

Making Research Data Repositories Visible – The re3data.org Registry

Frank Scholze | Karlsruhe Institute of Technology, KIT Library

Heinz Pampel, Paul Vierkant | GFZ German Research Centre for Geosciences, LIS

LIBER 2015 | London, June 26, 2015

June 25, 2015

Subscribe Today

Home

News

Global

Opinion & Ideas

Facts & Figures

Blogs

Advice

Forums

Jobs

Search The Chronicle

Go

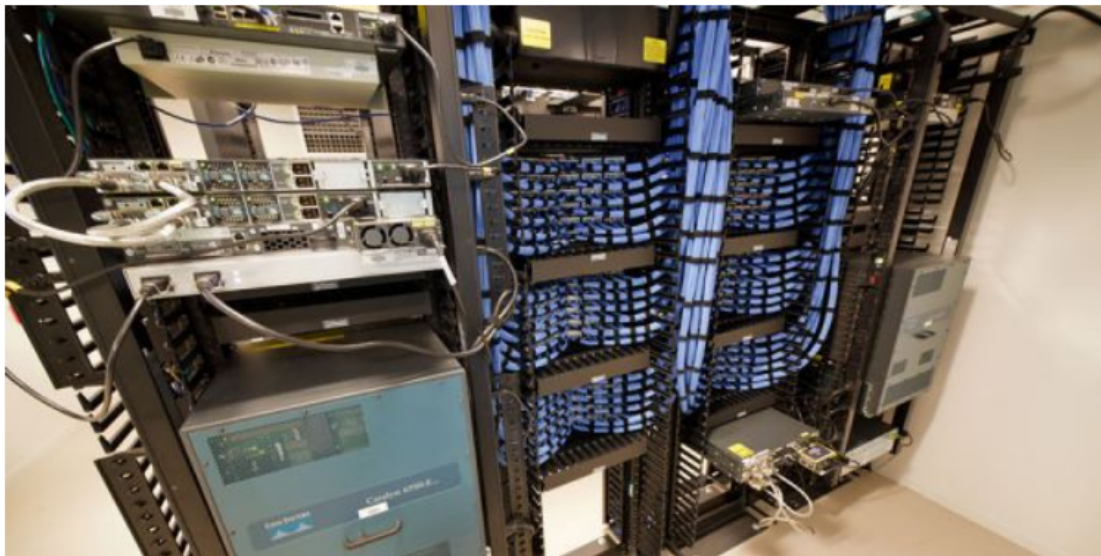
Advice

June 23, 2015



Where Should You Keep Your Data?

Formal policies on information storage are finally beginning to emerge from federal grant agencies



Roy Niswanger / Creative Commons

By Karen M. Markin

Most Popular

Most Viewed

Most Commented

1. Reining In a Fraternity Is No Easy Task, One University's Leaders Learn
2. The New York Public Library Wars
3. Why State Lawmakers Must Support Tenure at Public Universities
4. 'Let's Make This Work': Sweet Briar's New President Surveys the Task Ahead
5. Can a Longtime Fraud Help Fix Science?

ICLE CAREERS

Featured job from



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 **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 I);
- (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

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 >>

From

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley¹, Lee M. Ellis²,

Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

[◀ back to article](#)

Table 1: Reproducibility of research findings

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

* Source of citations: Google Scholar, May 2011.

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.org - 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

39

Based on Analyses,
Feedback and Experience

Version

2.2

Icons



The research data repository provides additional information on its service.

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.



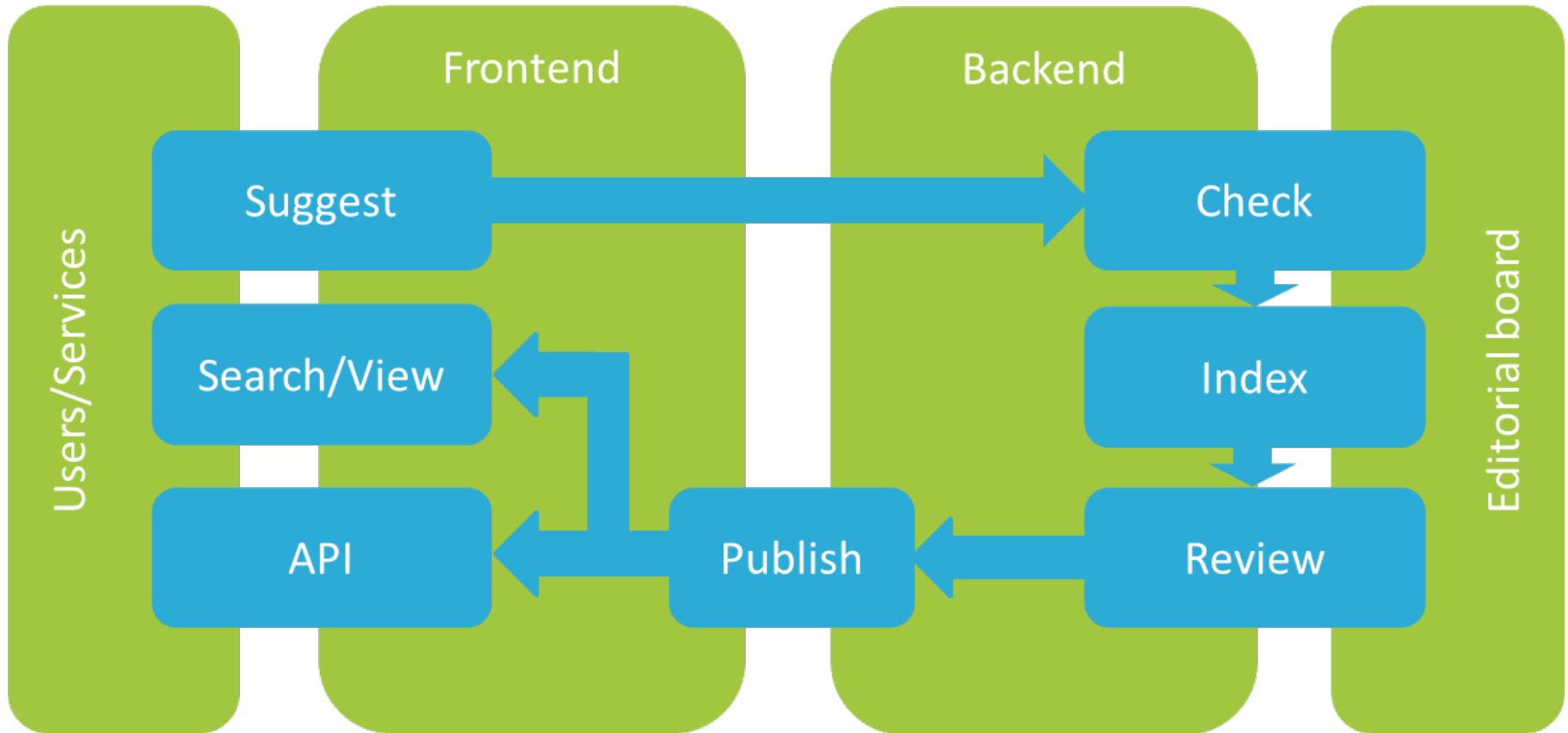
The terms of use and licenses of the data are provided by the research data repository.

Quality

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 **focus on research data**

Workflow



Search for Repositories (1275 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

Reset filter

3 results (filtered) (1 - 3)

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

Repository details



PANGAEA

General Institutions Terms Standards

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

Repository details



PANGAEA

General Institutions Terms Standards

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

Repository details



PANGAEA

General Institutions **Terms** Standards

Policies (1)

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

Database access

Type of access to research data repository	open
--	------

Data access (1)

Type of access to data	open
------------------------	------

Data licenses (1)

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

Data upload (1)

Type of data upload	restricted
Data upload restriction type(s)	registration

Data upload license (1)

Data upload license name	Data Submission
Data upload license URL	http://wiki.pangaea.de/wiki/Data_submission

Repository details



PANGAEA

General Institutions Terms Standards

Name of the repository software other

Versioning yes

Persistent identifier system(s) DOI

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

RDR Typology

- Institutional

- Disciplinary

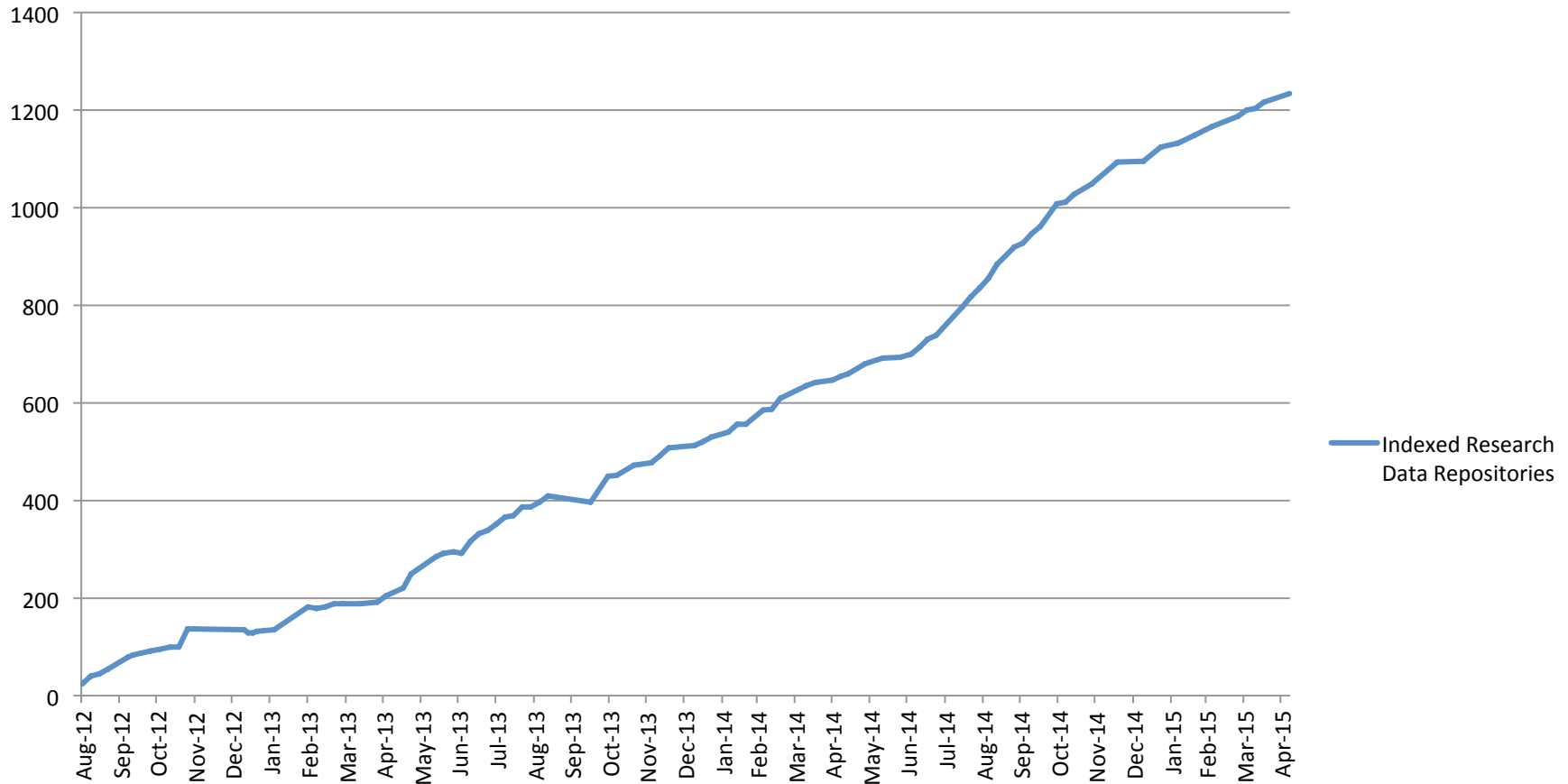
- Multidisciplinary

- Project

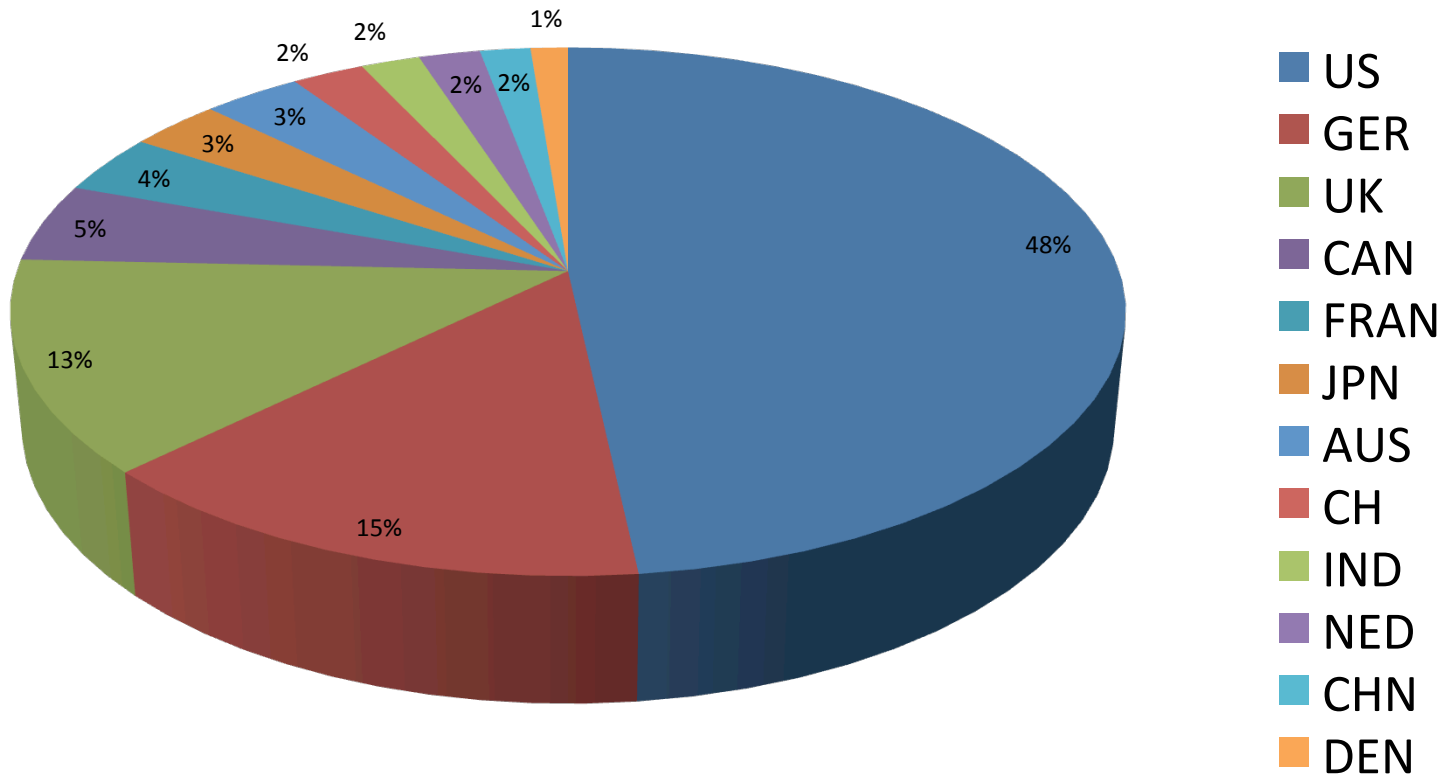
The collage features four overlapping screenshots of research data repositories:

- NCBI GEO:** Shows the 'Gene Expression Omnibus' homepage with navigation options like 'DataSets', 'Gene profiles', and 'GEO BLAST'. It displays statistics for public data: 9,127 Platforms and 604,248 Samples.
- Purdue University Research Repository (PURR):** Displays a dataset titled 'Phosphorus and Potassium Influence on Alfalfa Nutrition'. It includes a 'Download (XLS)' button, a 6.0 ranking, and a 'Share' section with social media icons.
- The Bern Digital Pantheon Project:** Shows a repository page for 'Herring Hepatic Transcriptome 34300 contigs.fa'. It features a large image of the Bern Cathedral dome, a 'Preview' button, and a 'Download' button. The page also lists categories like 'Molecular biology' and 'Marine Biology', and an author 'Steven Roberts'.
- Figshare:** Shows a 'My data' section with a 'Browse' button and an 'Upload' button. It also displays a 'Share' section with social media icons and a 'Cite this' section with a DOI handle.

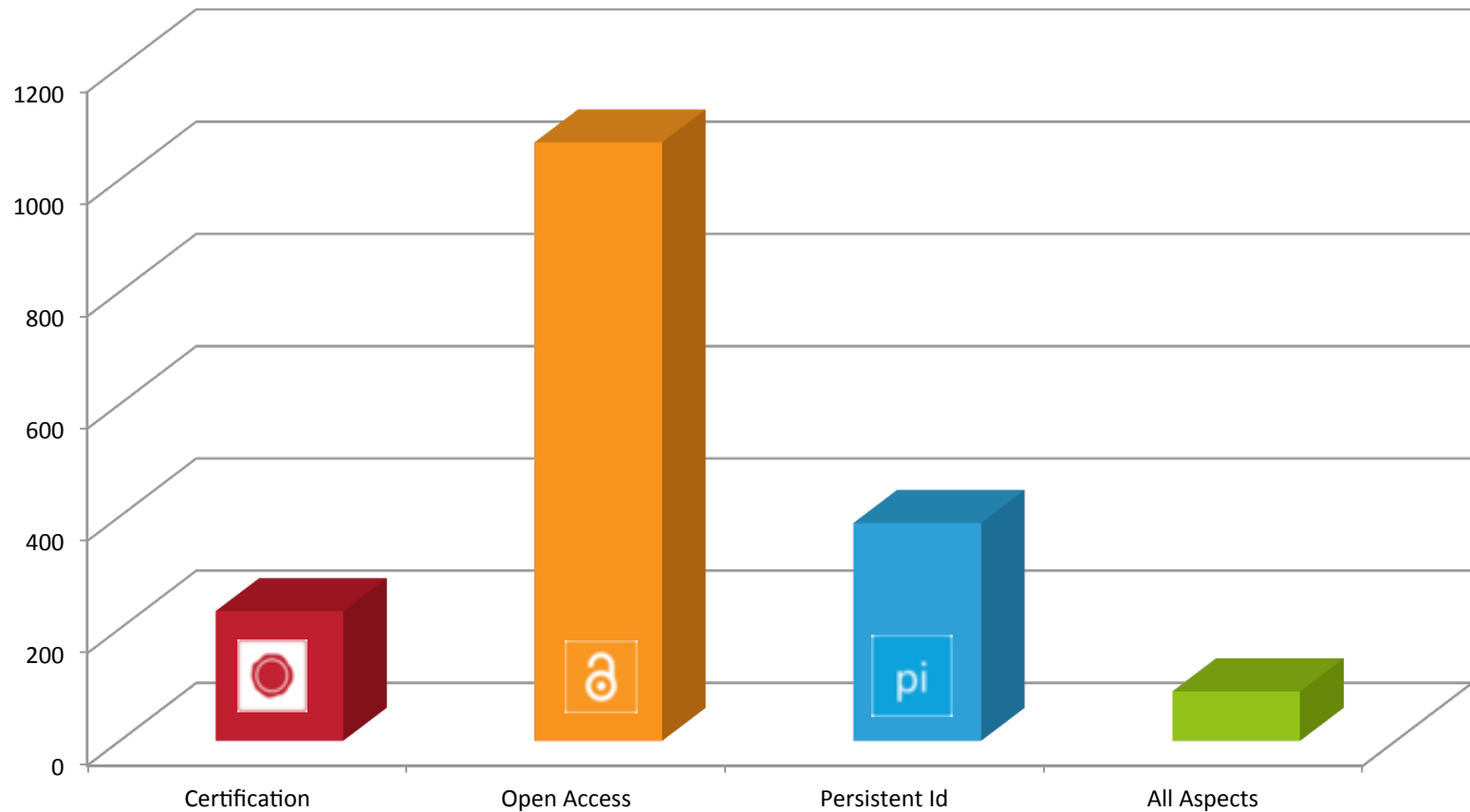
RDR indexed by re3data



RDR by Country

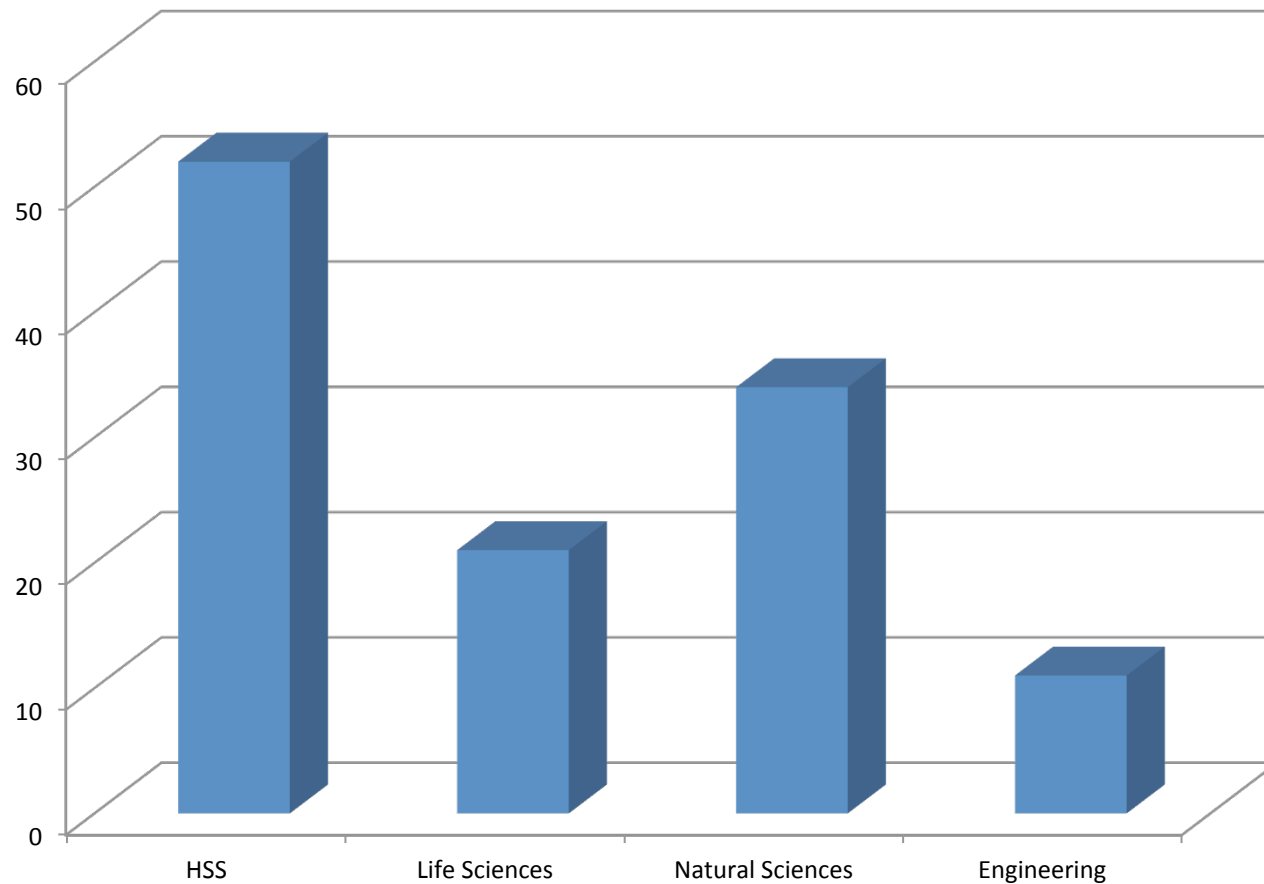


Icons and numbers



From a total of 1260 RDR in re3data (June 2015)

Champions by discipline



From a total of 88 RDR (June 2015)

Integration of re3data in Guidelines

- Funder Example: European Commission

1) **Step 1:** participating projects are required to deposit the research data described above, preferably into a research data repository. 'Research data repositories' are online archives for research data. They can be subject-based/thematic, institutional or centralised. Useful listings of research data repositories include the Registry of Research Data Repositories (www.re3data.org) and Databib (<http://databib.org>). In addition, it is expected that the Open Access Infrastructure for Research in Europe (OpenAIRE) will become an entry point for linking publications to underlying research data.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

- Institutional Example: Bielefeld University

- „Verzeichnisse, wie das DFG-geförderte "Registry of Research Data Repositories", bilden die Grundlage für die Suche nach geeigneten Publikationsorten für die Forschungsdaten.“

Universität Bielefeld (2011): Resolution zum Forschungsdatenmanagement. <https://data.uni-bielefeld.de/de/resolution>

- Publisher Example: Nature Publishing Group

- „Physics, astrophysics, astronomy and geoscience databases should be registered with re3data.org.“

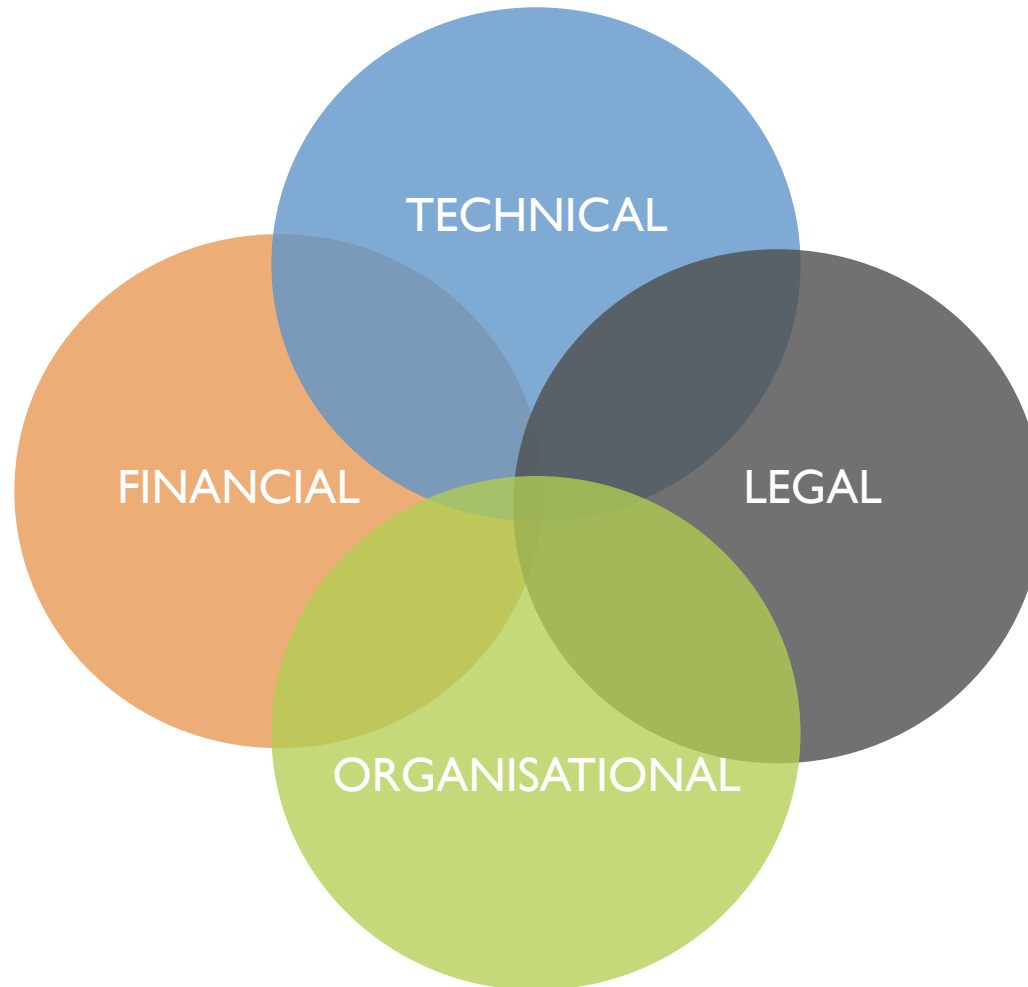
Scientific Data (2013): Data policies. <http://www.nature.com/sdata/data-policies>

Cooperation

- Deutsche Initiative für Netzwerkinformation (DINI)
- DataCite (MoU, April 2012)
- OpenAIRE (MoU, October 2013)
- BioSharing (MoU, November 2013)
- Databib (MoU, March 2014)
- DataCite (Formal cooperation, March 2015)

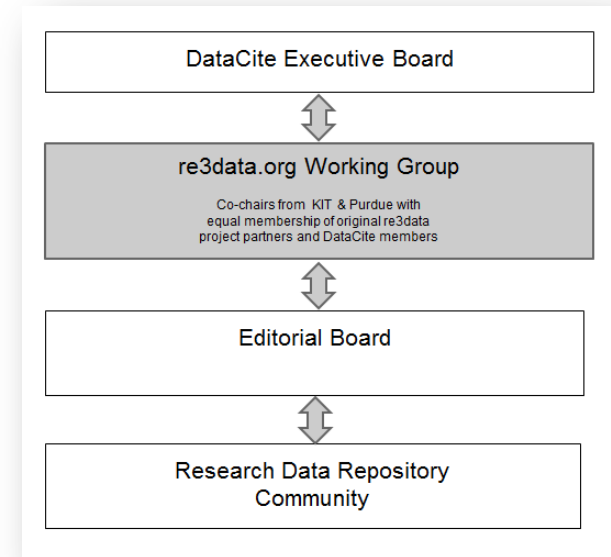


Dimensions of sustainability



Organizational sustainability

- Merger with DataBib under the auspices of DataCite
- re3data.org working group within DataCite
- International Editorial Board
- Cooperations within Research Data Alliance (RDA) and the research data repository community
- Community building and feedback loops during RFC phases (e.g. re3data.org schema)



Technical sustainability

- Open interfaces
 - RESTful API
 - OpenSearch
 - Documentation: <http://www.re3data.org/api/doc>
 - Used e.g. by OpenAIRE
- Open metadata
 - Documentation: <http://www.re3data.org/schema/>
- Long-term hosting commitment by KIT

Legal sustainability

- Open licenses
 - CC BY for the website
 - CC 0 for metadata

LICENSES



Except where otherwise noted, content on this site is licensed under a **Creative Commons Attribution 4.0 International License**.



To the extent possible under law, **re3data.org** has waived all copyright and related or neighboring rights to the database entries of re3data.org.

Financial sustainability

- Technical maintenance financed by DataCite from 2016
- Further development managed by DataCite
- Further project funding



Thanks to the team

- Michael Witt
Purdue University, Distributed Data Curation Center (D2C2)
- Roland Bertelmann, Claudio Fuchs, Heinz Pampel
GFZ German Research Centre for Geosciences, Library and Information Services (LIS)
- Maxi Kindling, Jessica Rücknagel, Peter Schirmbacher,
Paul Vierkant
Humboldt-Universität zu Berlin, Berlin School of Library and Information Science (BLIS)
- Hans-Jürgen Goebelbecker, Gabriele Kloska, Evelyn Reuter, Edeltraud Schnepf, Angelika Semrau, Michael Skarupianski, Robert Ulrich
Karlsruhe Institute of Technology (KIT), KIT Library

WHERE DO YOU STORE YOUR RESEARCH DATA?

- USB DRIVE
- DROPBOX
- RESEARCH DATA REPOSITORY

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/>

