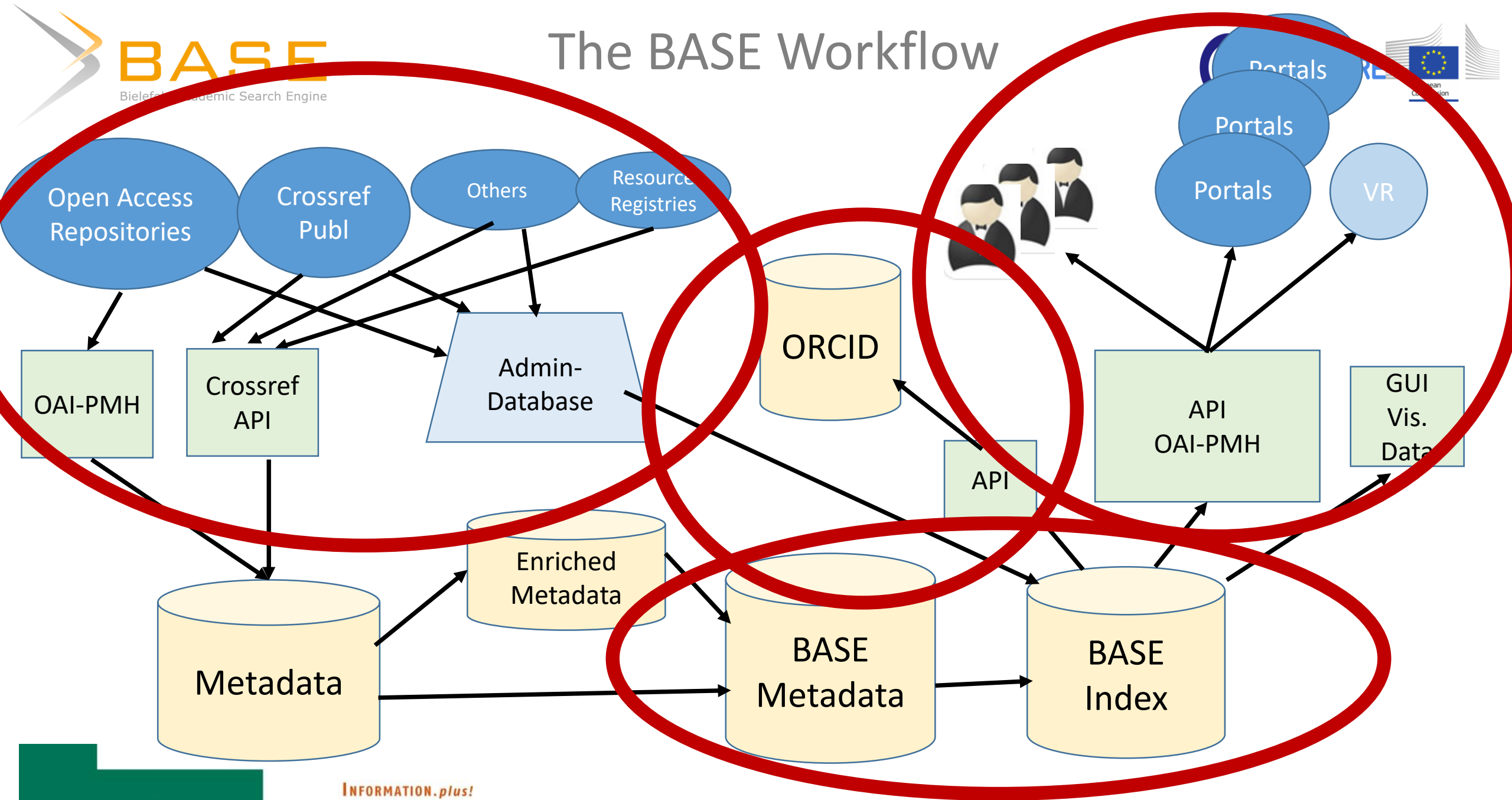


- Open Access Repositories world-wide
- Academic-valuable Contents
- Focus on Institutional Repositories
- Aggregators (RePEc)
- Subject Repositories (arXiv, CiteSeerX etc)
- Electronic Journals
- Digital Collections
- Data Repositories
- OA Publishers (via Crossref)





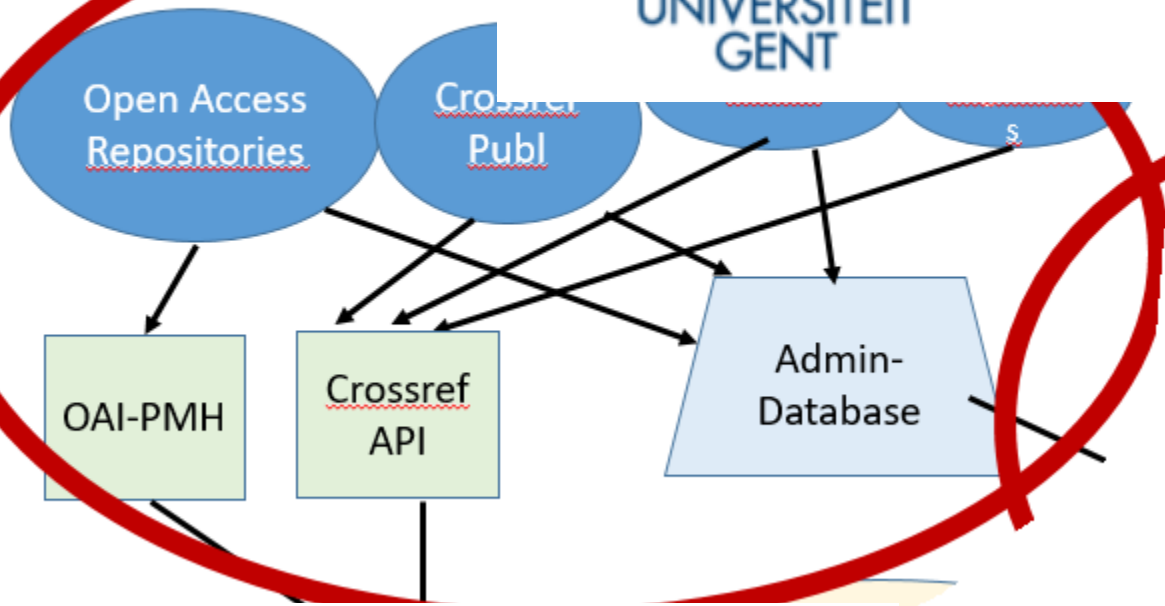
The BASE Backend



BASE Metadata Providers via OAI-PMH



Ghent University Academic Bibliography

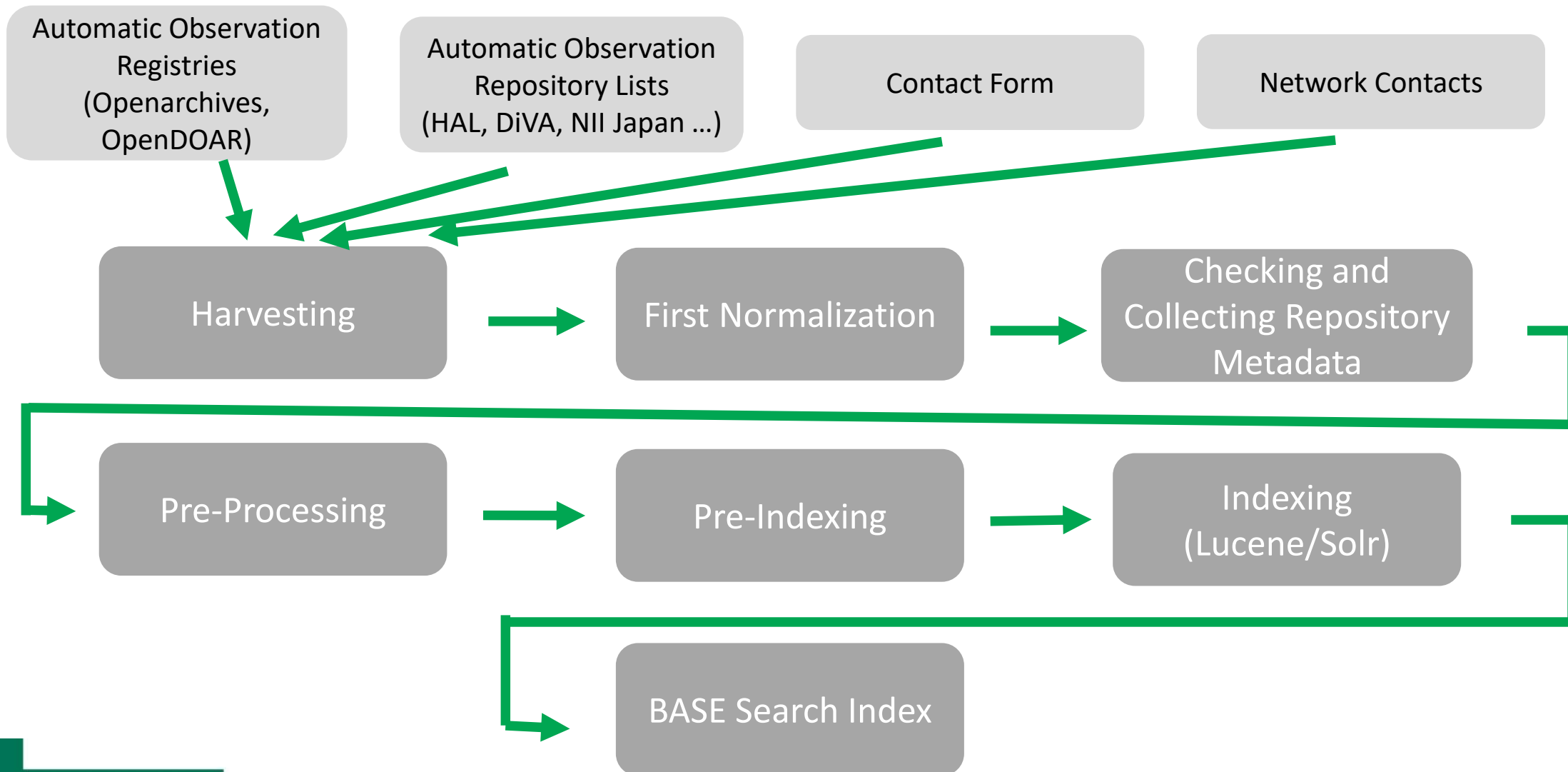


Cite
Seer
X
RePEc

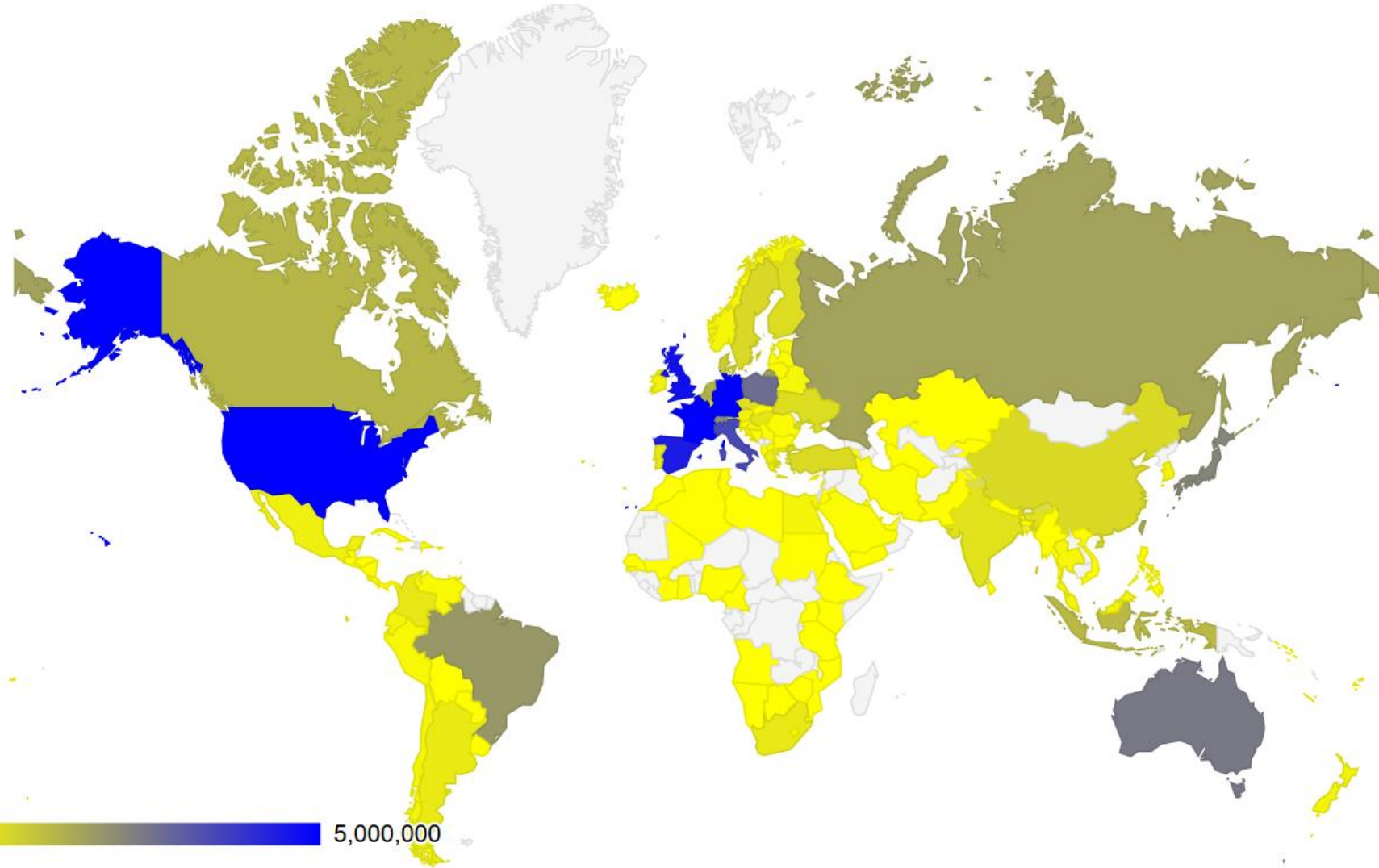


PubMed
PubMed comprises books. Citations ma



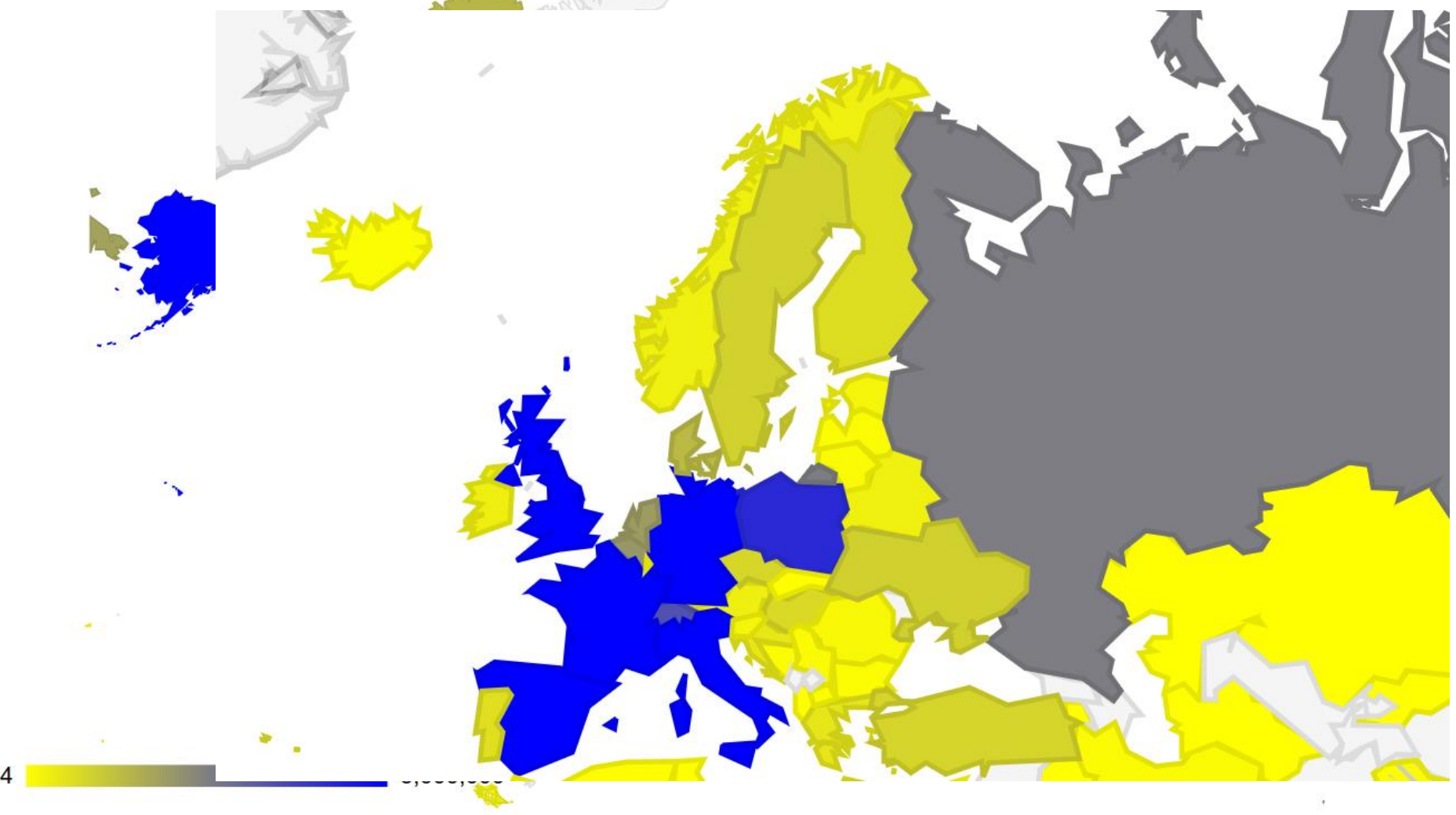


The Global Repository Network (documents per country)

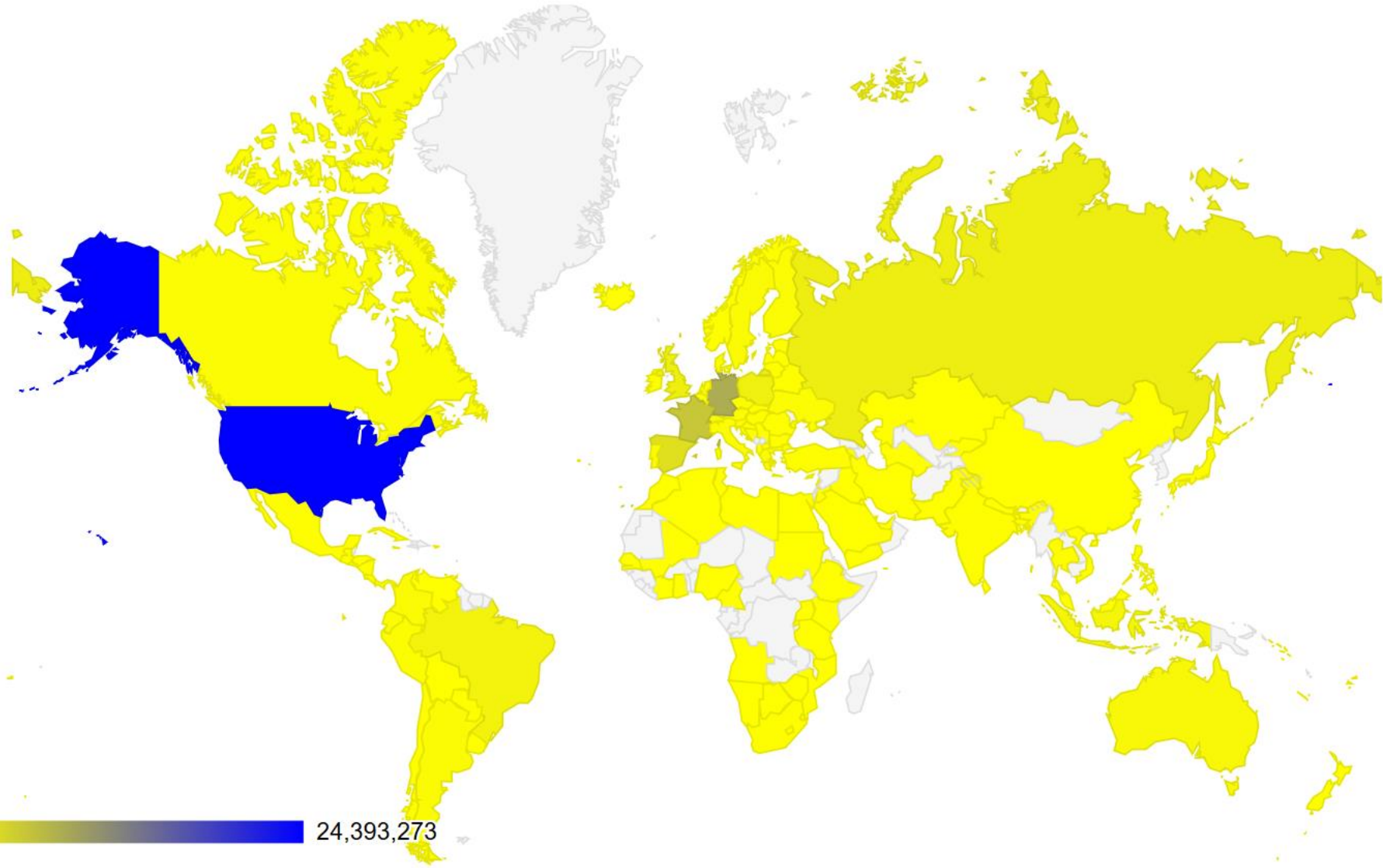


4 5,000,000

The Global Repository Network (documents per country)

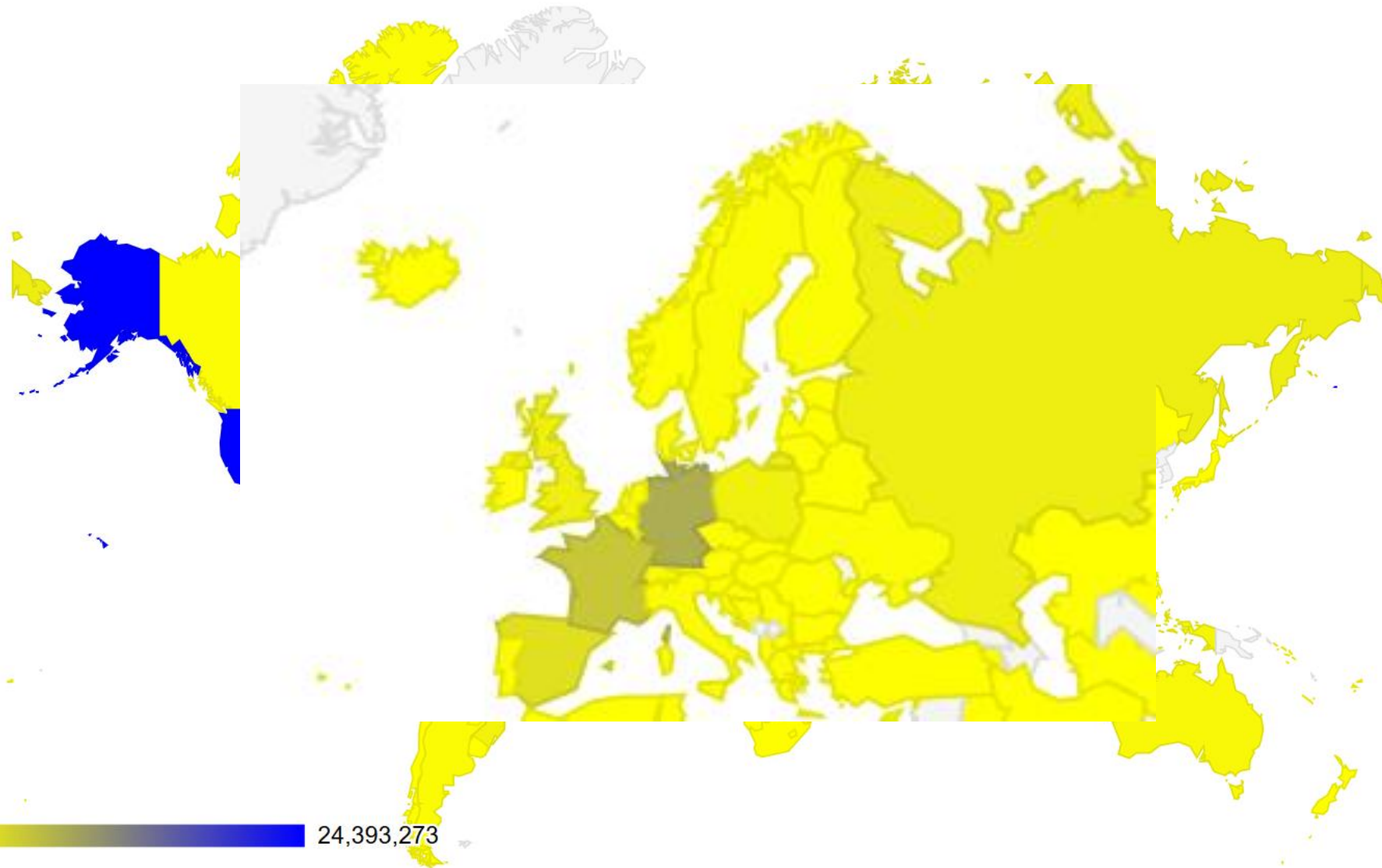


The Global Repository Network (OA documents per country)



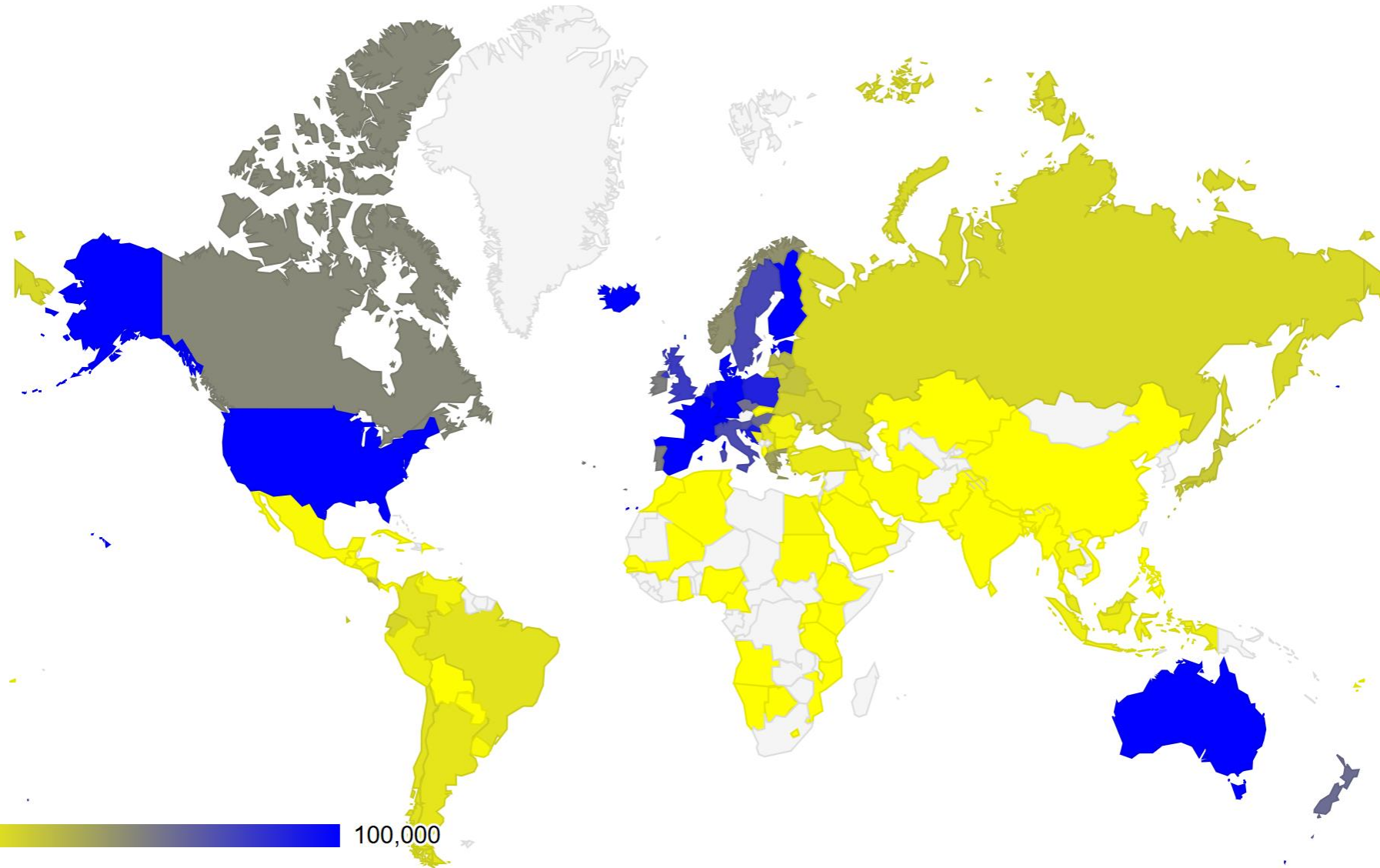
0 24,393,273

The Global Repository Network (OA documents per country)

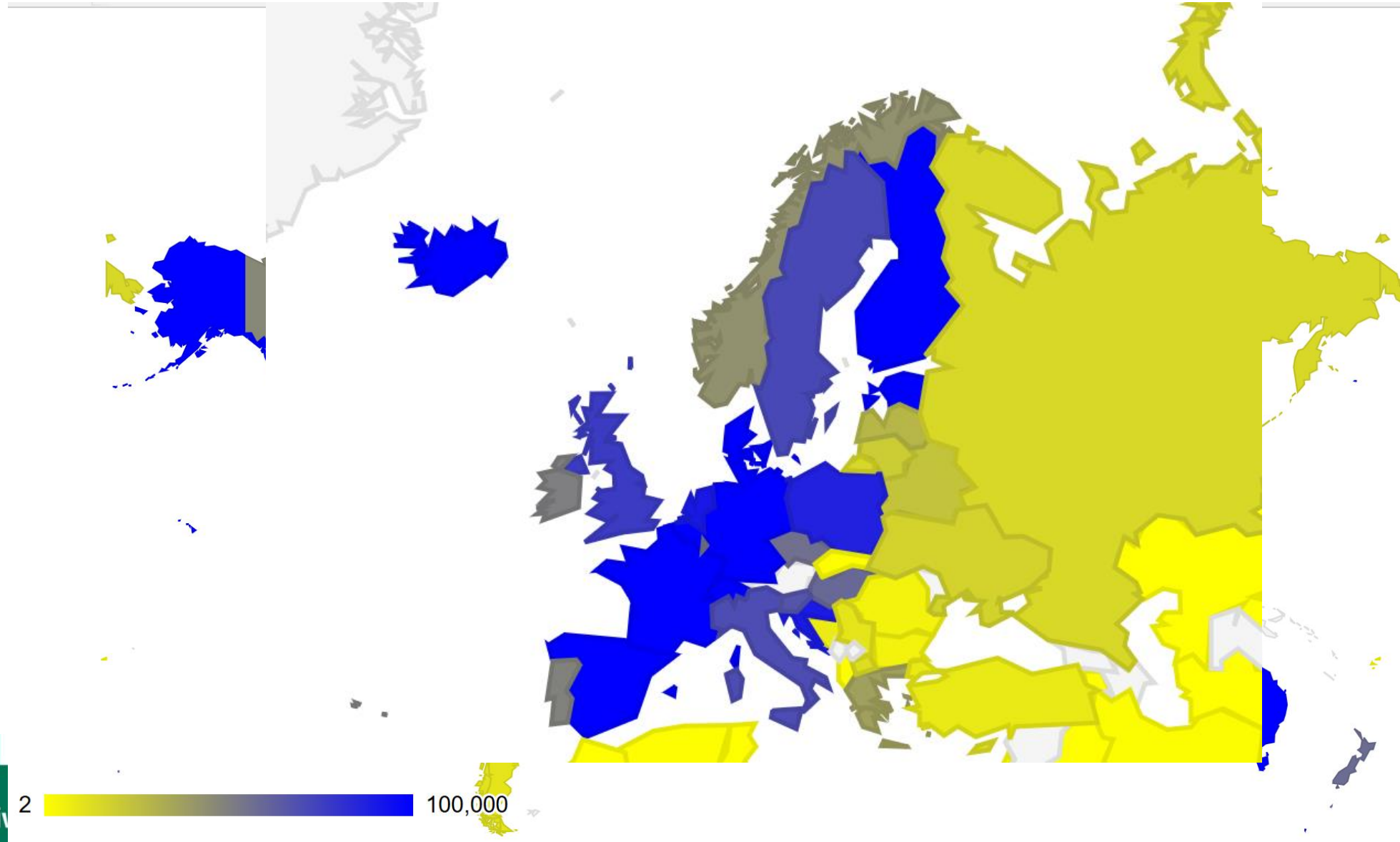


0 24,393,273

The Global Repository Network (documents per million people)



The Global Repository Network (documents per million people)

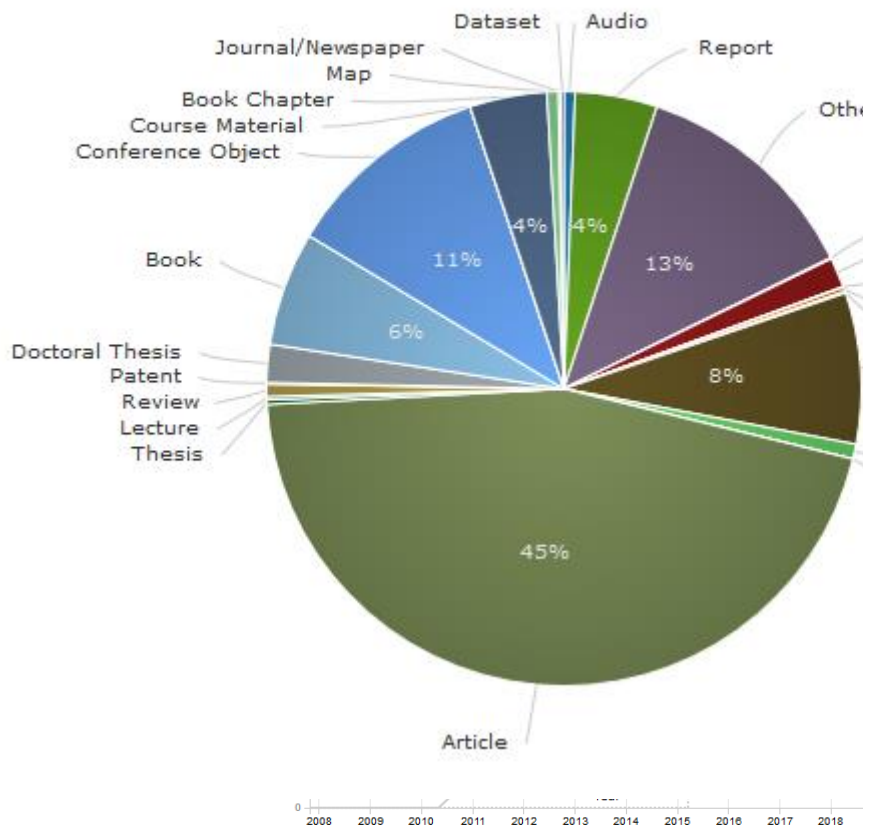


France
 10100424 Document
 7 % of Documents world-w

Nuxeo

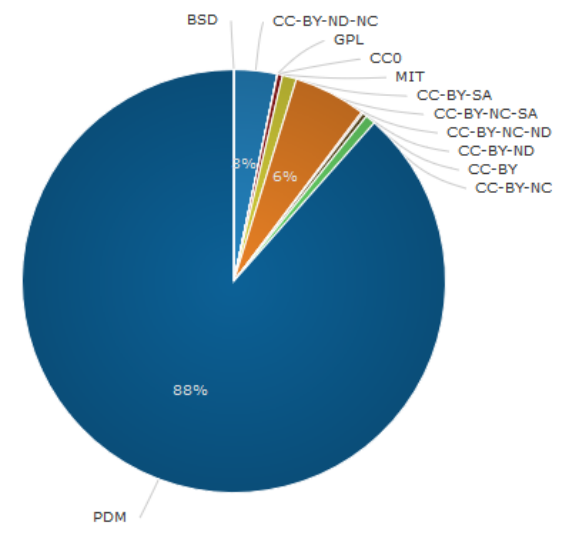
Publication Type Distribution

(number of documents)

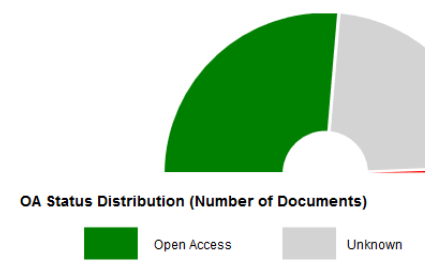
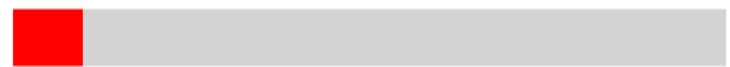


License Type Distribution

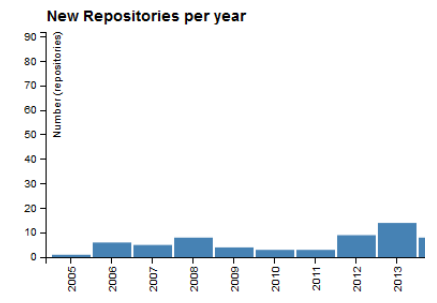
(number of documents)



Percentage Records with License Information = 12.31 %



Repository Development





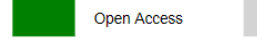
Germany

19568393 Documents from 393 Repositories

14 % of Do

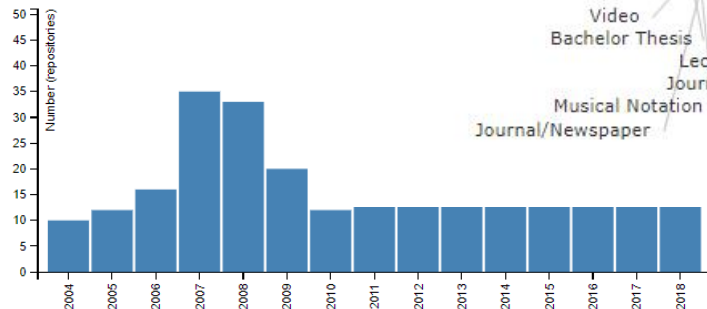


OA Status Distribution (Number of Doc)



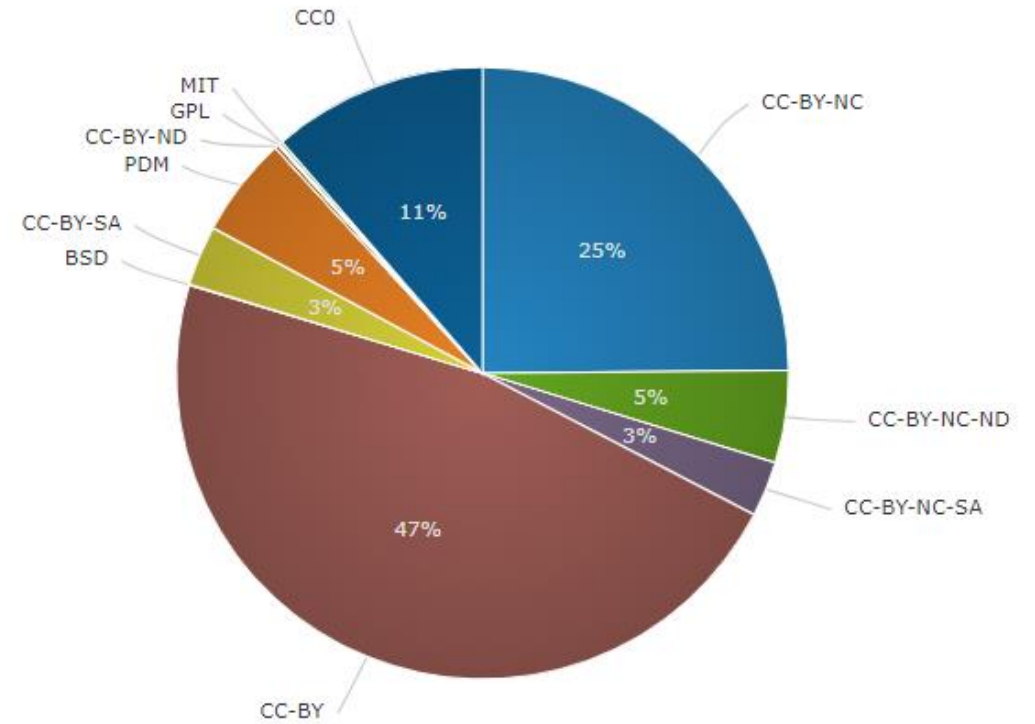
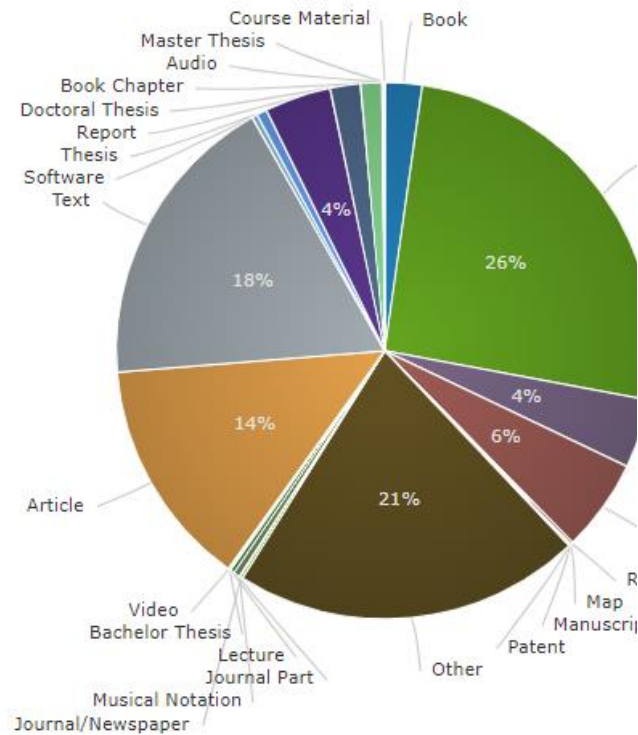
Repository Development

New Repositories per year

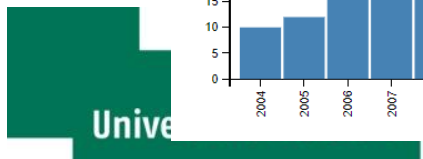
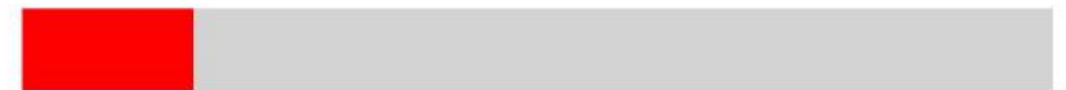


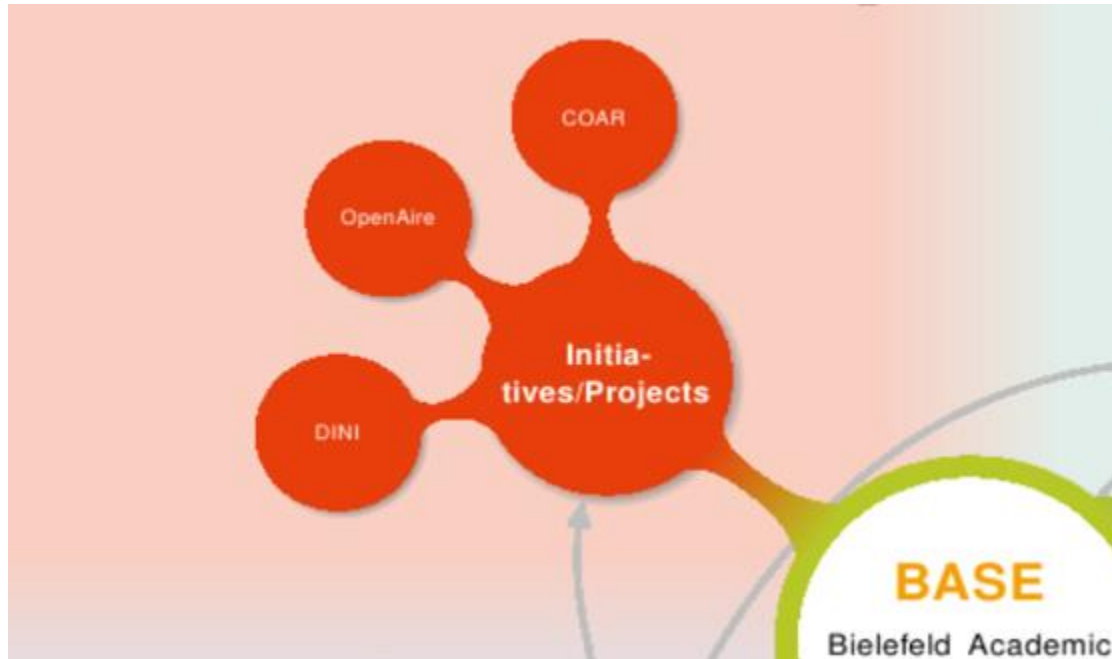
Publication Type Distributic

(number of documents)



Percentage Records with License Information = 20.84 %

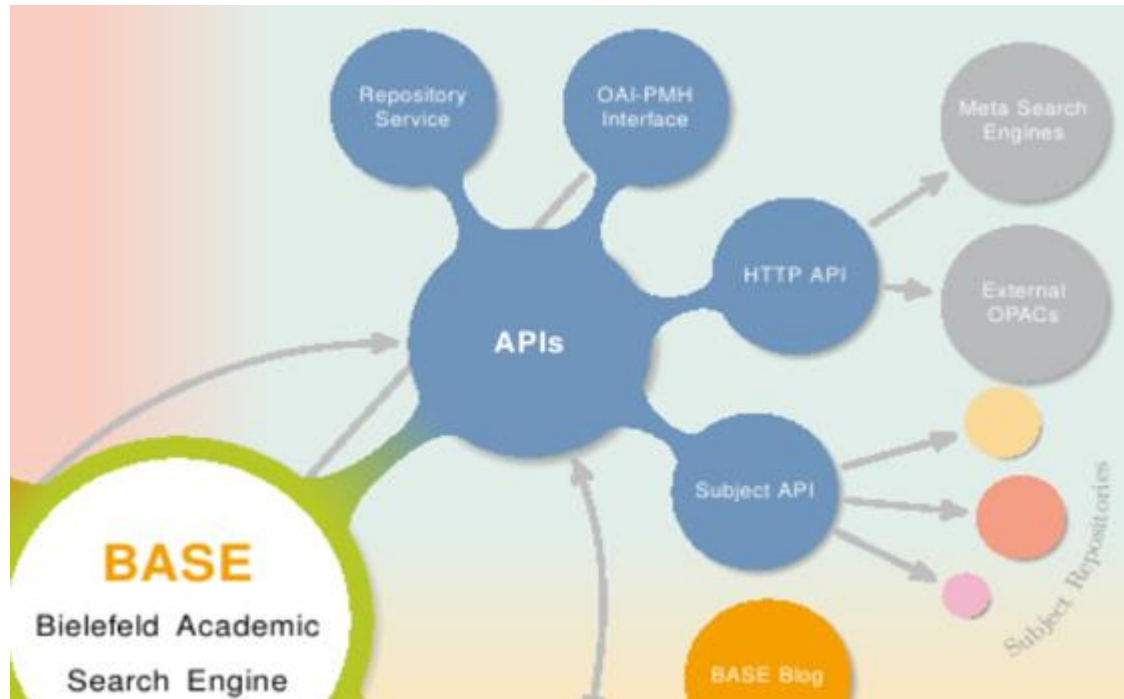




BASE Initiatives/Projects

BASE Network Integration (Projects, Partners, Communities)

- DRIVER/OpenAIRE
- DINI Epub Working Group (DINI-Zertifikat)
- COAR member
- Automatic Classification (DFG funded)
- Europeana
- ORCID-DE-Project (DFG funded)
- OpenAPC Project (DFG funded)
- Virtual-Reality-Solution for BASE Search
- Cooperation with National Science Library Beijing
- Cooperation with German National Library
- Local Working Group Bibliometrics
- National Office Open Access Transformation



The BASE APIs

- Search API (since 2006)
- OAI-PMH-API (since 2011)
- Data Dumps (since 2016)
- Repository Metadata API (in preparation)

Ihre Online Publikation verö

Elektronische Bibliothek

E-LIB

ECONBIZ

Find Economic Literature.

DEUTSCHE NATIONAL BIBLIOTHEK

Par dissemin











DOAI












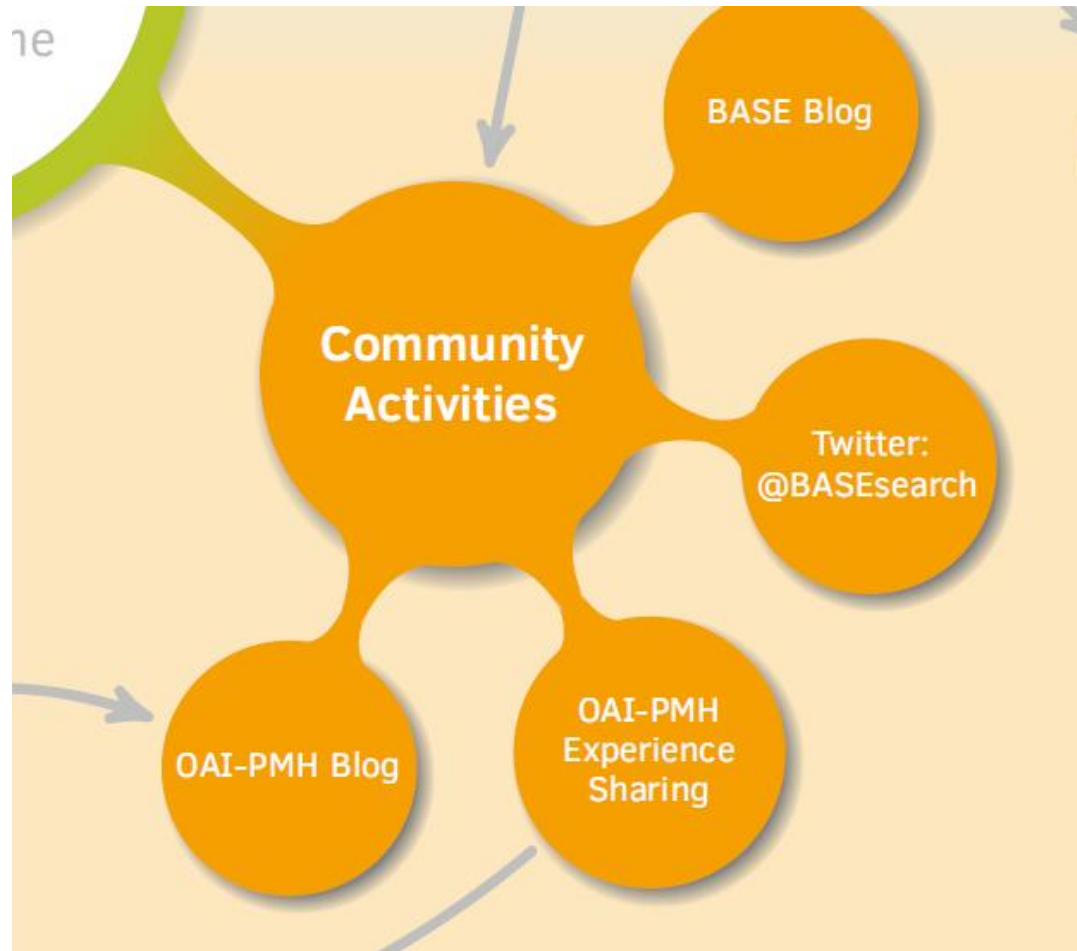
couperin.org
Consortium Unifié des Établissements Universitaires et de Recherche pour l'Accès aux Publications Numériques

s
and relies on the meta



Länder (Top 10) - Gesamte Liste					
	Länder		Seiten	Zugriffe	Bytes
	Germany	de	14.687.025	14.687.025	429.37 GB
	France	fr	1.134.189	1.134.189	4.64 GB
	Great Britain	gb	689.095	689.095	12.37 GB
	China	cn	76.109	76.109	503.42 MB
	United States	us	4.194	4.194	19.07 MB
	Switzerland	ch	1.774	1.774	210.17 KB
	Brazil	br	1.283	1.283	5.90 MB
	Russian Federation	ru	1.214	1.214	5.59 MB
	Unbekannt	unknown	1.107	1.107	17.72 MB
	Turkey	tr	304	304	1.26 MB
	Sonstige		5145	5145	72.67 MB

Länder (Top 10) - Gesamte Liste					
	Länder		Seiten	Zugriffe	Bytes
	Germany	de	110.975	522.751	17.73 GB
	United States	us	31.698	111.455	4.76 GB
	France	fr	30.304	98.637	3.54 GB
	China	cn	27.476	130.475	3.60 GB
	Bulgaria	bg	25.028	26.101	2.16 GB
	Russian Federation	ru	24.561	36.893	4.63 GB
?	Unbekannt	unknown	21.981	64.291	3.79 GB
	Poland	pl	16.482	58.900	2.63 GB
	Ukraine	ua	12.347	23.859	2.24 GB
	Great Britain	gb	12.097	48.317	1.90 GB
	Sonstige		204123	737383	34.72 GB



BASE Repository Community Activities (Blogs, Twitter, Emails)

Golden rules for repository managers

We are indexing all kinds of academically relevant resources - journals, institutional repositories, digital collections etc. - which provide an OAI interface and OAI-PMH for providing their contents (learn more about OAI at the [Open Archives Initiative](#) or [Wikipedia](#)). In case your source does not provide an OAI interface, you can upload your documents to aggregators like [DataCite](#) or [Zenodo](#), to subject repositories like [RePEC](#) or add your open access journal to [DOAJ](#). We are indexing these sources regularly.

However, the best way to get your documents indexed by BASE is to provide an OAI interface. We have compiled some *golden rules* that might be helpful to optimize your OAI interface. If your OAI interface complies with these rules, we can assure fast and smooth indexing of your source. Data from your source will be presented completely and in the best possible way.

You can check some of the following items using our [OAI-PMH validator OVAL](#).

• **OAI interface working**

💡 Your OAI interface is stable and responds to requests. *ListRecords* delivers results without timeout or other issues, e.g. an XML error.

⚠️ Otherwise, it is not possible to index your source.

• **Marking modified / deleted records**

💡 Any subsequent change to a record must be marked in your OAI interface by updating the document's datestamp. If a document is deleted from your source, the record in the OAI interface must be marked as "deleted". Under no circumstances may the record be completely deleted from the OAI interface.

⚠️ All indexed sources are regularly updated in BASE. If the "datestamp" is not updated or if a document is not marked as "deleted" (but instead completely removed from the OAI interface) an update in the BASE index is not possible and the document remains unchanged and therefore incorrect in the index.

• **Comprehensive metadata**

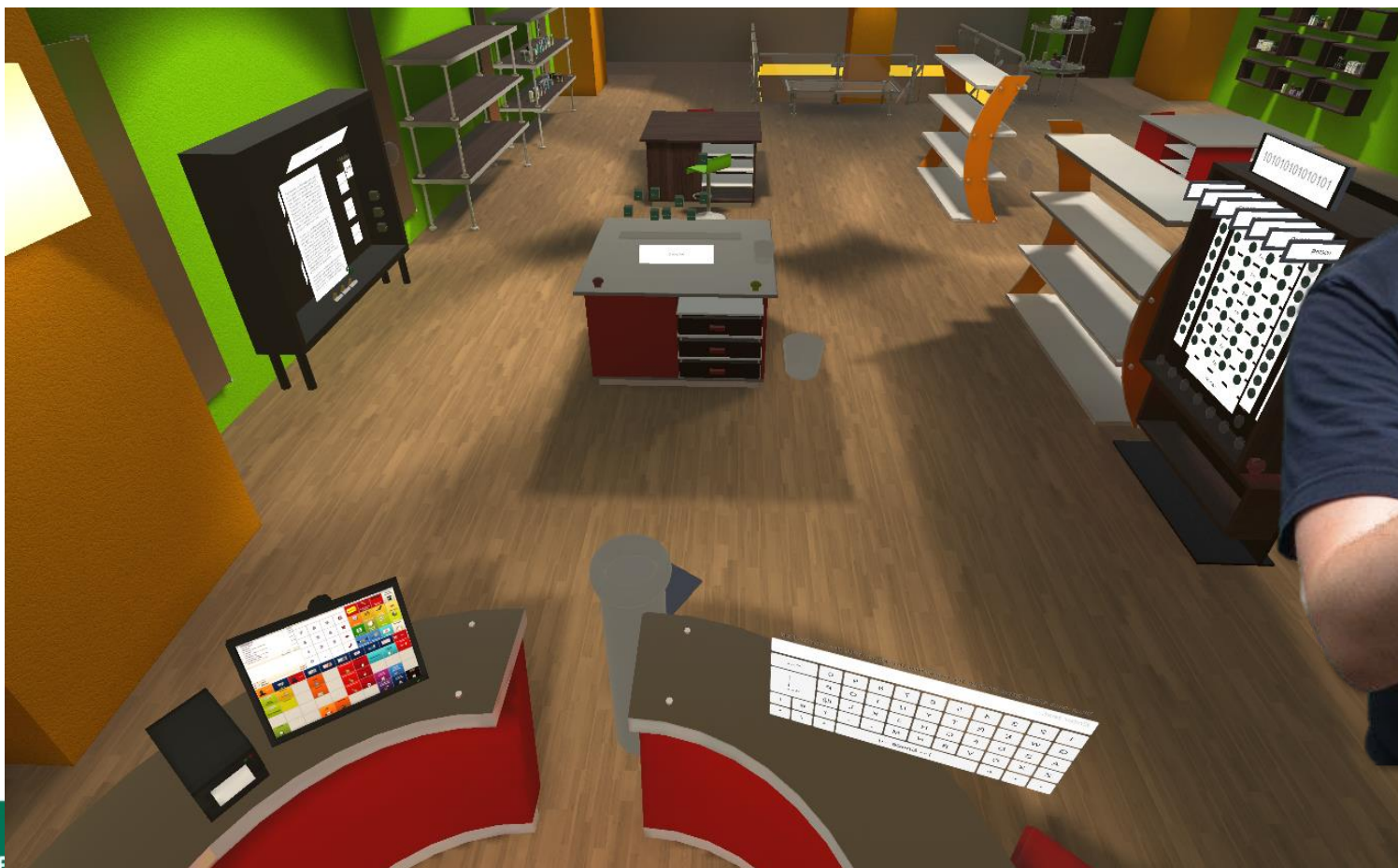
💡 Each item exposed via your OAI interface provides metadata as comprehensive as possible (title, author, abstract, publication date) using the [info-eu-r vocabulary](#).

⚠️ If important metadata is missing, documents from your source will be difficult to find in BASE. Using the info-eu-repo vocabulary makes sure that we can process and display hits from your source in the best way.

- Update Technical Infrastructure
(Multinode System in test)
- OA-Status, OA-Boosting, License information
- ORCID Integration (claiming)
- Crossref-OA-Daten
- Research Data
- Open Educational Resources
- VR Prototype
- Open Science Monitor

A Virtual Reality BASE Client

(Cooperative Development of a Client With Bielefeld University,
Cluster of Excellence Cognitive Interaction Technology)





www.openaire.eu

Project-Milestones

OpenAIRE

12/2009 – 11/2012

- Implementing the Open Access Pilot in FP7 (SC39)
- Setup of National Open Access Desks

OpenAIREplus

12/2011 – 12/ 2014

- Merge of DRIVER with OpenAIRE infrastructure
- Linking of publications with research data (subject-specific studies)

OpenAIRE2020

01/2015 – 06/ 2018

- OA-Mandat / Open Data Pilot in H2020
- FP7 Post Grant Pilot

OpenAIRE Advance

01/2018 - 12/2020

- consolidating current infrastructure & services (EOSC compatibility)
- Capacity building activities (Open Science policies, RDM, TDM)
- Towards a scholarly commons

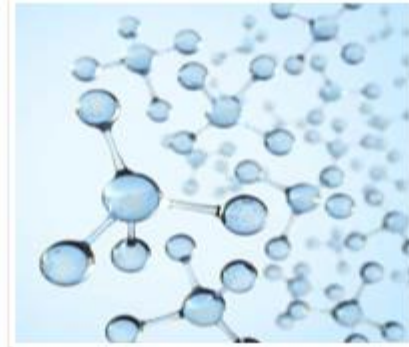
OpenAIRE is supporting:

OS policies



- Harmonization for policy makers
- Training
- Support

Infrastructure



- Interoperability
- Setup
- Connectivity
- Repositories

Open Research Data



- FAIR
- Open data
- Tools
- Legal
- Compliance

OA to publications



- Guides
- Tools/repositories
- Licenses
- Compliance

15+ Services

From basic infrastructure level
to value added



From Basic Infrastructure to Value Added Services



Search & Share



Publish OA



OpenAIRE | MONITOR



OA Broker



Validator



Research Analytics



Institution Monitoring Dashboard



OpenAIRE | CONNECT



Helpdesk



Usage Analytics



Funder Monitoring Dashboard



Open Science Observatory



OpenAIRE | EXPLORE



APIs



Interoperability Guidelines



easyDMP



Universitätsbibliothek



OpenAIRE | PROVIDE



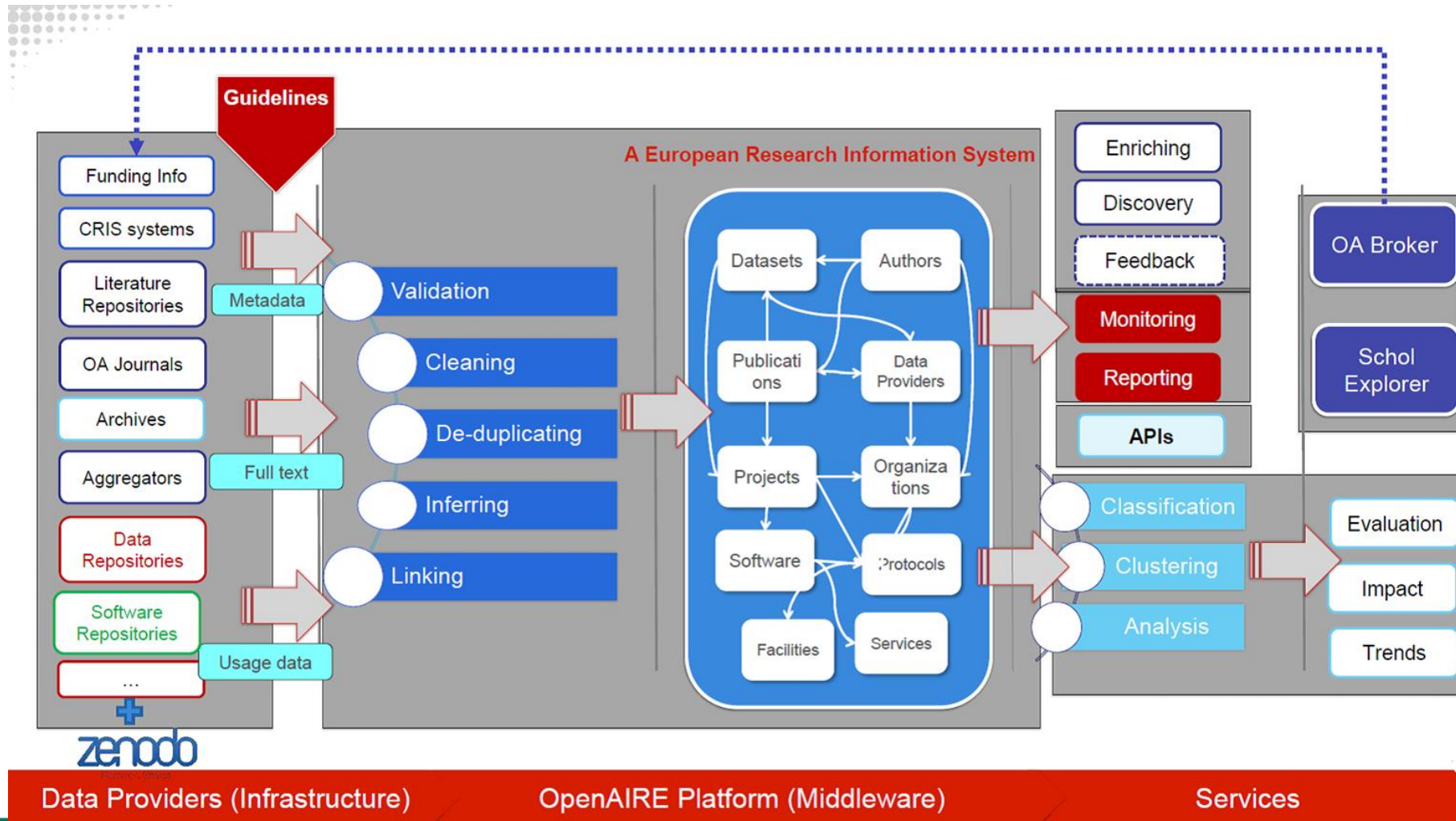
Zenodo



Training



OpenAIRE | DEVELOP



Data Providers (Infrastructure)

OpenAIRE Platform (Middleware)

Services

INFORMATION.plus!

OpenAIRE Information Space in Numbers



SEARCH SHARE LINK CONTENT PROVIDERS

Our community

21,411,573 PUBLICATIONS

609,083 RESEARCH DATA

13 K
CONTENT PROVIDERS

17
FUNDERS

2,544,493
PROJECTS

Our data policies

DATA ACQUISITION

How we retrieve data, how often, what processes it goes through (aggregating, cleaning, transforming, inferring, de-duplicating), what are the quality checks along all data processing stages.

DATA USAGE

Who is able to retrieve our data, what are the licences, what about specific SLAs.

[Learn more](#)

Content provider?

Join OpenAIRE, use our tools and make your content more visible around the world.

[OpenAIRE.PROVIDE](#)



Validate

Interoperable metadata is key for effective content sharing.



Register

Reach a wider audience around the world



Enrich

Improve your metadata. Get more connections



Measure

Open research impact empowers Open Science



Interoperable metadata is key for effective content sharing

Use our validation service and see how you can apply the OpenAIRE Guidelines to expose your contents using global standards.

VALIDATE

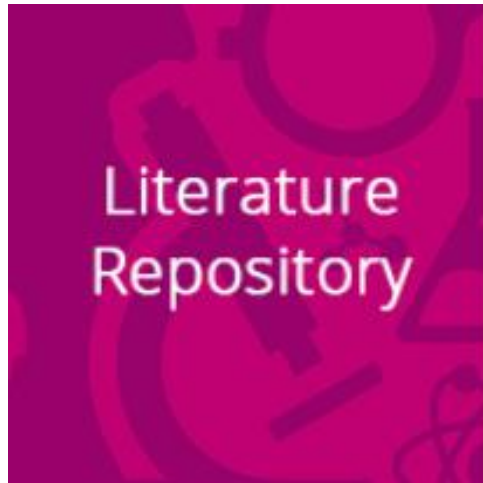




Reach a wider audience around the world

Register your datasource in OpenAIRE and be part of a global interlinked network.

REGISTER





ENRICH

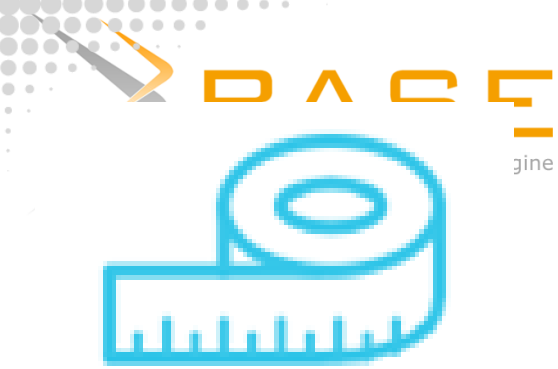
Improve your metadata. Get more connections

OA Broker service offers a wealth of information on scholarly communication data.

Find out what interests you and subscribe to enrich your records.

More & Missing events that may enrich your Repository:

- Persistent identifiers
- Open Access Versions
- Projects
- Subjects
- Abstracts



Open research impact empowers Open Science



MEASURE

Open Metrics service by sharing your usage data.
Get the benefit of an aggregated environment to broaden the mechanisms for impact assessment.

Get usage statistics reports for your datasource



Article Report 1, number of successful article download requests by month and repository.



Item Report 1, number of successful item download requests by month and repository.



Repository Report 1, number of successful item downloads for all repositories participating in the usage statistics service.



Book Report 1, number of successful title requests by month and title.



Book Report 2, number of successful section requests by month and title.

Interoperability Guidelines

Evolution of the OpenAIRE-Guidelines

2010
Literature
Guidelines v1

2013
Literature
Guidelines
v3

2015
CRIS-CERIF
Guidelines v1

2012
- Literature
Guidelines v2
- Data
Guidelines v1

2014
Data
Guidelines
v2

2018
new releases
of Content
Provider
Guidelines

BASE FROM Guidelines for Data Providers TO Guidelines for Open Science Content Providers



RESEARCH LITERATURE

- Thematic Repositories
- Institutional Repositories
- E-journals

RESEARCH SOFTWARE

- Software Repositories
- Catch-all Repositories



RESEARCH DATA

- Data Repositories
- Catch-all Repositories
- Institutional & thematic repositories

OTHER RESEARCH PRODUCTS

- Catch-all Repositories

RESEARCH INFORMATION

CRIS



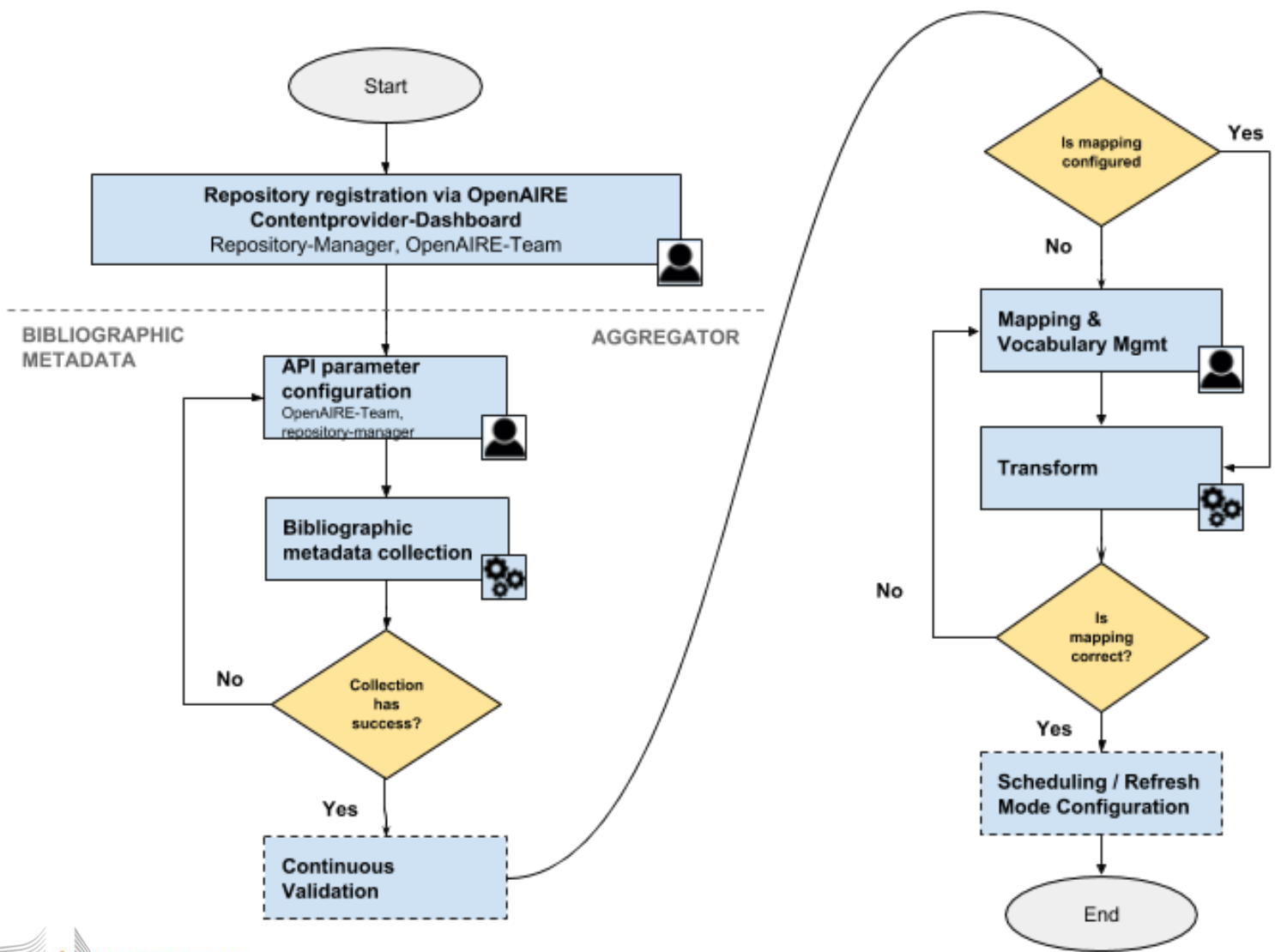
<https://guidelines.openaire.eu>

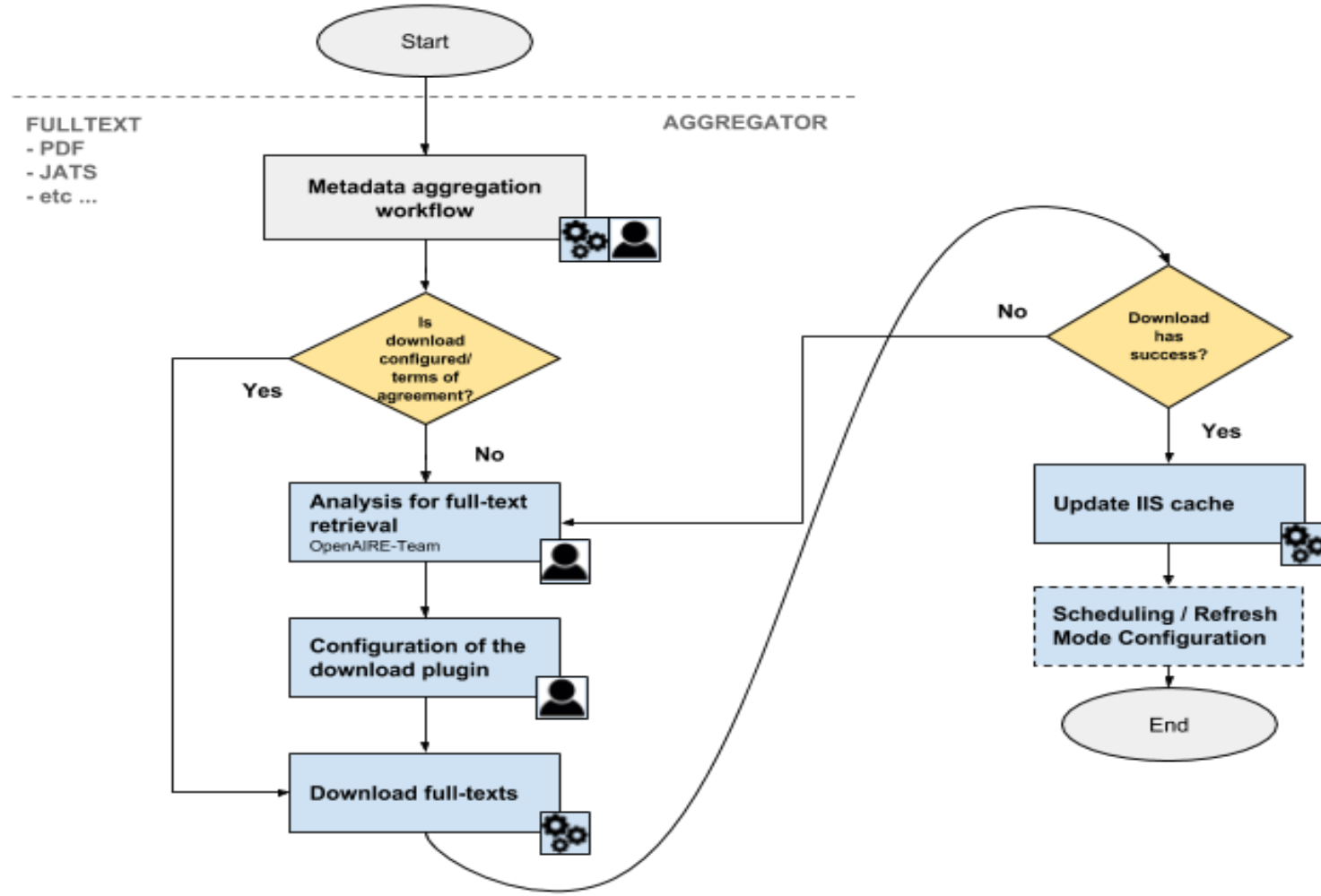
Guidelines for Literature Repository Managers

- **Funder perspective**
 - Link funding information with research results
- **Author and Reader perspective**
 - Link authors and contributors with their research output and ease name disambiguation
- **Service provider perspective**
 - Avoid overloading of oai_dc metadata
 - Make maintenance and mappings of controlled vocabularies easier by help of identifiers; follow linked data principles if possible
 - Make identification of resources easier (e.g. for TDM)
 - Improve alignment with other regional repository networks
 - **Agree on a shared set of metadata properties and controlled vocabularies**
 - Examples: LA Referencia, JAIRO



- to improve interoperability of metadata information exchange
- to support FAIR Data Principles
- by defining application profiles to describe different kinds of research products
- by re-using existing standards (Dublin Core, DataCite)
- by re-using vocabularies and extending them when necessary
 - e.g. for PID types





BASE

started as local initiative, not 3rd party funded, bringing up expertise for networking, door opener for cooperative projects with funding of additional services

OpenAIRE

EC funded project since 2009, firstly aiming to implement the European Commission Open Access policy (SC39 in FP7) then the Open Access mandate (since Horizon2020) and now one of the EC key research infrastructures to implement the European Open Science Cloud

- * quite large consortium (48 partners across Europe)
- * forming a legal entity, officially launched in Sept-2018

Technical Infrastructure

BASE: small, efficient, stable, flexible, pragmatic, weekly re-indexing, small set of index-focused services as linguistic tools, OA boosting, ORCID claiming

OpenAIRE: broad and complex infrastructure of services, from metadata aggregation, enrichment, deduplication, indexing, monitoring (number of (OA) publications per project/funder), end-user feedback (claiming links between research results and projects/funder; links between research literature and data); value added services: broker service to notify content providers about suggested metadata additions and enrichments in their repositories; usage statistics service; participation in the ScholeXplorer - data-literature interlinking service
(<https://diservice.research-infrastructures.eu/#/>)

Content Scope:

BASE: Global, all kind of science-related material including digital collections, archives contents, research data, open educational resources

OpenAIRE: all kind of science-related material (literature, research data, software, "other research products") in Europe and beyond; EU-funded context and other national funders in Europe and beyond; contextualization by interlinking research works with projects, persons, organisations (CRIS like)

User Scope

BASE: end-users and API partners, specific tools for repository managers (validator, rules)

OpenAIRE: end-users and API partners; tool for (European) project coordinators to generate publication lists per project and dissemination to the EC participant portal; European scientists esp. in the context of research communities (OpenAIRE-Connect project), extending the dissemination services

Network and Interoperability

BASE: Many national and international partners, specific communication channels with repository managers, participating in COAR and DINI, regional and local embeddin

OpenAire: European research infrastructure, NOAD infrastructure, participation of LA Referencia and COAR in the project consortium to liaise with the global repository community

- dashboards that target specific stakeholder groups
- content provider - <https://provide.openaire.eu>
- funding organizations, project coordinators, institutions - <https://monitor.openaire.eu/> *
- research communities - <https://connect.openaire.eu/>
- API users and developers - <https://develop.openaire.eu/>

BASE <-> OpenAire

Cooperative Approach – different working groups with overlapping of members

Exchange of information related to specific resources (detecting new ones, technical specifics, error situations, change management, communication with repository managers),

Exchange of expertise on repository network developments

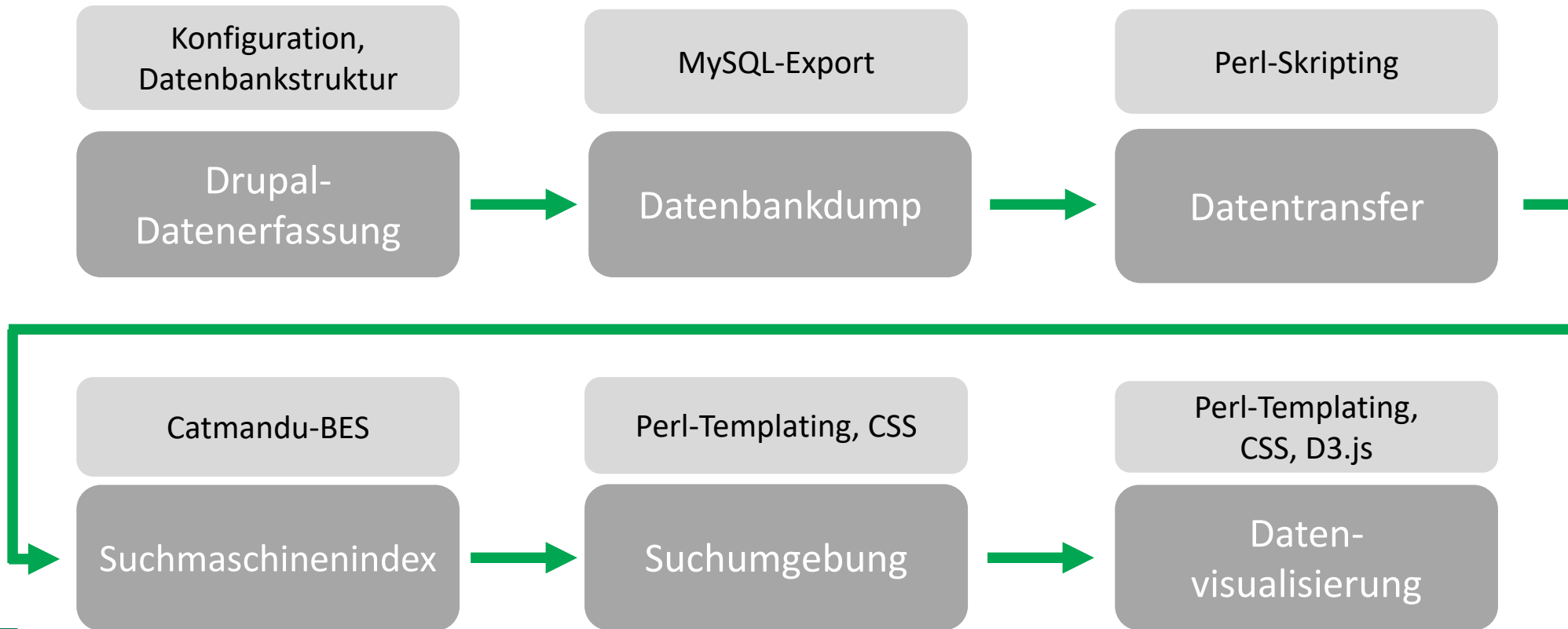
Thank You for Your Attention!

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

<https://www.openaire.eu/>

friedrich.summann@uni-bielefeld.de

Datenfluss vom n der Erfassung zur Präsentation



Vergleich BASE- OpenAire

BASE – OpenAIRE

BASE local initiative, not 3rd party funded, bringing up expertise for networking, door opener for cooperative projects with funding

OpenAIRE EC funded project since 2009, firstly aiming to implement the European Commission Open Access policy (SC39 in FP7) then the Open Access mandate (since Horizon2020) and now one of the EC key research infrastructures to implement the European Open Science Cloud

* quite large consortium (48 partners across Europe)

* forming a legal entity, officially launched in Sept-2018

Technical Infrastructure

INFORMATION .plus!

Universitätsbibliothek

BASE: small, efficient, relativ, weekly re-indexiung, small set of

Neue Materialien (beyond Journal Articles)
Digital Collections (Scientific Research Material, I
Research Data
OER

<option value='f_dcoa:"2"'>(72 902 529) Inconnu</option> <option value='f_dcoa:"1"'>(59 235 632)
Accès libre (Open Access)</option> <option value='f_dcoa:"0"'>(4 142 088) Non-Open Access</option>



World-Wiede

<option value='f_dcrightsnorm:"PDM"'>(6 187 163) PDM</option> <option value='f_dcrightsnorm:"CC-BY"'>(5 410 176) CC-BY</option> <option value='f_dcrightsnorm:"CC-BY-NC-ND"'>(2 643 913) CC

Verteilung Publikationstypen

Embedding

Lokal (OPAC, VR,)

Regional

National

International

Branding of the Repository

Aggregation of Research Results by OpenAIRE

