

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Making research data repositories
visible and discoverable

Robert Ulrich – Karlsruhe Institute of Technology

Outline

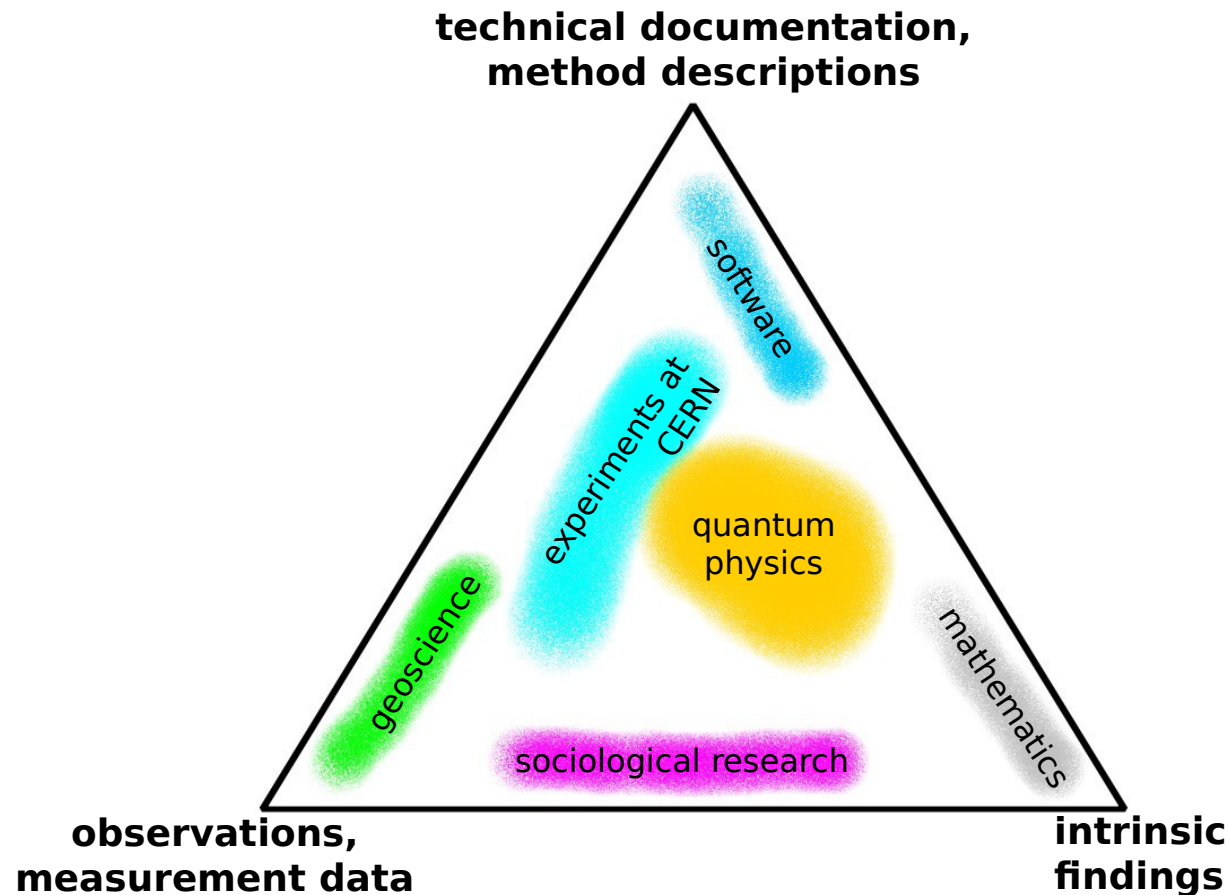
- Background
- Mission
- Schema, Icons, Quality and Workflow
- Interface
- Growth
- Cooperations
- Experiences & Best practices

Background

- Research data are valuable and ubiquitous
- New technologies facilitate data-intensive science
- Broad discussion about the permanent access to research data
- Increasing requirements from funders to make research data openly available
- Growing demand for trustable and sustainable research data repositories
- Trend: data journals

Background

Generic range of research data (with examples)



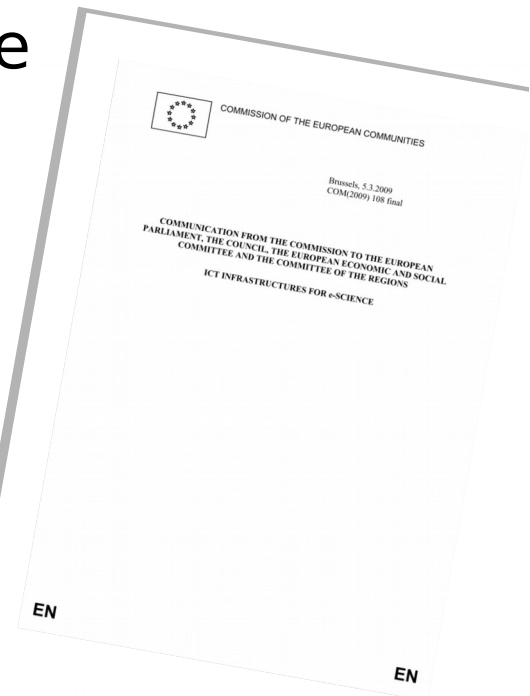
Background

- Research data are of most varied nature.
- Research data can only imperfectly be treated by an information management like conventional information/library media.
- Research data repositories (RDRs) often represent an essential stage of compression, abstraction and summary of research data, authorized and authenticated by the producers.
- RDRs can be operated centrally (institutional RDRs) or locally (disciplinary RDRs).
- In particular local or disciplinary RDRs are very popular in science because they represent a kind of a bottom up approach in research data management by the research groups themselves.

Research Data Repositories

- Highly heterogeneous landscape of research data repositories
- Different communities and different approaches
- EC (2009): ICT infrastructures for e-science

„The landscape of data repositories across Europe is fairly heterogeneous, but there is a solid basis to develop a coherent strategy to overcome the fragmentation and enable research communities to better manage, use, share and preserve data.“



The RDR Landscape



funders



research data
repositories



journals



scientists



universities and
research labs

[RRZE Icon Set](#) (CC: BY-SA)

The RDR Landscape

Investigators are expected to share their data!



funders



research data
repositories

Underlying data must be accessible!



journals

Where can I find data and store mine?

How can we set up repositories?



scientists

Should we offer repositories for all disciplines?



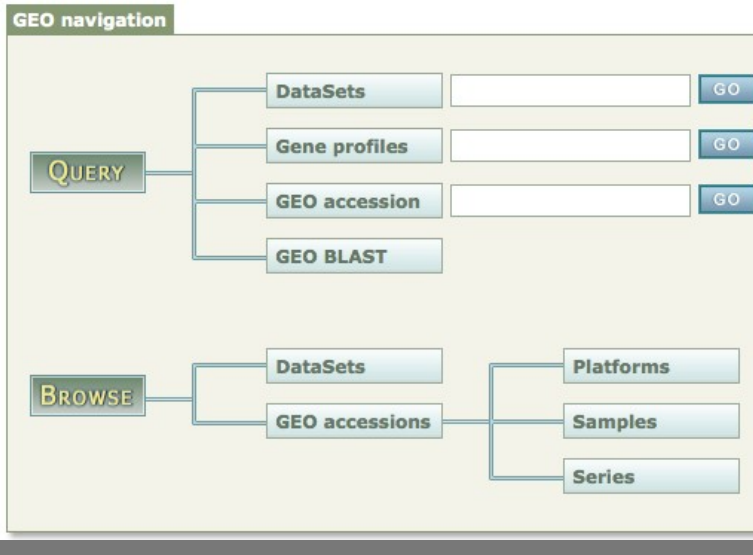
universities and
research labs

[RRZE Icon Set](#) (CC: BY-SA)

Research Data Repository



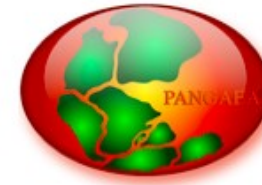
Gene Expression Omnibus: a public functional genomics data repository supporting submissions. Array- and sequence-based data are accepted. Tools are provided to help with experiments and curated gene expression profiles. [More information »](#)



GEO,
<http://www.ncbi.nlm.nih.gov/geo/>

PANGAEA®

Data Publisher for Earth & Environmental Science



[About](#) – [Submit Data](#) – [Projects](#) – [Software](#) – [WDC-MARE](#) – [Contact](#)

This work is licensed under a [Creative Commons License](#)

PANGAEA, <http://www.pangaea.de>

Mission







- is a global registry of research data repositories
- covers research data repositories from all academic disciplines
- helps researchers, funding bodies, publishers and scholarly institutions to find research data repositories
- aims to promote a culture of sharing, increased access and better visibility of research data

Schema

- general information (e.g. short description of the RDR, content types, keywords)
- responsibilities (e.g. institutions responsible for funding, content or technical issues)
- policies (e.g. policies of the RDR, incl. their URL)
- legal aspects (e.g. licenses of the database and datasets)
- technical standards (e.g. APIs, versioning of datasets, software of the RDR)
- quality standards (e.g. certificates, audit processes)



Icons

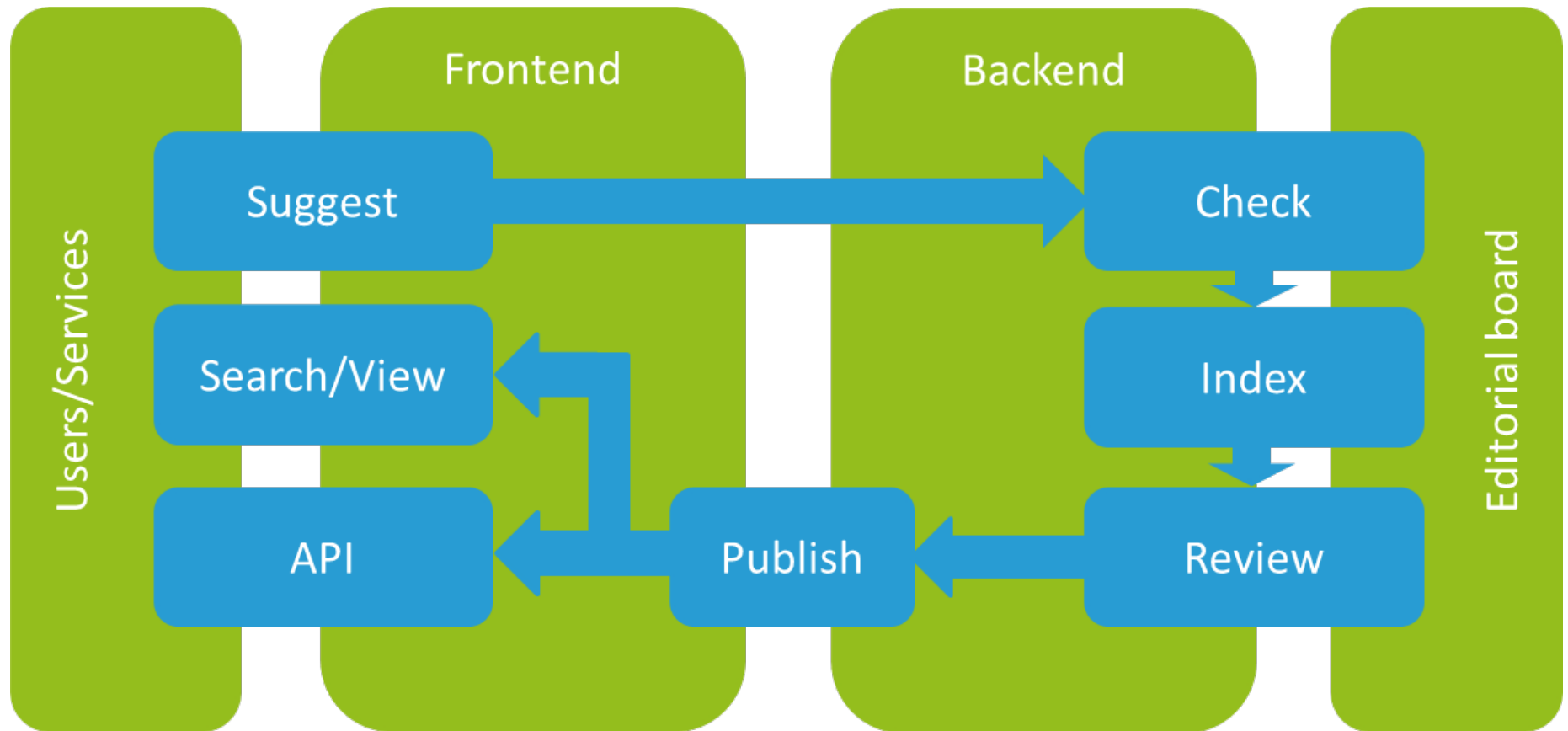
-  Additional information on its service
-  Information on the terms of access to its data, database and upload
-  Terms of use and licenses of the data
-  Persistent identifier system to make provided data unique and citable
-  Is certified or supports a repository standard.
-  Provides a policy

Quality

Requirements

- 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
- (Have an english graphical user interface)

Workflow



Interface

Search for Repositories (1234 Reviewed Repositories)

Q Search



Subject

Content Type

Country (of the responsible institutions)

☐ Certificates ☐ Open Access ☐ Persistent Identifier

Reset filter

1234 results (1 – 25)

Sort by weight

«	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	»				

3TU.Datacentrum

3TU.DC



Subjects: Agriculture, Forestry, Horticulture and Veterinary Medicine Agriculture, Forestry, Horticulture and Veterinary Medicine
Analytical Chemistry, Method Development (Chemistry) Basic Biological and Medical Research Bioinformatics and Theoretical Biology
Biology Biophysics Chemistry Computer Science Computer Science, Electrical and System Engineering
Construction Engineering and Architecture Construction Engineering and Architecture Engineering Sciences
Geochemistry, Mineralogy and Crystallography Geography Geophysics Geophysics and Geodesy Geosciences (including Geography)
Life Sciences Materials Science Materials Science and Engineering Mathematics Natural Sciences Physics Soil Sciences
Systems Engineering Traffic and Transport Systems, Logistics
Urbanism, Spatial Planning, Transportation and Infrastructure Planning, Landscape Planning Water Research

Content types: Archived data Audiovisual data Images Plain text Raw data Scientific and statistical data formats Standard office documents
Structured text

Interface

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

Interface

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

Interface

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

Interface

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

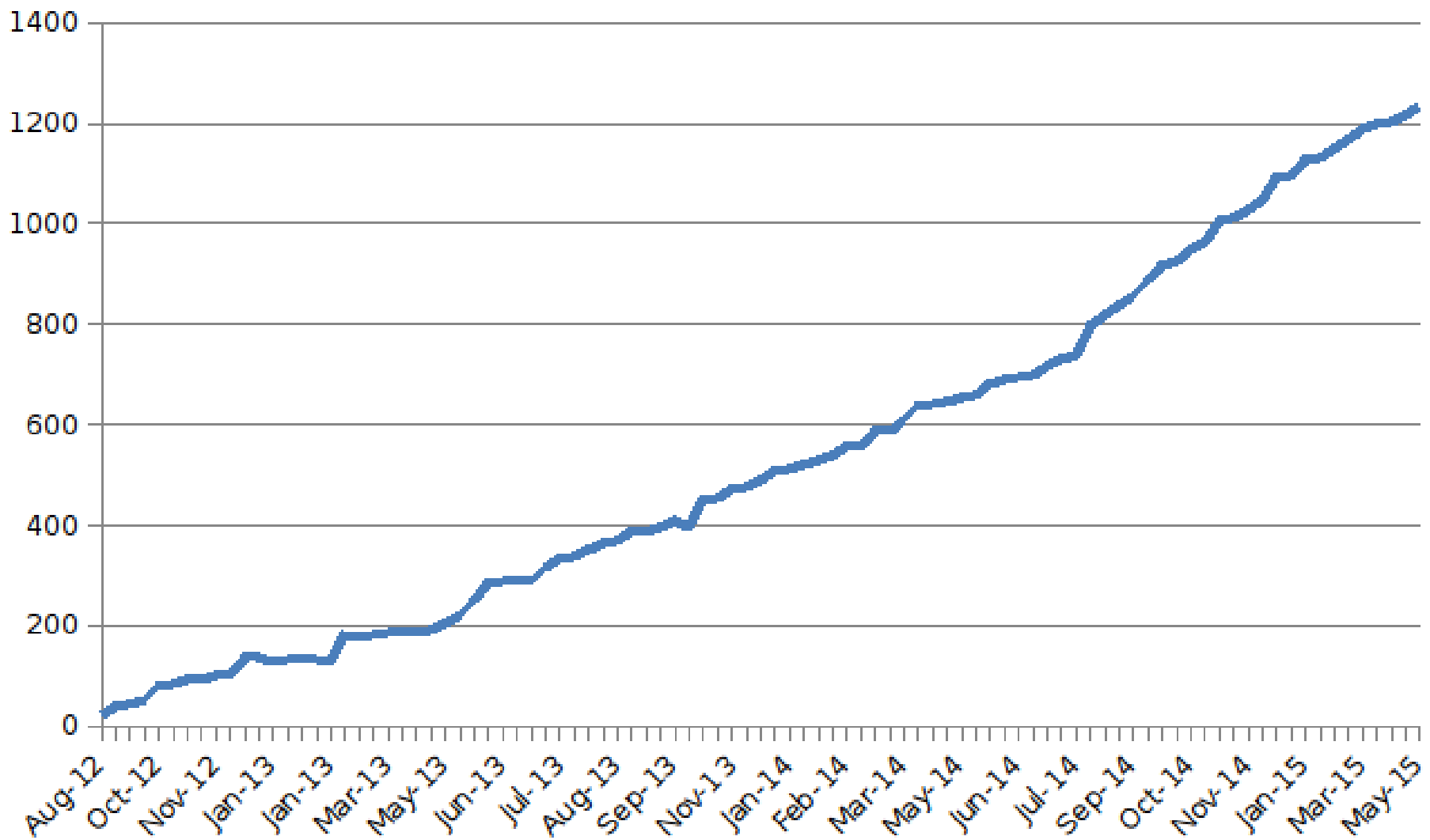
Remarks Data of World Data Center for Marine Environmental Sciences (WDC-MARE) are available via the data library PANGAEA which will be operated as a member of the new WDS (World Data System)

Growth

“Datasets are more likely to be seen, reused, and have impact if they can be found where potential reusers are likely to look. If you are unsure where that might be, the Registry of Research Data Repositories (re3data.org) provides a list of repositories organised by subject, content type and country.”

Alex Ball (DCC) and Monica Duke (DCC) - How to Measure the Impact of Research Data / A Digital Curation Centre ‘working level’ guide

Growth



Initial Partners

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



Cooperations

- German Initiative for Network Information (DINI)
- DataCite (MoU, April 2012)
- OpenAIRE (MoU, October 2013)
- BioSharing (MoU, November 2013)
- Databib (MoU, March 2014)
- RDA/WDS IG on Certification



Join forces

Databib and re3data.org have agreed to the following five principles for successful cooperation:

- Openness
- Optimal quality assurance
- Development of innovative functionalities
- Shared leadership
- Sustainability

Work together

- ➔ Databib and re3data.org merged in spring 2015
- ➔ Will become a service of DataCite
- Prevent doing the same work twice, it saves limited resources
- Results and experience gathered by one project member can be shared, leading to a better overall service
- Both, the project and the repositories gain more visibility due to the international character of the collaboration

Technical infrastructure

- Excel
- ZIP-File
- And what is this RDF thing?

Technical infrastructure

- Keep it small & simple
- Supply multiple convenient data formats if possible
- Provide tools and documentation to consume the data

Raise awareness

- Among service units & researchers
- Teach students & young scientists
 - Not during lectures
 - But during projects & thesis's

Clarify responsibilities

Responsible	Responsibility
Scientist	Correctness of Data
Service units	Data management

Thank you for your attention!

WHERE DO YOU STORE
YOUR RESEARCH DATA?

- ☐ USB DRIVE
- ☐ DROPBOX
- ☐ RESEARCH DATA REPOSITORY

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES