Making Research Data Repositories Visible: The re3data.org Registry

Heinz Pampel (heinz.pampel@gfz-potsdam.de), Deutsches GeoForschungsZentrum GFZ, Library and Information Services (LIS), Germany

Paul Vierkant (paul.vierkant@hu-berlin.de), Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Germany

Frank Scholze (frank.scholze@kit.edu), Karlsruhe Institute of Technology (KIT), KIT Library, Germany

Roland Bertelmann (roland.bertelmann@gfz-potsdam.de), Deutsches GeoForschungsZentrum GFZ, Library and Information Services (LIS), Germany

Maxi Kindling (maxi.kindling@hu-berlin.de), Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Germany

Jens Klump (jens.klump@gfz-potsdam.de), Deutsches GeoForschungsZentrum GFZ, Centre for GeoInformation Technology (CeGIT), Germany

Hans-Jürgen Goebelbecker (hans-juergen.goebelbecker@kit.edu), Karlsruhe Institute of Technology (KIT), KIT Library, Germany

Jens Gundlach (jens.gundlach@kit.edu), Karlsruhe Institute of Technology (KIT), KIT Library, Germany

Peter Schirmbacher (schirmbacher@cms.hu-berlin.de), Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Germany

Uwe Dierolf (uwe.dierolf@kit.edu), Karlsruhe Institute of Technology (KIT), KIT Library, Germany

Abstract

Researchers require infrastructures that ensure a maximum of accessibility, stability and reliability to facilitate working with and sharing of research data. Such infrastructures are being increasingly summarized under the term Research Data Repositories (RDR). The project re3data.org – Registry of Research Rata Repositories has begun to index research data repositories in 2012 and offers researchers, funding organizations, libraries and publishers an overview of the heterogeneous research data repository landscape. Information icons help researchers to easily identify an adequate repository for the storage and reuse of their data. This article describes the RDR landscape, outlines the practicality of re3data.org as a service, and shows how this service helps to find research data.

1. Introduction

In the debate on the access to and reuse of research data we currently witness a dynamic development. Already in 2003 leading research organizations worldwide declared the importance of "Open Access to Knowledge in the Sciences and Humanities" and the relevance of research data as an integral part of scholarly knowledge in the Berlin Declaration (Berlin Declaration, 2003). In 2007 the Organization for Economic Co-operation and Development (OECD) published "Principles and Guidelines for Access to Research Data from Public Funding", which "are intended to promote data access and sharing among researchers" (OECD, 2007). These are only two early references of many to follow in a widespread and ongoing debate that concerns diverse stakeholders in the scholarly system.

The Royal Society joined this debate through their notable report "Science as an open enterprise" that was published in 2012. In this report, the Royal Society asks scientists to make their data accessible and usable in the sense of an "intelligent openness": "Where data justify it, scientists should make them available in an appropriate data repository." (The Royal Society, 2012). This statement is echoed on a political level. The European Commission demands from member states that they pass policies to ensure that "research data that result from publicly funded research become publicly accessible, usable and re-usable through digital e-infrastructures." (European Commission, 2012a). The U.S. government went one step further. It obliged its national research agencies to maximize access to digital research data. The Office of Science and Technology Policy (OSTP) directs that "digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze." (OSTP, 2013). The European Commission is planning a similar requirement in their 8th Framework Programme HORIZON 2020 (European Commission, 2012b). Finding a definition of the term research data that is valid for various scholarly disciplines remains a challenge comparable with the challenge to define a research data repository. What is meant by research data differs according to research methods and the character of research objects in the disciplines. Nevertheless, it is important to examine the concept of research data as research data repositories need to serve different academic and disciplinary communities with their respective concepts of research data. Information infrastructure requirements arise from these contents and user requirements.

In the following, the term research data is defined as digital data being a (descriptive) part or the result of a research process. This process covers all stages of research, ranging from research data generation, which may be in an experiment in the sciences, an empirical study in the social sciences or observations of cultural phenomena, to the publication of research results. Digital research data occur in different data types, levels of aggregation and data formats, informed by the research disciplines and their methods. With regards to the purpose of access for use and re-use of research data, digital

research data are of no value without their metadata and proper documentation describing their context and the tools used to create, store, adapt, and analyze them (Kindling & Schirmbacher, 2013).

Data policies increasingly affect scientists and their handling of research data and recommendations and mandates from funders and journals show to be the most effective influence. Data policies require receivers of grants and authors of papers to ensure the accessibility to the data generated within the scope of a project or as the basis of a publication (Pampel & Bertelmann, 2011). As an example, a "Data Sharing Policy" requires National Science Foundation (NSF) applicants "to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants." (National Science Foundation, 2011a) The NSF further requires that measures for the implementation of this policy be specified in a "Data Management Plan" (National Science Foundation, 2011b). The German Research Foundation (DFG) also expects similar statements concerning the handling of research data in project proposals since 2010, whereby, whenever possible, consideration should be given to "existing standards and data repositories" (Deutsche Forschungsgemeinschaft, 2012). A similar requirement can be found in the "Editorial Policies" of scholarly journals publishers, such as Nature Publishing Group: "authors are required to make materials, data and associated protocols promptly available to others without undue qualifications." Accessibility of the research data is supposed to be achieved "via public repositories." (Nature, 2013).

Although scientists agree with the potential benefit of data sharing for the scientific progress, the majority is reserved when it comes to practical implementations (Kuipers & Van der Hoeven, 2009; Tenopir et al., 2011). Incentives, such as proper citation of data, can help to foster the transition (Nature Biotechnology, 2009). The integration of data sharing in scholarly communication offers high potential in this respect.

Research data could be made openly accessible via three publication strategies (Pampel & Dallmeier-Tiessen, in Press):

- Publication of research data, as an independent information object, through a repository (Klump et al., 2006).

- Publication of research data with textual documentation as a so-called data paper (Chavan & Penev, 2011).

- Publication of research data as enrichment of an interpretive text publication ("enriched publication") (Woutersen-Windhouwer & Brandsma, 2009).

These publication strategies have in common that an information infrastructure is required, which ensures storage and accessibility of the data with a maximum of persistence and reliability. Such infrastructures can include so called data archives, data centers, digital libraries, digital collections and the like. They are collectively summarized under the term Research Data Repositories (RDR).

2. Research Data Repositories Landscape

In 2009 the European Commission concluded that: "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." (European Commission, 2009). The Commission characterizes the present landscape of information infrastructures appropriately and clearly describes the need for integrated and homogeneous research services.

RDR, and their corresponding services, are characterized by their content (see section 1). They currently store a wide variety of file formats under different conditions for access and reuse. Compared to the storage of research data, activities concerning the standardization of repositories providing research publications are far more established. The Open Archives Initiative (OAI) was set up early to promote the standardization and networking of institutional or disciplinary repositories providing open access to textual information objects, such as research papers (pre- and post-prints), theses and dissertations (Lagoze & Sompel, 2001). In contrast the RDR community lacks a comparable degree of standardization.

Up to now, only a few studies investigated the global landscape of research data repositories. A study on the characteristics of 100 RDR was published by Marcial & Hemminger in 2010. Schaaf made a similar attempt in 2011. A look at disciplinary studies shows a large diversity among RDR, even on a disciplinary level, with biomedicine offering an impressive number of RDR, thereby shaping today's landscape of research data infrastructures. A closer look at the practices in biomedicine shows that many of its requirements can also be seen in other communities. The 2013 edition of the "Molecular Biology Database Collection" (http://www.oxfordjournals.org/nar/database/a/) of the Nucleic Acids Research journal presents 1512 infrastructures where biomedical research data can be deposited (Fernández-Suárez & Galperin, 2013). 200 of these infrastructures were closely examined in the scope of the European Life Science Infrastructure for Biological Information Project (ELIXIR). These 200 repositories are operated by about 100 institutions with a total staff of at least 350. A community of several hundred thousand scientists uses these RDR. The annual direct cost for these 200 RDR is approximately 30 million euro (ELIXIR, n.d.). In order to guarantee the sustained operation of the biomedical RDR landscape, ELIXIR has been included in the European Strategy Forum on Research Infrastructures (ESFRI). ESFRI is dedicated to the strategic promotion of research infrastructures that are of central importance for the competitiveness of the European Research Area (ERA). It was already clear from the start of ESFRI in 2004 that research infrastructures are not only to be classified as physical infrastructures, such as research ships or particle accelerators, but also as digital information infrastructures, such as, "electronic archiving systems for scientific publications and databases" (European Strategy Forum on Research Infrastructures, 2006).

2.1 Typology of Research Data Repositories

In the following, we present a first attempt to characterize RDR more systematically as institutional, disciplinary, multidisciplinary, and project specific infrastructures (Pampel, Goebelbecker & Vierkant, 2012):

2.1.1 Institutional Research Data Repositories

Institutional research data repositories are run by an institution such as an university or research institution. On university level the scope is multidisciplinary. Edinburgh DataShare (http://datashare.is.ed.ac.uk) is an example of an institutional RDR from the United Kingdom. This "online digital repository of multi-disciplinary research data sets produced at the University of Edinburgh" (Edinburgh DataShare, n.d.) is based on the DSpace software framework and was developed in the period 2007 to 2009 (Rice, 2009). A total of 61 data sets were stored on the repository as of March 2013. Open Data LMU (http://data.ub.uni-muenchen.de) is another example of an institutional RDR from Germany. It was started in 2010 by using the software ePrints and is available for all members of the LMU Munich as a publication platform for research data (Schallehn, 2010). This repository stored 35 data sets as of March 2013.

2.1.2 Disciplinary Research Data Repositories

Prominent examples in the field of disciplinary RDR are GenBank or PANGAEA. GenBank (http://www.ncbi.nlm.nih.gov/genbank) started its service in 1982 (Cravedi, 2008) and defines itself as being a "public database of nucleotide sequences and supporting bibliographic and biological annotation." The National Center for Biotechnology Information (NCBI) operates this infrastructure, providing information on the nucleotide sequences of more than 250,000 species (Benson et al., 2012). PANGAEA – Data Publisher for Earth & Environmental Science (http://www.pangaea.de) is defined as an "[o]pen access library aimed at archiving, publishing and distributing georeferenced data from earth system research." (PANGAEA, n.d.). This RDR is operated by Alfred Wegener Institute for Polar and Marine Research (AWI) and MARUM – Center for Marine Environmental Sciences at the University of Bremen. PANGAEA started with the scope of being a "Paleoclimate Data Center" and was funded by the German Federal Ministry of Education and Research (BMBF) from 1994 to 1997 (Diepenbroek et al., 1999). In 2011, PANGAEA stored "about half a million data sets from all fields of geosciences" (Schindler, Diepenbroek, & Grobe, 2012).

2.1.3 Multidisciplinary Research Data Repositories

Alongside disciplinary and institutional approaches there are also research data repositories serving multidisciplinary needs. Figshare (http://figshare.com) is a research data repository example that „allows researchers to publish all of their data in a citable, searchable and sharable manner." (Figshare, n.d.). Figshare started in 2011 and is operated by Digital Science, a Macmillan Publishers company (Fenner, 2012). A second example is LabArchives (http://www.labarchives.com), a "web-based

electronic notebook software", operated by a private company, that allows scientists "to store, organize, and publish [their] research data" (LabArchives, n.d.).

2.1.4 Project Specific Research Data Repositories

The landscape of RDR with a specific focus on the research data resulting from particular research projects is also diverse. The Scientific Drilling Database (SDDB) (http://www.scientificdrilling.org) operated by GFZ German Research Centre for Geosciences can be named as being exemplary here. It provides drilling data that is created in the scope of the Scientific Continental Drilling Program (ICDP) openly reusable (Klump & Conze, 2007). The RDR of the Bern Digital Pantheon Project (http://www.digitalpantheon.ch/repository) is another example. In this RDR high resolution images and visualizations of the Pantheon in Rome are made freely accessible.

The represented typology attempt shows how heterogeneous the RDR landscape is. Together with a description of potential user needs it is a first step towards a more unified and integrated service description of RDR.

2.2 Need for research data management services and tools

Researchers want to find research data, they want to know how to access the data and they want to know whether if and how reuse is possible or not. Further, researchers want to know about possibilities to store their own data and therefore, information on RDR features, are essential for them. Today, in most academic disciplines it is difficult to gain a comparative overview of existing RDR. In the last years websites like the OAD – Open Access Directory (http://oad.simmons.edu/oadwiki/Data_repositories) or DataCite (http://www.datacite.org/repolist) started to list RDR. However, these and similar lists only provide basic information about a RDR and their services, such as a short description containing the repository operator, discipline and URL. More detailed descriptions and new services are necessary to meet the needs of the research communities. These include tools for the identification of suitable RDR to facilitate discovery and access to research data, including information on the conditions of research data reuse. Besides sharing data, the discussion on research data publishing is strongly gaining momentum. Quickly selecting RDR, supplying which supply persistent identifiers like a Digital Object Identifier (DOI) will be more and more a basic need for researchers looking for a RDR to publish their data.

3. re3data.org – Registry of Research Data Repositories

RDR are expected to be of high relevance for researchers in the near future. The scientific and political demand for Open Science (Kroes, 2012a, 2012b), including open access to publicly funded research data and results, is bound to fail without trustworthy, persistent and sustainable infrastructures that

support researchers to share their research data. Surveys of RDR operators revealed uncertainty about the financial security of these infrastructures for periods longer than five years (Kuipers & Van der Hoeven, 2009; Pfeiffenberger et al., 2012). Thus, the current strategic research development and funding can be considered as being inadequate.

A vision of how research data will be handled in 2030 was described in a study commissioned by the European Commission in 2010. According to this, researchers will then be able "to find, access and process the data they need." In addition to this, researchers gathering data will be able to "deposit their data with confidence in reliable repositories", working on the basis of international standards (High Level Expert Group on Scientific Data, 2010). A glance at the present heterogeneous RDR landscape shows that the realization of this vision is a central challenge for the scholarly system.

Against the background of the growing demand for data sharing (see section 1) and the heterogeneous landscape of RDR (see section 2), the re3data.org – Registry of Research Data Repositories initiative (http://re3data.org) aims to develop and operate a directory of RDR. The indexed and structured description of RDR of all domains in a web-based registry is the target of this project. Added value is created by adding to these descriptions a quick and easy to use system of information icons that describes elementary features of each RDR.

Project partners in re3data.org are the Library and Information Services department (LIS) of the GFZ German Research Centre for Geosciences, the Berlin School of Library and Information Science at the Humboldt-Universität zu Berlin and the KIT Library at the Karlsruhe Institute of Technology (KIT). The three project partners have a long standing working relationship with the German Initiative for Network Information (DINI). Under the auspices of DINI a policy paper on research data was already published in 2009 (Dallmeier-Tiessen et al., 2009). The first phase of the project from January 2012 to December 2013 is funded by the German Research Foundation (DFG).

The main goal of re3data.org is to offer researchers orientation in the heterogeneous landscape of RDR, both in their role as data producers and as data users. Other target groups are research funders and infrastructure facilities, such as data centers and academic libraries. Furthermore, re3data.org aims to make a contribution to establishing a more coherent and integrated "eco system of data repositories" (Van der Graaf & Waaijers, 2011). The registry portraits the development of the worldwide RDR scene. This global overview could, for example, be used to identify disciplines in which the RDR landscape is still underdeveloped.

When the project re3data.org started only a few lists of RDR existed and listed only basic information, such as the name of the repository, its operator, and its disciplinary focus. As a first step the project collected and recorded information of approximately 400 infrastructures storing research data by December 2012. Already existing lists of repositories, such as the lists of the OAD – Open Access

Directory were used alongside our own investigations. All three project partners independently examined a subset of twenty randomly selected RDR. This first analysis confirmed the impression of an extremely heterogeneous RDR landscape and served as a basis for the creation of a first draft of a schema to describe RDR. The absence of a suitable schema required us to develop a new metadata schema to describe RDR. In a second step the schema was aligned with similar metadata schemes, vocabulary elements were modified, and basic requirements for RDR introduced.

In July 2012 Version 1.0 of the vocabulary for the description of RDR was published together with a documentation (Vierkant et al., 2012a). To ensure transparency in the development of the vocabulary, as well as to gain input and acceptance by the community of RDR operators, comments on the vocabulary were not only requested on the project website, but also via emails to various mailing lists. The feedback was very positive and in some cases very elaborate. The project received feedback from reBIND (http://rebind.bgbm.org), DataCite (http://www.datacite.org), and OpenAIREplus (http://www.openaire.eu), among others. All comments were analyzed and discussed by the project team and the suggestions taken into consideration in the revision of the vocabulary leading to version 2.0 of the vocabulary which was published in December 2012 (Vierkant et al., 2012b). The vocabulary covers the following aspects: general information, responsibilities, policies, legal aspects, technical standards, metadata standards, quality standards, and services (see Figure 1).

The current requirements for certification and auditing procedures for RDR were also examined (Braun et al., 2011; Consultative Committee for Space Data Systems, 2011; Data Seal of Approval, 2010; ESF & EUROHORCs, 2011; ICSU World Data System, 2011). It became clear that many of these requirements are not universally applicable to RDR due to the heterogeneous needs in different academic communities and a general lack of RDR standardization. Consequently, re3data.org defined a low entry barrier for RDR to included in the registry. For a repository to be indexed in re3data.org, details on the access to and licensing of the research data are indispensable. If these basic requirements are met the RDR will be indexed and reviewed in autopsy by the project team.

A set of icons has been developed to show the main characteristics of a repository (see Figure 2). This icon system helps users to identify a suitable repository for the storage of their data. In re3data.org researchers can clearly see the terms of access and use of each RDR and other characteristics. The icons and their meaning are explained on the website (http://www.re3data.org/faq).

The icons create an incentive for RDR operating institutions to register their services on re3data.org. However, the icon system is not only useful to the researcher but also to the operators of RDR, allowing them to compare repositories and identify strengths and weaknesses of their own infrastructures. This makes re3data.org also a tool to follow trends in the development of RDRs.

The search on re3data.org is designed with one simple search field (Figure 3). Each entry in the list of search results includes the name of the RDR, the subjects covered by the repository, a brief description of the infrastructure followed by a set of icons (Figure 4), including whether it has already been reviewed by re3data.org.

Filters can be used to narrow down the search. The following search facets can be used: subjects, content type, country, certification, open access, persistent identifier and review status. A comprehensive view of the respective RDR entry can be obtained by clicking on the name of the repository (Figure 5).

RDR operators can suggest their infrastructures to be listed in re3data.org via a simple application form. The project team reviews and lists the proposed repositories in the directory. A repository is indexed when the minimum requirements are met, meaning that mode of access to the data and repository as well as the terms of use must be clear explained on the repository pages. Due to the diverging structure of RDR websites the indexing process is time-consuming. Only very few RDR have policies containing essential information on its services, the designated community and the terms of use and in some cases the operators of RDR have to be contacted to obtain this information. Based on these first lessons learned, the workflows will be optimized and ways of obtaining feedback from RDR operators improved.

4. Outlook

With her statement "We start the era of open science" Neelie Kroes, EU commissioner for the Digital Agenda (http://ec.europa.eu/digital-agenda), shows that openness will be the paradigm of digital science (Kroes, 2012a). The promotion of this development will require a permanent information infrastructure that supports scientists in the sharing of their research data and guarantees access and reuse of research data for the next-generation of researchers.

To provide a persistent registry all re3data.org project partners guarantee the long-term operation of the registry. Based on the feedback from stakeholders re3data.org will go on to develop new features and services in the realm of research data management. Against this setting, a Memorandum of Understanding with DataCite was signed in Spring 2012. DataCite, the initiative for persistent identification of research data, is the result of a project on data publication funded by the German Research Foundation (DFG) in which one re3data.org partner was a consortium member (Klump et al., 2006). In the scope of this cooperation, the flow of information between the two partners is particularly important. Consultation with related initiatives like Databib (http://databib.org) is ongoing. In addition to the technical and structural development of the registry, re3data.org and its project partners will continue to contribute to a closer integration and bigger coherence of RDR. The

project has, for instance, co-organized both the Research Data Infrastructure Symposium 2013 (Schäfer et al. 2013) in Germany and a workshop for librarians on RDR at the 5th Congress Library and Information in Germany in March 2013.

Although re3data.org is still in its starting phase, as of April 2013, 338 RDR have already been indexed in re3data.org out of which 171 have been reviewed. In the next project phase the focus will be on improving the usability and implementing new features. Beyond the development of the registry, the project is dedicated to the networking and standardization of research data depositories. The project strives to make all metadata in the registry available for open use under the Creative Commons license CC0. In doing so, re3data.org contributes to the challenging path to Open Science.
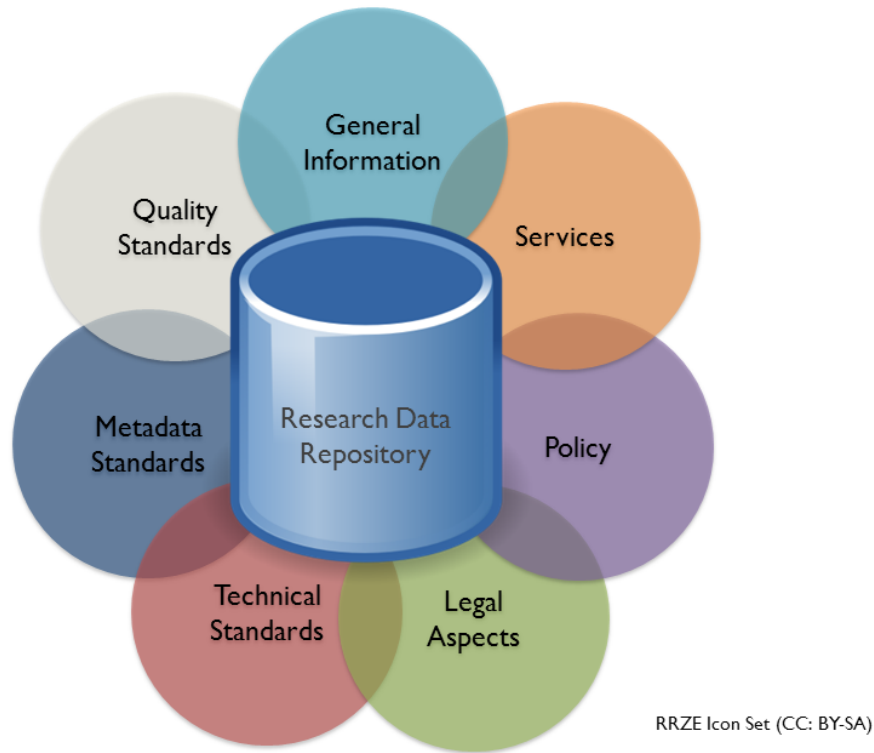
Figures

Figure 1: The aspects of a Research Data Repository documented in the re3data.org vocabulary.

Possible values for each icon

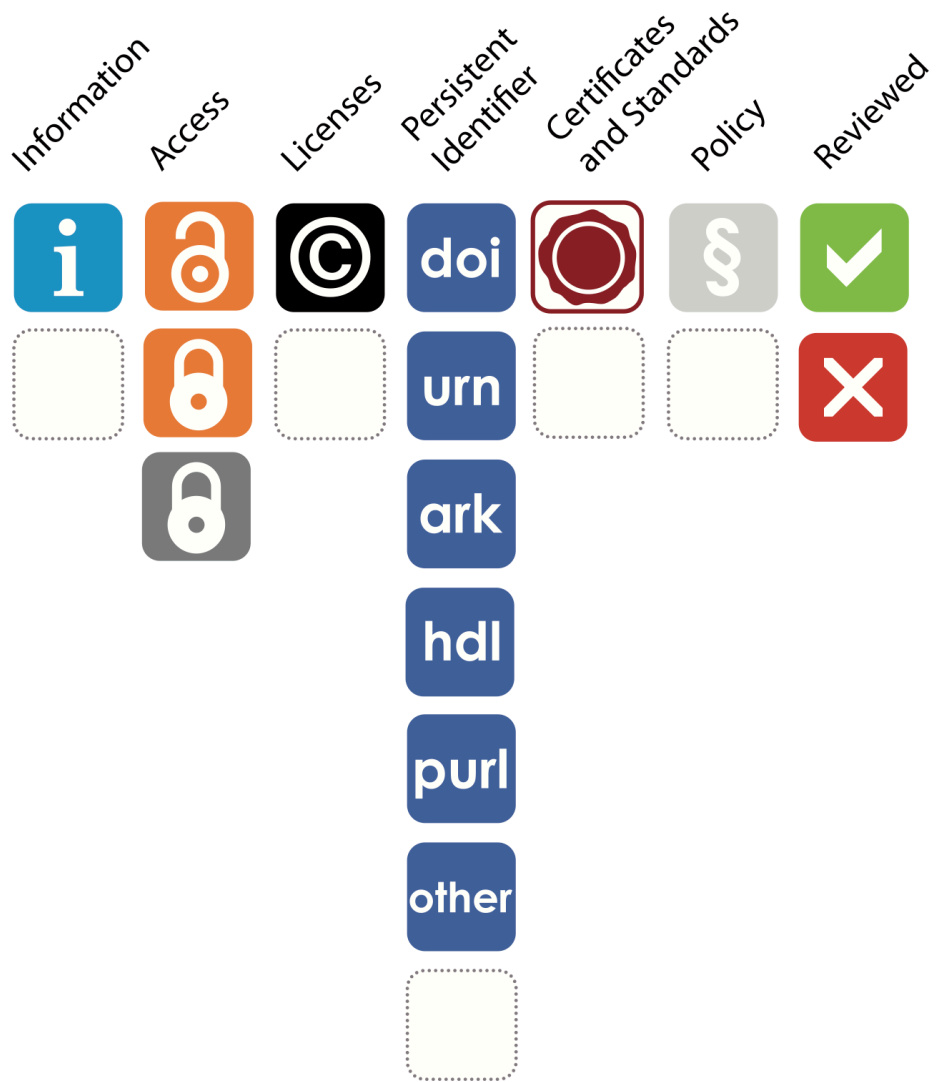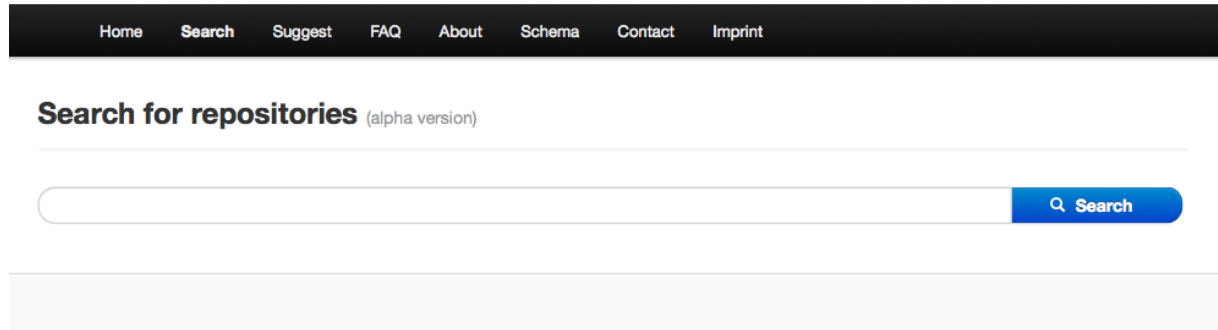**Figure 2: The re3data.org icon system.**

Figure 3: The home page of the re3data.org search.

Figure 4: The hit list of a search.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

| Home | Search | Suggest | FAQ | About | Schema | Contact | Imprint |

Repository details

PANGAEA

Back to results

General information

Name of repository	**PANGAEA**
Additional name	Publishing Network for Geoscientific and Environmental Data
Repository URL	http://www.pangaea.de
Subjects	Biology Geochemistry Geodesy Geoinformatics Geology and Palaeontology Geophysics Mineralogy and Crystallography Oceanography Remote Sensing
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	disciplinary
Research data repository language(s)	eng
Data and/or service provider	dataProvider

Responsible institutions (2)

Institution name	**Alfred Wegner Institute for Polar and Marine Research**
Additional name	AWI
URL	http://www.awi.de/
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

Policies (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

Technical standards

Name of the repository software	other
Versioning	yes

Metadata standards

Persistent identifier system	DOI
Data citation guideline	http://wiki.pangaea.de/wiki/Citation

Quality standards

Quality management	yes
Certificates and Standards	WDS

Alerting services (1)

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

Versioning

Entry date	2012-07-16
Last update	2012-07-16

Back to results

Figure 3: A detailed description of a Research Data Repository.

References

Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, *40*(Database issue), D48–D53. doi:10.1093/nar/gkr1202

Berlin Declaration. (2003). Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Retrieved October 27, 2009, from http://oa.mpg.de/openaccess-berlin/berlindeclaration.html

Braun, K., Buddenbohm, S., Dobratz, S., Herb, U., Müller, U., Pampel, H., Schmidt, B., et al. (2011). *DINI Certificate Document and Publication Services 2010* (3.0 ed.). Göttingen. doi:urn:nbn:de:kobv:11-100182800

Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, *12*(Suppl 15), S2. doi:10.1186/1471-2105-12-S15-S2

Consultative Committee for Space Data Systems (CCSDS). (2011). *Audit and Certification of Trustworthy Digital Repositories. Recommended Practice. CCSDS 652.0-M-1. Magenta Book. Practice*. Retrieved from http://public.ccsds.org/publications/archive/652x0m1.pdf

Cravedi, K. (2008). *GenBank Celebrates 25 Years of Service with Two-Day Conference. Leading Scientists Will Discuss the DNA Database at April 7-8 Meeting*. Retrieved from http://www.nih.gov/news/health/apr2008/nlm-03.htm

Dallmeier-Tiessen, S., Dobratz, S., Gradmann, S., Horstmann, W., Kleiner, E., Pampel, H., Schirmbacher, P., et al. (2009). *Positionspapier Forschungsdaten* (1.0 ed.). doi:urn:nbn:de:kobv:11-10098082

Data Seal of Approval. (2010). *Data Seal of Approval. Guidelines version 1*. Retrieved from http://assessment.datasealofapproval.org/documentation/

Deutsche Forschungsgemeinschaft. (2012). *Proposal Preparation Instructions. DFG form 54.01. 2012*. Retrieved from http://www.dfg.de/formulare/54_01/54_01_en.pdf

Diepenbroek, M., Grobe, H., Reinke, M., Schlitzer, R., & Sieger, R. (1999). Data management of proxy parameters with PANGAEA. In G. Fischer & G. Wefer (Eds.), *Use of Proxies in Paleoceanography. Examples from the South Atlantic* (pp. 715–727). Berlin, Heidelberg & New York: Springer-Verlag. Retrieved from http://hdl.handle.net/10013/epic.11224

Edinburgh DataShare. (n.d.). *What is Edinburgh DataShare?* (pp. 6–8). Retrieved from http://datashare.is.ed.ac.uk

ELIXIR. (n.d.). *The ELIXIR Strategy for Data Resources. Draft Report from Workpackage 2. The ELIXIR Preparatory Phase*. Retrieved from http://www.elixir-europe.org/bcms/elixir/Documents/reports/ELIXIR Strategy for Data Resources report.pdf

ESF & EUROHORCs. (2011). *Basic Requirements for Research Infrastructures in Europe*. Retrieved from http://www.dfg.de/download/pdf/foerderung/programme/wgi/basic_requirements_research_infra structures.pdf

European Commission. (2009). ICT infrastructures for e-science. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM(2009) 108 final. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF

European Commission. (2012a). *Commission Recommendation on access to and preservation of scientific information. C(2012) 4890 final*. Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

European Commission. (2012b). *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions Towards Better Access To Scientific Information: Boosting The Benefits Of Public Investments In Research*. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52012DC0401:EN:NOT

European Strategy Forum on Research Infrastructures. (2006). *European Roadmap for Research Infrastructures. Report 2006*. Retrieved from

ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_200 6_en.pdf

Fenner, M. (2012). Figshare: Interview with Mark Hahnel. *Gobbledygook*. Retrieved from http://blogs.plos.org/mfenner/2012/02/16/figshare-interview-with-mark-hahnel/

Fernández-Suárez, X. M., & Galperin, M. Y. (2013). The 2013 Nucleic Acids Research Database Issue and the online molecular biology database collection. *Nucleic Acids Research*, *41*(Database issue), D1–7. doi:10.1093/nar/gks1297

Figshare. (n.d.). *FAQs*. Retrieved from http://figshare.com/faqs

High Level Expert Group on Scientific Data. (2010). *Riding the wave. How Europe can gain from the rising tide of scientific data*. Retrieved from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

ICSU World Data System (WDS). (2011). *Certification of World Data System Members Introduction*. Retrieved from http://icsu-wds.org/images/files/Certification_summary_6_Jul_2011.pdf

Kindling, M. & Schirmbacher, P. (2013). "Die digitale Forschungswelt" als Gegenstand der Forschung. Lehrstuhl Informationsmanagement. (Research on Digital Research. The information management department.) *Information. Wissenschaft & Praxis* 2-3 (64), 127–136. doi: 10.1515/iwp-2013-0017

Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., Lautenschlager, M., et al. (2006). Data publication in the open access initiative. *Data Science Journal*, *5*, 79–83. doi:10.2481/dsj.5.79

Klump, J., & Conze, R. (2007). The Scientific Drilling Database (SDDB) - Data from Deep Earth Monitoring and Sounding. *Scientific Drilling*, (4), 30–31. doi:10.2204/iodp.sd.4.06.2007

Kroes, N. (2012a). *Opening Science Through e-Infrastructures. SPEECH/12/258*. Retrieved from europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258

Kroes, N. (2012b). *Making Open Access a reality for Science. SPEECH/12/392*. Retrieved from http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/392&format=HTML&age d=0&language=EN&guiLanguage=en

Kuipers, T., & Van der Hoeven, J. (2009). *Insight into digital preservation of research output in Europe. Survey Report*. Retrieved from http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

LabArchives. (n.d.). *Electronic Notebook Software for Laboratory Notebooks FAQ's*. Retrieved from http://www.labarchives.com/faqs.php

Lagoze, C., & Sompel, H. Van De. (2001). The Open Archives Initiative: Building a Low-Barrier Interoperability Framework. *First ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'01)* (pp. 54–62). Roanoke, Virginia. doi:10.1109/JCDL.2001.10061

Marcial, L. H., & Hemminger, B. M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology*, *61*(10), 2029–2048. doi:10.1002/asi.21339

National Science Foundation. (2011a). *Proposal and Award Policies and Procedures Guide. Chapter VI - Other Post Award Requirements and Considerations*. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

National Science Foundation. (2011b). *Proposal and Award Policies and Procedures Guide. Grant Proposal Guide. Chapter II - Proposal Preparation Instructions*. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

Nature. (2013). *Availability of data and materials*. Retrieved from http://www.nature.com/authors/policies/availability.html

Nature Biotechnology. (2009). Credit where credit is overdue. *Nature Biotechnology*, *27*(7), 579. doi:10.1038/nbt0709-579

Office of Science and Technology Policy. (2013). *Increasing Access to the Results of Federally Funded Scientific Research*. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.p df

Organisation for Economic Co-operation and Development. (2007). *Principles and Guidelines for Access to Research Data from Public Funding* (p. 24). Paris. Retrieved from Principles and Guidelines for Access to Research Data from Public Funding

Pampel, H., & Bertelmann, R. (2011). „Data Policies" im Spannungsfeld zwischen Empfehlung und Verpflichtung. In S. Büttner, H.-C. Hobohm, & L. Müller (Eds.), *Handbuch Forschungsdatenmanagement* (pp. 49–61). Bad Honnef: Bock + Herchen. doi:urn:nbn:de:kobv:525-opus-2287

Pampel, H., & Dallmeier-Tiessen, S. (in Press). Open Research Data - From Vision to Practice. In S. Bartling & S. Friesike (Eds.), *Opening Science – The Evolving Guide On How The Internet Is Changing Research, Collaboration and Scholarly Publishing*. Heidelberg: Springer.

Pampel, H., Goebelbecker, H.-J., & Vierkant, P. (2012). re3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositorien. Ein Werkstattbericht. In B. Mittermaier (Ed.), *Vernetztes Wissen – Daten, Menschen, Systeme. WissKom 2012* (pp. 61–73). Jülich: Verlag des Forschungszentrums Jülich. Retrieved from http://hdl.handle.net/2128/4699

PANGAEA. (n.d.). About / Imprint. Retrieved July 29, 2011, from http://www.pangaea.de/about/

Pfeiffenberger, H., Pampel, H., Schäfer, A., Guidetti, V., Bruch, C., Tzitzikas, Y., Pröll, S., et al. (2012). *Report and Strategy on Annotation, Reputation and Data Quality*. Retrieved from http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=D26.1+Report+and+Strategy+on+Annotation%2C+Reputation+and+Data+Quality

Rice, R. (2009). *DISC-UK DataShare*. Retrieved from http://ie-repository.jisc.ac.uk/336/

Schaaf, I. (2011). *Forschungsdaten im Netz – Untersuchung von online verfügbaren Repositorien*. Hochschule der Medien Stuttgart.

Schäfer, L., Pampel, H., Klump, J., & Häsler, T. (2013). *Bericht Symposium "Forschungsdaten-Infrastrukturen (FDI 2013)". Workshop Report*. doi:10.2312/radieschen_003

Schallehn, V. (2010, October 13). Open Data LMU. *Universitätsbibliothek München*. Retrieved from http://www.ub.uni-muenchen.de/no_cache/aktuelles/einzelne-nachricht/article/open-data-lmu/

Schindler, U., Diepenbroek, M., & Grobe, H. (2012). PANGAEA - Research Data enters Scholarly Communication. *Geophysical Research Abstracts*, *14*, EGU2012–13378–1. Retrieved from http://meetingorganizer.copernicus.org/EGU2012/EGU2012-13378-1.pdf

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., et al. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, *6*(6), e21101. doi:10.1371/journal.pone.0021101

The Royal Society. (2012). *Science as an open enterprise. The Royal Society Science Policy Centre report 02/12*. Retrieved from http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

Van der Graaf, M., & Waaijers, L. (2011). *A Surfboard for Riding the Wave. Towards a four country action programme on research data*. Retrieved from http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=/Files/Filer/downloads/Primary+Research+Data/Surfboard+for+Riding+the+Wave/KE_Surfboard_Riding_the_Wave_Screen.pdf

Vierkant, P., Spier, S., Rücknagel, J., Gundlach, J., Fichtmüller, D., Pampel, H., Kindling, M., et al. (2012). *Vocabulary for the Registration and Description of Research Data Repositories. Version 2.0*. Retrieved from http://dx.doi.org/10.2312/re3.002

Vierkant, P., Spier, S., Rücknagel, J., Gundlach, J., Kindling, M., Pampel, H., Manova, S., et al. (2012). *Vocabulary for the Registration and Description of Research Data Repositories. Version 1.0*. doi:10.2312/re3.001

Woutersen-Windhouwer, S., & Brandsma, R. (2009). Enhanced Publications, State of the Art. In M. Vernooy-Gerritsen (Ed.), *Enhanced Publications. Linking Publications and Research Data in Digital Repositories*. Amsterdam: Amsterdam University Press. Retrieved from http://dare.uva.nl/aup/nl/record/316849