

# 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 Commission. (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](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf)

# Background

- 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.”

## 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  
Yes – all data are fully available without restriction  
No – some restrictions will apply

Please select a response.

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>

# Reproducibility and trust

From: [Infect Immun. 2011 October; 79\(10\): 3855–3859.](#)

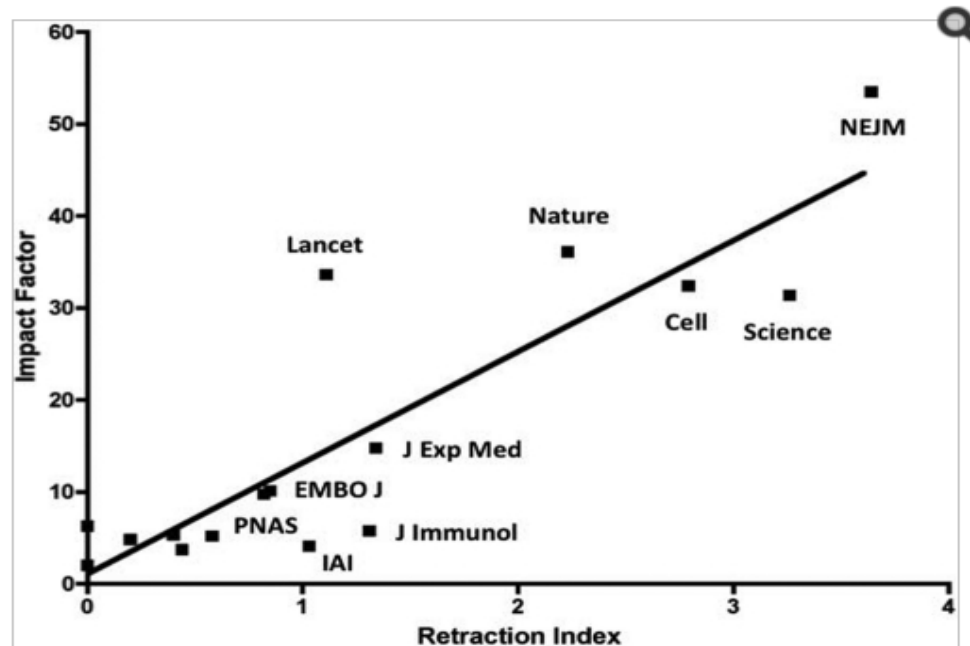
doi: 10.1128/IAI.05661-11

<< Prev Next >>

[Copyright/License ▶](#)

[Request permission to reuse](#)

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.





# Quality

Possible values for each icon

Information	Access	Licenses	Persistent Identifier	Certificates and Standards	Policy	Reviewed

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.





## Search for Repositories (1028 Reviewed Repositories)

geosciences

Search

Subject

Content Type

Country (of the responsible institutions)

Add subjects

Add content types

Add countries

Germany

Certificates

Open Access

Persistent Identifier

Clear

2 results (filtered) (1 - 2)

Sort by weight

### PANGAEA

Publishing Network for Geoscientific and Environmental Data



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.

simple search box

filters

results

icons


[← Back to results](#)
[General](#)
[Institutions](#)
[Terms](#)
[Standards](#)

### General information

Name of repository	<b>PANGAEA</b>
Additional name	Publishing Network for Geoscientific and Environmental Data
Repository URL	<a href="http://www.pangaea.de">http://www.pangaea.de</a>
Subjects	<a href="#">Q Atmospheric Science and Oceanography</a> <a href="#">Q Biology</a> <a href="#">Q Geochemistry, Mineralogy and Crystallography</a> <a href="#">Q Geochemistry, Mineralogy and Crystallography</a> <a href="#">Q Geology and Palaeontology</a> <a href="#">Q Geology and Palaeontology</a> <a href="#">Q Geophysics</a> <a href="#">Q Geophysics and Geodesy</a> <a href="#">Q Geosciences (including Geography)</a> <a href="#">Q Life Sciences</a> <a href="#">Q Natural Sciences</a> <a href="#">Q Oceanography</a>
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	<a href="#">Q Archived data</a> <a href="#">Q Audiovisual data</a> <a href="#">Q Images</a> <a href="#">Q Plain text</a> <a href="#">Q Standard office documents</a>
Keywords	<a href="#">Q Earth Science</a> <a href="#">Q Environmental Science</a>
Repository type	disciplinary
Research data repository language(s)	eng
Data and/or service provider	dataProvider

[← Back to results](#)
[Text Edit](#)
[GUI Edit](#)

[← Back to results](#)[General](#)[Institutions](#)[Terms](#)[Standards](#)**Responsible institutions** (2)

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

Institution name	<b>Center for Marine Environmental Sciences (MARUM)</b>
URL	<a href="http://www.marum.de/">http://www.marum.de/</a>
Contact	<a href="mailto:mdiepenbroek@pangaea.de">mdiepenbroek@pangaea.de</a>
Country	<b>Germany</b>
Type(s) of responsibility	general
Type of institution	non-profit

[← Back to results](#)[Text Edit](#)[GUI Edit](#)

[← Back to results](#)[General](#)[Institutions](#)[Terms](#)[Standards](#)

### Terms (1)

Policy name	<b>Data policy of the information system PANGAEA</b>
URL	<a href="http://www.pangaea.de/curator/files/pangaea-data-policy.pdf">http://www.pangaea.de/curator/files/pangaea-data-policy.pdf</a>

### 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	<b>CC</b>
License URL	<a href="http://wiki.pangaea.de/wiki/License">http://wiki.pangaea.de/wiki/License</a>

#### Data upload

Type of data upload	restricted
Data upload restriction type	registration
Data upload license name	Data Submission
Data license URL	<a href="http://wiki.pangaea.de/wiki/Data_submission">http://wiki.pangaea.de/wiki/Data_submission</a>

[← Back to results](#)[Text Edit](#)[GUI Edit](#)

[← Back to results](#)[General](#) [Institutions](#) [Terms](#) [Standards](#)

### Standards

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

### Application programming interfaces <sup>(1)</sup>

API type	OAI-PMH
URL	<a href="http://ws.pangaea.de/oai/">http://ws.pangaea.de/oai/</a>

### Alerting services <sup>(1)</sup>

Type of alerting service	RSS
Alerting service	<a href="http://www.pangaea.de/tools/latest-datasets.rss">http://www.pangaea.de/tools/latest-datasets.rss</a>

### 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

[← Back to results](#)[Text Edit](#)[GUI Edit](#)

# RDR Typology

- Institutional

- Disciplinary

- Multidisciplinary

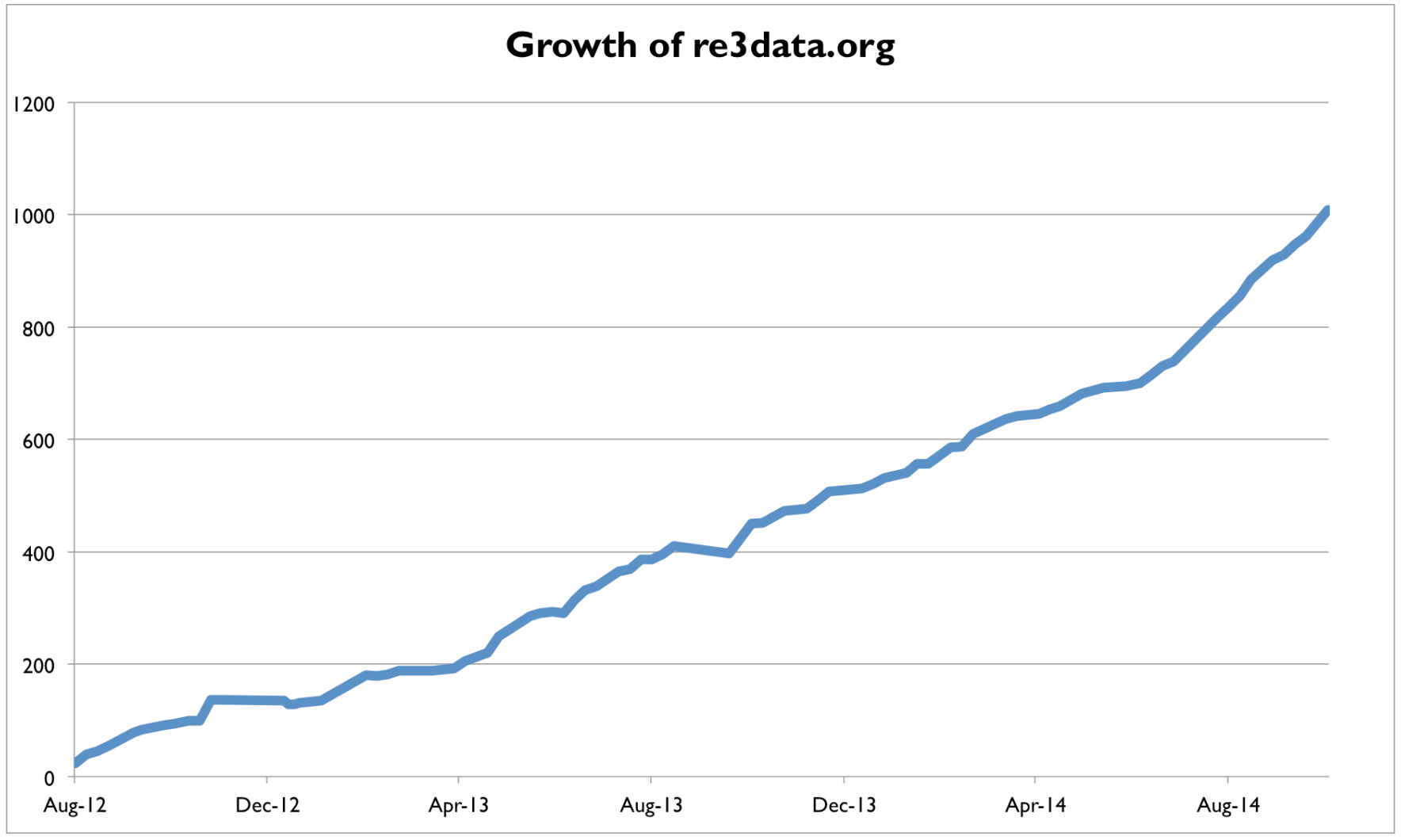
- Project

The collage illustrates four types of Research Data Repositories (RDRs):

- NCBI GEO (Institutional):** A screenshot of the Gene Expression Omnibus website, showing search options like 'DataSets', 'Gene profiles', and 'GEO BLAST'.
- Purdue University Research Repository (PURR) (Disciplinary):** A screenshot of a research paper page titled 'Phosphorus and Potassium Influence on Alfalfa Nutrition', featuring a 'Download' button and a 'SEE ALSO' section.
- The Bern Digital Pantheon Project (Multidisciplinary):** A screenshot of a repository page for 'Herring Hepatic Transcriptome 34300 contigs.fa', displaying a list of contigs with their average coverage and a 'Preview' button.
- Figshare (Project):** A screenshot of a Figshare page for the same transcriptome data, showing a 'Download' button and a 'Usage of the repository' section.

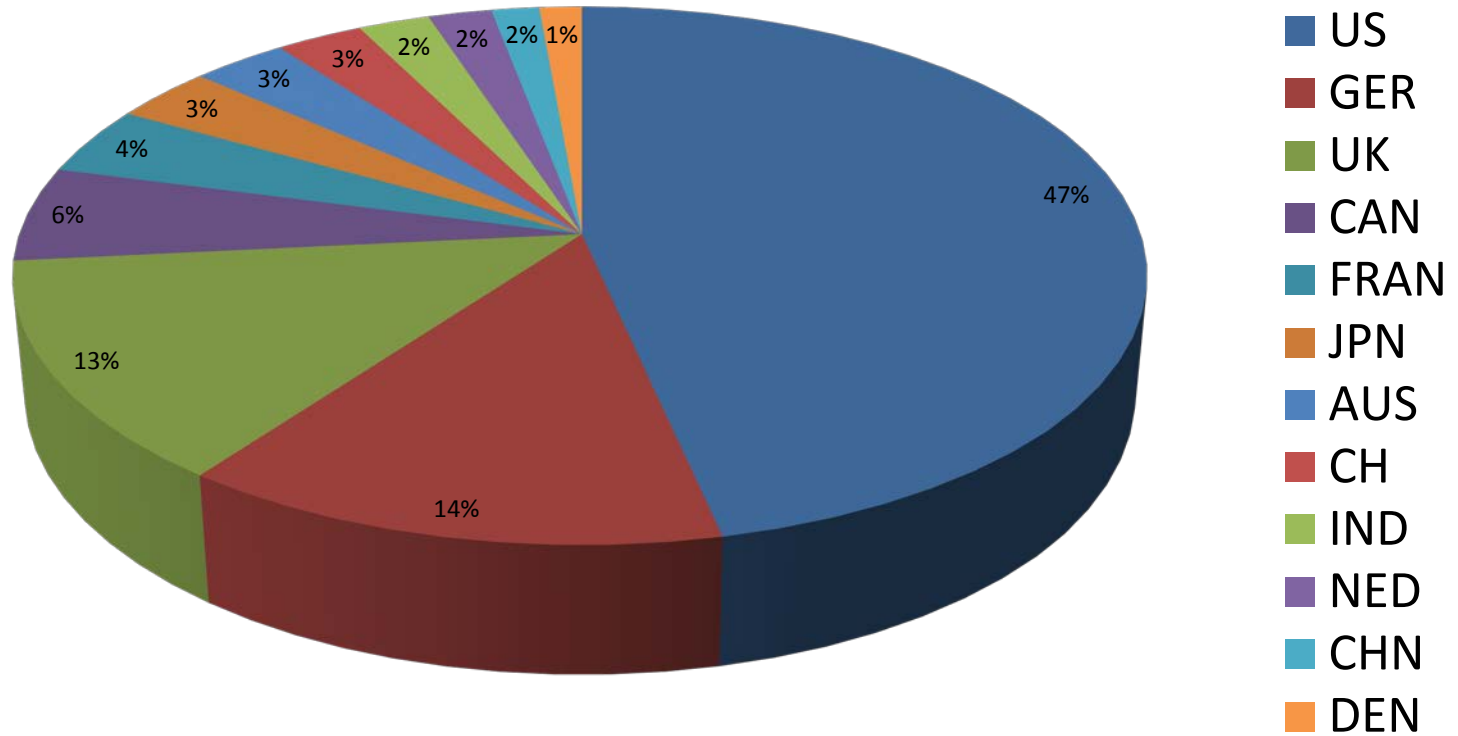


# Growth



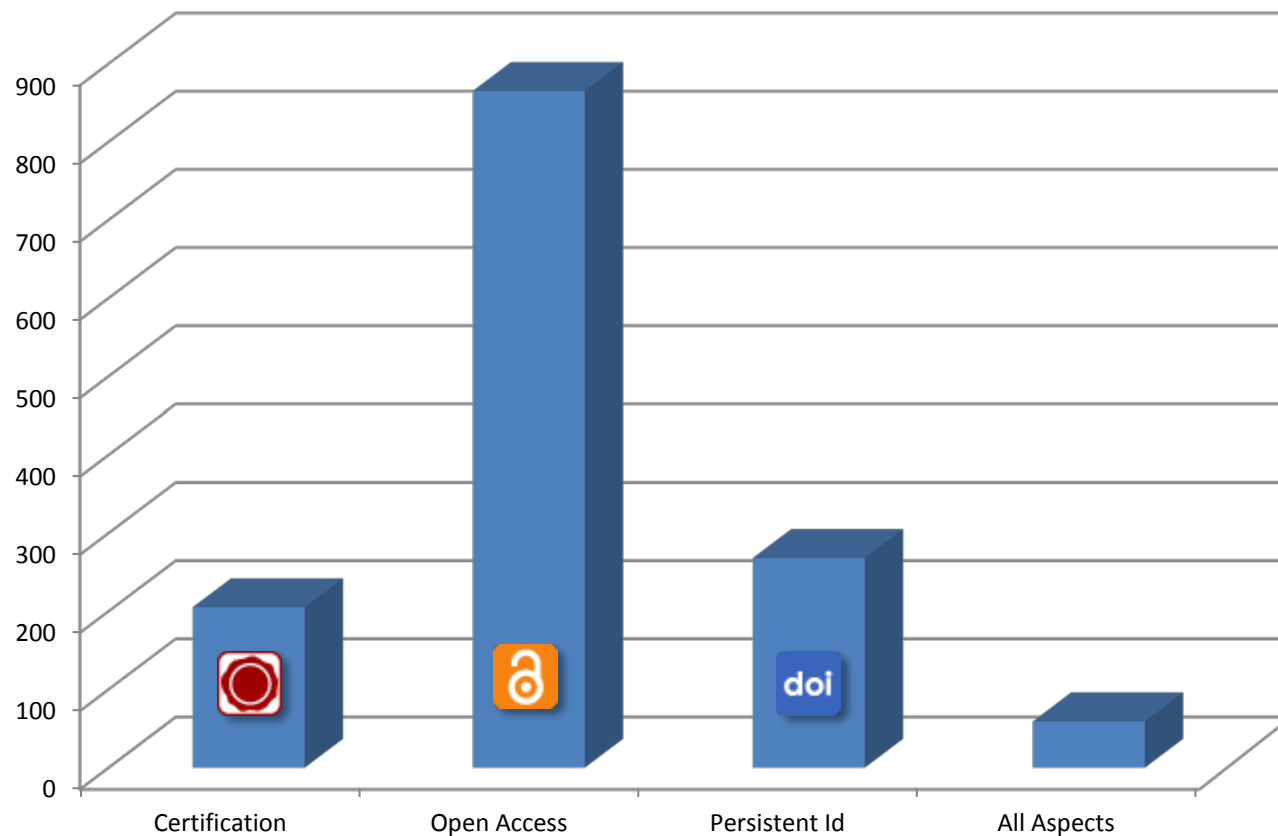


# RDR by Country



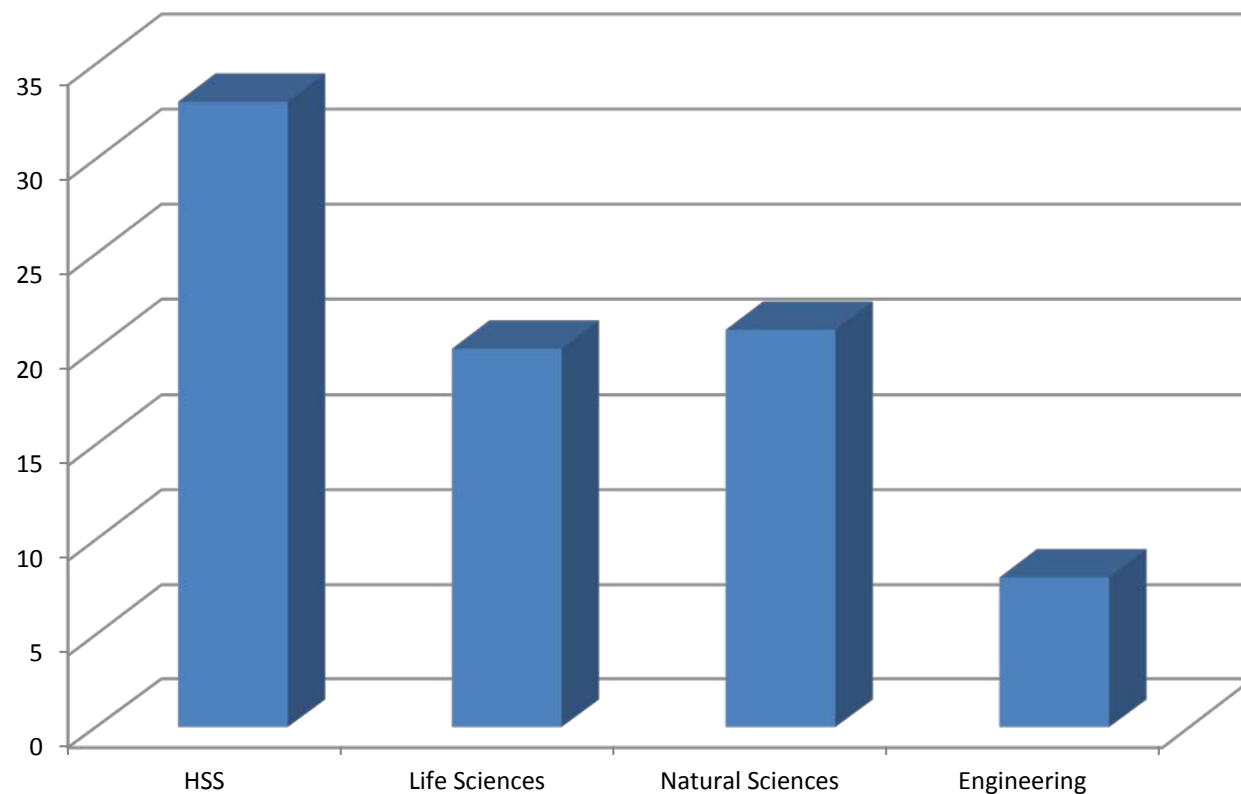


# Icons and numbers



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)




OpenAIRE  
Open Access Infrastructure for Research in Europe



biosharing

Home Members FAQs Services Resources Events Contact us

 Helping you to find, access, and reuse data  
DataCite

**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

Why cite data?  
What is DataCite?  
What do we do?

# 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
- Further engagement with the data repository community

info@re3data.org

http://re3data.org



With the exception of all photos and graphics, this slides are licensed under the “Attribution 4.0 International (CC BY 4.0)“ Licence:  
<http://creativecommons.org/licenses/by/4.0/>





# Backup Slides

# re3data.org & RDA



- Platform for networking and exchange
  - Schema: requests for comments
- Repository Audit and Certification DSA–WDS Partnership WG
  - Heterogeneous landscape of data repositories
  - 20 % of the indexed repositories are certified
- Vision: “Data Fabric IG”



# Champions



- CLARIN-ERIC
- Archaeology Data Service
- EASY (DANS)
- IQSS Dataverse network
- CrystalEye (beta)
- Durham HepData Project
- figshare
- Global Change Master Directory (NASA)
- NeuroMorpho
- Neuroscience Information Framework
- ClinicalTrials.gov
- Ecological Archives (ESA)
- PANGAEA
- WDC for Remote Sensing of the Atmosphere

# Cooperation

- Databib and re3data.org have agreed to the following five principles for successful cooperation:
  1. Openness
  2. Optimal quality assurance
  3. Development of innovative functionalities
  4. Shared leadership
  5. Sustainability

<http://www.datacite.org/node/115>

# Project Team (Status: Oct. 2014 )

- Bertelmann, R. <sup>a</sup>; Fuchs, C. <sup>a</sup>; Goebelbecker, H.-J. <sup>c</sup>;  
Kindling, M. <sup>b</sup>; Kloska, G. <sup>c</sup>; Pampel, H. <sup>a</sup>; Reuter, E. <sup>c</sup>;  
Rücknagel, J. <sup>b</sup>; Schirnbacher, P. <sup>b</sup>; Schnepf, E. <sup>c</sup>; Scholze,  
F. <sup>c</sup>; Semrau, A. <sup>c</sup>; Skarupianski, M. <sup>c</sup>; Ulrich, R. <sup>c</sup>; Vierkant,  
P. <sup>b</sup>; Witt, M. <sup>d</sup>

a GFZ German Research Centre for Geosciences, Library and Information Services (LIS)

b Humboldt-Universität zu Berlin, Berlin School of Library and Information Science (BLIS)

c Karlsruhe Institute of Technology (KIT), KIT Library

d Purdue University, Distributed Data Curation Center (D2C2)

# Partners

- Germany

- Berlin School of Library and Information Science
- GFZ German Research Centre for Geosciences, Library and Information Services (LIS)
- Karlsruhe Institute of Technology (KIT), KIT Library
- Funded by the German Research Foundation

- USA

- Purdue University, Purdue Libraries
- Funded by the Institute of Museum and Library Services (IMLS)



# DATABIB EDITORIAL BOARD

- Amy West, University of Minnesota, United States
- Andras Micsik, MTA SZTAKI, Hungary
- Catherine Jones, Science and Technology Facilities Council, United Kingdom
- Cristina Ribeiro, University of Porto, Portugal
- Dorothea Salo, University of Wisconsin, United States
- Fco. Javier Hernández San Miguel, University of Valencia, Spain
- Gail Steinhart, Cornell University, United States
- Jiban K. Pal, Indian Statistical Institute, India
- Jochen Schirrwagen, Bielefeld University, Germany
- Kassim S. Mwitondi, Sheffield Hallam University, United Kingdom
- Kei Kurakawa, National Institute of Informatics, Japan
- Lynn Yarmey, Snow and Ice Data Center, United States
- Michael Witt, Editor-in-Chief, Purdue University, United States
- Sarah Williams, University of Illinois, United States
- Steve Marks, University of Toronto, Canada
- Xiuling Qing, Chinese National Academy of Sciences, China



# WHERE DO YOU STORE YOUR RESEARCH DATA?

- USB DRIVE
- DROPBOX
- RESEARCH DATA REPOSITORY

**re3**data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

## Thanks for your attention!