Making Research Data Repositories Visible – The re3data.org Registry

Frank Scholze | Karlsruhe Institute of Technology, KIT Library

Heinz Pampel | GFZ German Research Centre for Geosciences, LIS

Paul Vierkant | Humboldt-Universität zu Berlin, BLIS

Roland Bertelmann | GFZ German Research Centre for Geosciences, LIS

IMCW 2014 | Antalya, 25.11.2014

Background

- Funders' data policies

  - Example: European Commission

  29.3   Open access to research data

  [OPTION for actions participating in the open Research Data Pilot: Regarding the digital research data generated in the action ('data'), the beneficiaries must:

  (a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

  (i)   the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

  (ii)   other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);

  (b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

European Commisson. (2014). Horizon 2020 Annotated Model Grant Agreements. Version 1.6.2 .Retrieved from
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

Background

Data Availability

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available, without restriction and from the time of publication, with only rare exceptions to address legal and ethical concerns (see the PLOS Data Policy and FAQ for further details). When submitting a manuscript, authors must provide a Data Availability Statement that describes where the data underlying their manuscript can be found.

Your answers to the following constitute your statement about data availability and will be included with the article in the event of publication. **Please note that simply stating 'data available on request from the author' is not acceptable. If, however, your data are only available upon request from the author(s), you must answer "No" to the first question below, and explain your exceptional situation in the text box provided.**

Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?

Answer Required: | Please select a response | Please select a response.
Please select a response
Yes – all data are fully available without restriction
No – some restrictions will apply

- Journal Data Policies

  - Nature Publishing Group
    - "[...] authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications."

  - PLOS
    - "PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception."

NPG (2013). Availability of data and materials. Retrieved from http://www.nature.com/authors/policies/availability.html
PLOS (2014). PLOS Editorial and Publishing Policies. Retrieved from http://www.plosone.org/static/policies.action

Reproducability and trust

Fig. 1.
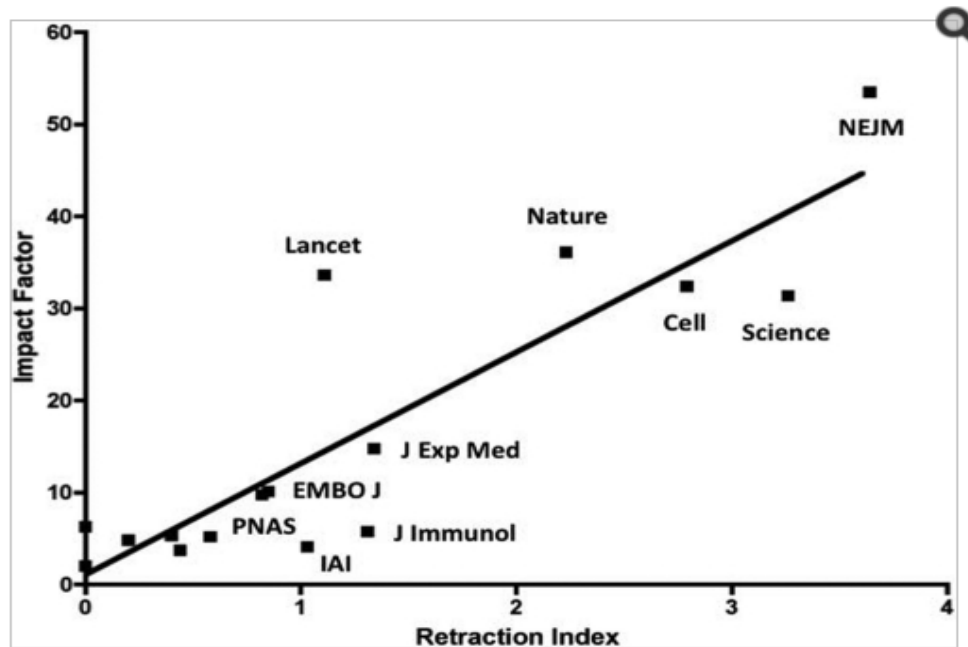
Correlation between impact factor and retraction index. The 2010 journal impact factor (37) is plotted against the retraction index as a measure of the frequency of retracted articles from 2001 to 2010 (see text for details). Journals analyzed were *Cell, EMBO Journal, FEMS Microbiology Letters, Infection and Immunity, Journal of Bacteriology, Journal of Biological Chemistry, Journal of Experimental Medicine, Journal of Immunology, Journal of Infectious Diseases, Journal of Virology, Lancet, Microbial Pathogenesis, Molecular Microbiology, Nature, New England Journal of Medicine, PNAS,* and *Science.*

re3data - Mission

- **global registry** of research data repositories

- covers **all** academic **disciplines**

- helps researchers, funding bodies, publishers, libraries and scholarly institutions to **find** research data repositories

- promotes a culture of **sharing**, increased **access** and better **visibility** of research data
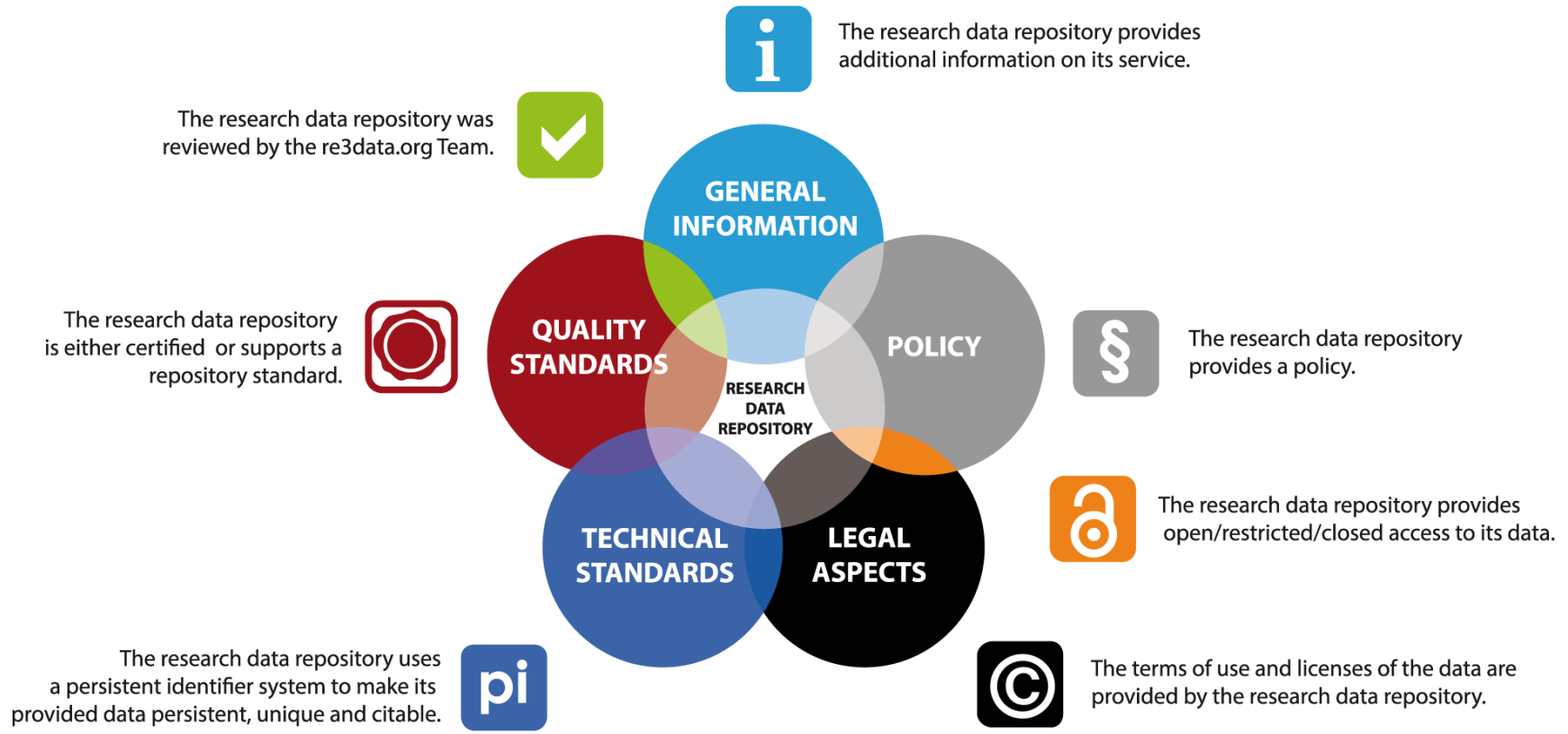
Schema

Properties **37**

Based on Analyses, Feedback and Experience
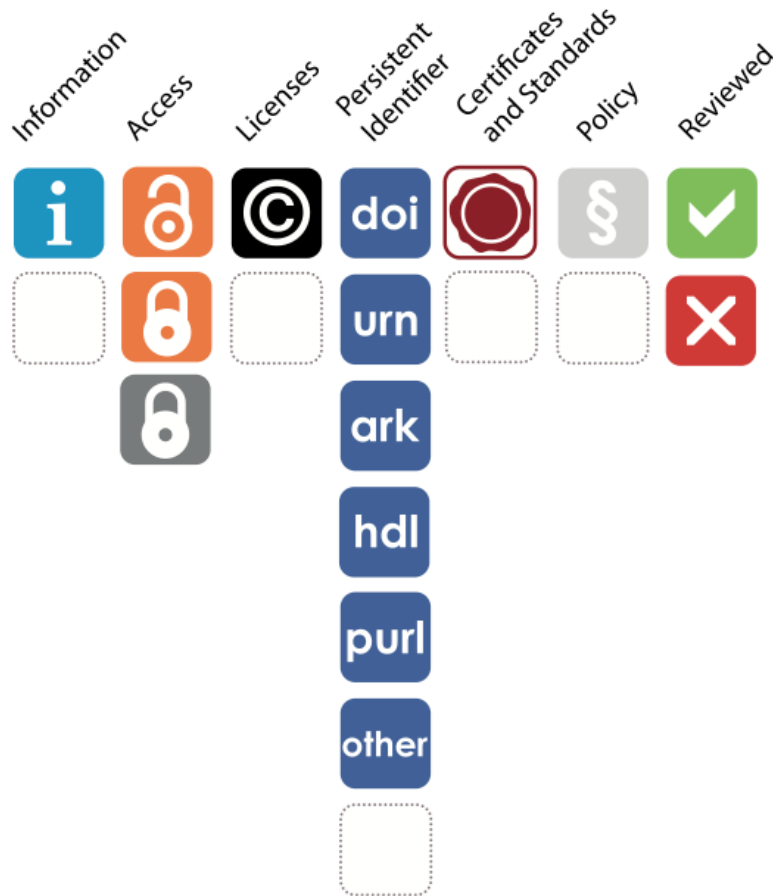
Version **2.1**

Icons

- Information icons help researchers to easily identify an adequate repository for the storage and reuse of their data.
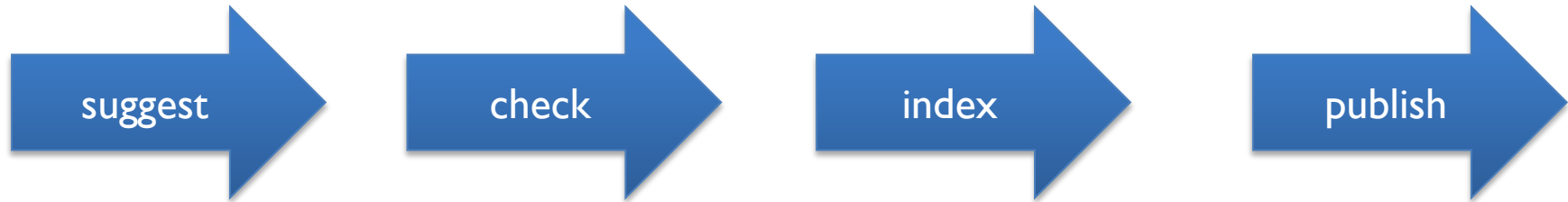

The research data repository provides additional information on its service.

The research data repository was reviewed by the re3data.org Team.

The research data repository is either certified or supports a repository standard.

The research data repository provides a policy.

The research data repository provides open/restricted/closed access to its data.

The research data repository uses a persistent identifier system to make its provided data persistent, unique and citable.

The terms of use and licenses of the data are provided by the research data repository.

GENERAL INFORMATION

QUALITY STANDARDS

POLICY

RESEARCH DATA REPOSITORY

TECHNICAL STANDARDS

LEGAL ASPECTS

pi

GFZ
Helmholtz Centre
POTSDAM

HUMBOLDT-UNIVERSITÄT ZU BERLIN

KIT
Karlsruhe Institute of Technology

PURDUE
UNIVERSITY

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Quality

Possible values for each icon

Information	Access	Licenses	Persistent Identifier	Certificates and Standards	Policy	Reviewed
i	🔓	©	doi	◎	§	✔
	🔒		urn			✖
	🔒		ark			
			hdl			
			purl			
			other			

Requirements:

- be run by a **legal entity**, such as a sustainable institution (e.g. library, university);

- **clarify access conditions** to the data and repository as well as the **terms of use**;

- have an **English GUI**;

- have **focus** on research data.

GFZ Helmholtz Centre POTSDAM · HUMBOLDT-UNIVERSITÄT ZU BERLIN · KIT Karlsruhe Institute of Technology · PURDUE UNIVERSITY · re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

Workflow

suggest → check → index → publish
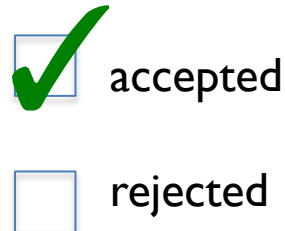
Suggest a Data Repository
* Erforderlich

Repository Name *
The full name of the research data repository.

Repository URL *
The URL, which gives reference to the research data repository.

Description *
A textual description containing additional information about the data repository (primary language is English).

Language
The user interface language of the research data repository.

http://www.re3data.org/suggest/

✔ accepted

☐ rejected

Possible values for each icon

Information · Access · Licenses · Persistent Identifier · Certificates and Standards · Policy · Reviewed

doi
urn
ark
hdl
purl
other

http://www.re3data.org/schema/

PANGAEA

General Institutions Terms Standards

eneral information

name of repository	PANGAEA
additional name	Publishing Network for Geoscientific and Environmental Data
repository URL	http://www.pangaea.de

description: The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.

content types
keywords
repository type   disciplinary
research data repository   eng
anguage(s)
ata and/or service   dataProvider
rovider

http://service.re3data.org/repository/r3d100010134

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Home   **Search**   Browse   Suggest   FAQ   About   Schema   Contact   Imprint

Search for Repositories (1028 Reviewed Repositories)

simple search box

geosciences   ✕   🔍 **Search**

filters

Subject	Content Type	Country (of the responsible institutions)
Add subjects ▾	Add content types ▾	Add countries ▾
		⊕ Germany

☑ 🔴 Certificates   ☑ 🔒 Open Access   ☑ **pi** Persistent Identifier

Clear

results

2 results (filtered) (1 – 2) ▾

icons

Sort by weight ▾

PANGAEA
Publishing Network for Geoscientific and Environmental Data

ⓘ 🔒 © doi 🔴 §

Subjects: Atmospheric Science and Oceanography   Biology   Geochemistry, Mineralogy and Crystallography   Geochemistry, Mineralogy and Crystallography   Geology and Palaeontology   Geology and Palaeontology   Geophysics   Geophysics and Geodesy   Geosciences (including Geography)   Life Sciences   Natural Sciences   Oceanography

Content types: Archived data   Audiovisual data   Images   Plain text   Standard office documents

Countries: Germany

The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.

PANGAEA

‹ Back to results

General · Institutions · Terms · Standards

General information

Name of repository	**PANGAEA**
Additional name	Publishing Network for Geoscientific and Environmental Data
Repository URL	http://www.pangaea.de
Subjects	🔍 Atmospheric Science and Oceanography 🔍 Biology 🔍 Geochemistry, Mineralogy and Crystallography 🔍 Geochemistry, Mineralogy and Crystallography 🔍 Geology and Palaeontology 🔍 Geology and Palaeontology 🔍 Geophysics 🔍 Geophysics and Geodesy 🔍 Geosciences (including Geography) 🔍 Life Sciences 🔍 Natural Sciences 🔍 Oceanography
Description	The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.
Content types	🔍 Archived data 🔍 Audiovisual data 🔍 Images 🔍 Plain text 🔍 Standard office documents
Keywords	🔍 Earth Science 🔍 Environmental Science
Repository type	disciplinary
Research data repository language(s)	eng
Data and/or service provider	dataProvider

‹ Back to results   ✏ Text Edit   👁 GUI Edit

PANGAEA

General   **Institutions**   **Terms**   **Standards**

Responsible institutions (2)

Institution name	**Alfred Wegener Institute for Polar and Marine Research**
Additional name	AWI Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung
URL	http://www.awi.de/en/home/
Contact	hgrobe@pangaea.de
Country	Germany
Type(s) of responsibility	general technical
Type of institution	non-profit

Institution name	**Center for Marine Environmental Sciences (MARUM)**
URL	http://www.marum.de/
Contact	mdiepenbroek@pangaea.de
Country	Germany
Type(s) of responsibility	general
Type of institution	non-profit

✎ Text Edit   👁 GUI Edit

PANGAEA

General   **Institutions**   **Terms**   **Standards**

Terms (1)

Policy name	Data policy of the information system PANGAEA
URL	http://www.pangaea.de/curator/files/pangaea-data-policy.pdf

Legal aspects

Database access

Type of access to research data repository	open

Data access

Type of access to data	open

Data licences (1)

License name	CC
License URL	http://wiki.pangaea.de/wiki/License

Data upload

Type of data upload	restricted
Data upload restriction type	registration
Data upload license name	Data Submission
Data license URL	http://wiki.pangaea.de/wiki/Data_submission

✏ Text Edit   👁 GUI Edit

PANGAEA

General   **Institutions**   **Terms**   **Standards**

Standards

Name of the repository software	other
Versioning	yes
Persistent identifier system	DOI
Data citation guideline	http://wiki.pangaea.de/wiki/Citation
Quality management	yes
Certificates and Standards	WDS

Application programming interfaces (1)

API type	OAI-PMH
URL	http://ws.pangaea.de/oai/

Alerting services (1)

Type of alerting service	RSS
Alerting service	http://www.pangaea.de/tools/latest-datasets.rss

Remarks

Remarks	Data of World Data Center for Marine Environmental Sciences (WDC-MARE) are available via the data library PANGAEA which will be operated as a member of the new WDS (World Data System)
Entry date	2012-07-16
Last update	2014-04-14

✎ Text Edit   👁 GUI Edit

RDR Typology

- Institutional

- Disciplinary

- Multidisciplinary

- Project

Growth

Growth of re3data.org

RDR by Country

Legend:
- US — 47%
- GER — 14%
- UK — 13%
- CAN — 6%
- FRAN — 4%
- JPN — 3%
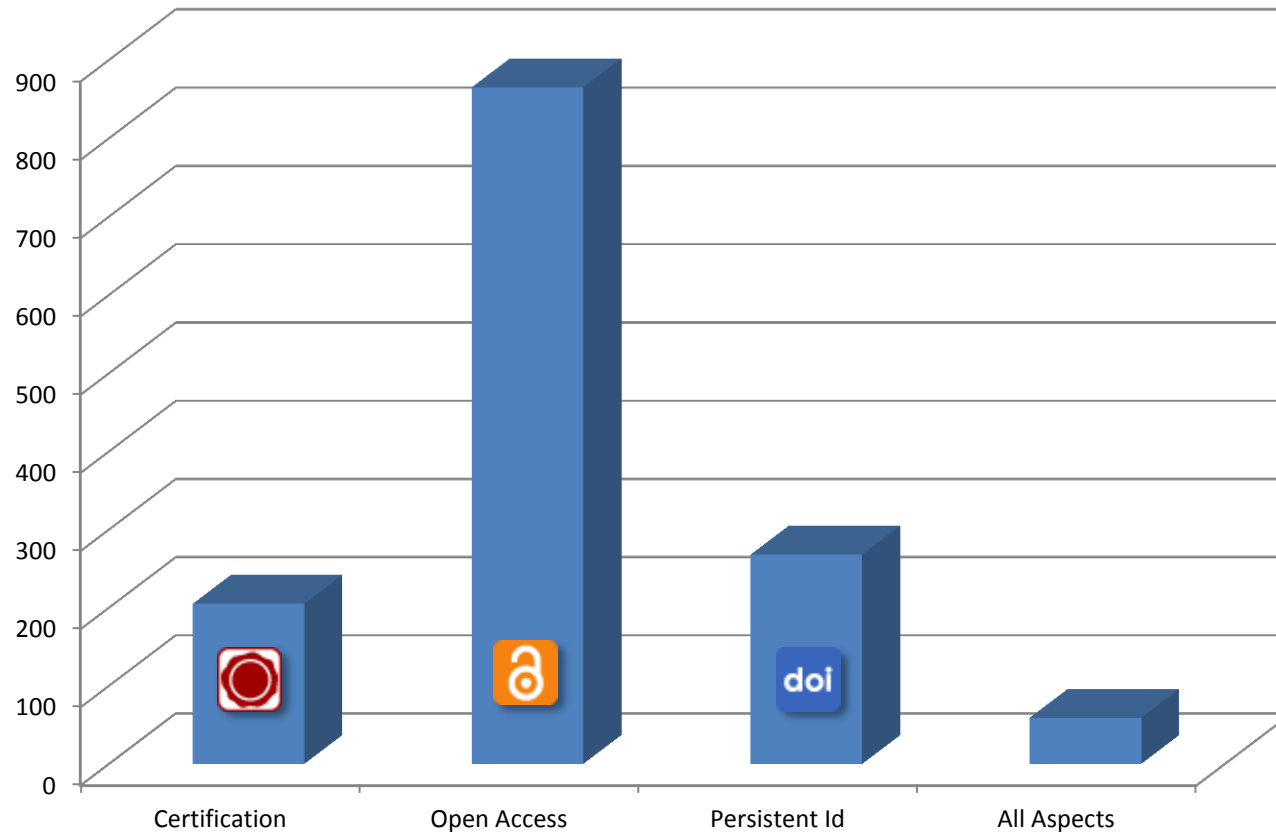- AUS — 3%
- CH — 3%
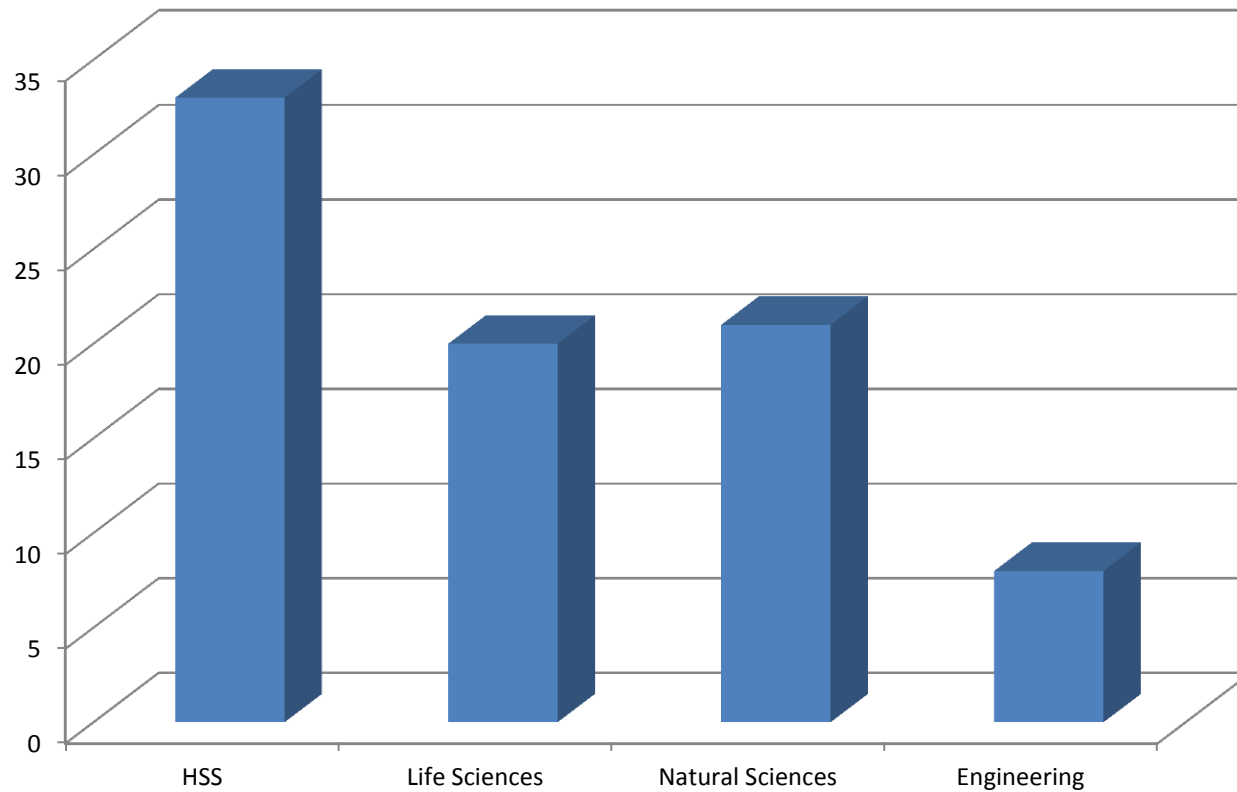- IND — 2%
- NED — 2%
- CHN — 2%
- DEN — 1%

Icons and numbers

From a total of 1029 RDR in re3data

Champions by discipline

Cooperation

- German Initiative for Network Information (DINI)

- DataCite (MoU, April 2012)

- OpenAIRE (MoU, October 2013)

- BioSharing (MoU, November 2013)

- Databib, DataCite (MoU, March 2014)

DataCite, re3data.org, and Databib Announce Collaboration

Published by Sergio Ruiz on 26 March 2014 - 11:46am

Databib and "re3data.org - Registry of Research Data Repositories" are pleased to announce their plan to merge their two projects into one service that will be managed under the auspices of DataCite by the end of 2015. Their joint proposal to the DataCite General Assembly was approved at the General Assembly on March 25th, in advance of the 3rd Plenary Meeting of the Research Data Alliance (RDA) in Dublin, Ireland.

The aim of this merger is to reduce duplication of effort and to better serve the research community with a single, sustainable registry of research data

Next steps

- Operate workflow system

- Develop API (Open Data)

- Finalize the merger between Databib and re3data.org

- Implement an international Editorial Board (based on the Databib Board)

- Roadmap for running the service under the auspices of DataCite from 2016 on

- Further engagement with the data repository community

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

info@re3data.org

http://re3data.org

GFZ
Helmholtz-Zentrum
POTSDAM

HUMBOLDT-UNIVERSITÄT ·ZU·BERLIN·

KIT
Karlsruhe Institute of Technology

PURDUE
UNIVERSITY
LIBRARIES