Making Research Data Repositories Visible: The re3data.org Registry

Heinz Pampel[1]*, Paul Vierkant[2], Frank Scholze[3], Roland Bertelmann[1], Maxi Kindling[2], Jens Klump[1], Hans-Jürgen Goebelbecker[3], Jens Gundlach[3], Peter Schirmbacher[2], Uwe Dierolf[3]

1 Deutsches GeoForschungsZentrum GFZ, Library and Information Services (LIS), Potsdam, Germany, 2 Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Berlin, Germany, 3 Karlsruhe Institute of Technology (KIT), KIT Library, Karlsruhe, 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 Data 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. In July 2013 re3data.org lists 400 research data repositories and counting. 288 of these are described in detail using the re3data.org vocabulary. Information icons help researchers to easily identify an adequate repository for the storage and reuse of their data. This article describes the heterogeneous RDR landscape and presents a typology of institutional, disciplinary, multidisciplinary and project-specific RDR. Further the article outlines the features of re3data.org, and shows how this registry helps to identify appropriate repositories for storage and search of research data.

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 [1]. 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" [2]. 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" [3]. 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" [4]. 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" [5]. The European Commission is planning a similar requirement in their 8th Framework Programme HORIZON 2020 [6]. 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 [7].

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 [8]. 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." [9] The NSF further requires that measures for the implementation of this policy be specified in a "Data Management Plan" [10]. 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" [11]. 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" [12].

Although scientists agree with the potential benefit of data sharing for the scientific progress, the majority is still reserved when it comes to practical implementations [13,14]. Incentives, such as proper citation of data, can help to foster the transition [15]. 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 [16]:

- Publication of research data, as an independent information object, through a repository [17].
- Publication of research data with textual documentation as a so-called data paper [18].
- Publication of research data as enrichment of an interpretive text publication ("enriched publication") [19].

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).

Up to now there was no comprehensive overview of this infrastructures and their functionalities available. The project re3data.org – Registry of Research Data Repositories change this adverse situation. The project has begun to index research data repositories in 2012 and offers researchers, funding organizations, libraries and publishers a systematic overview of the heterogeneous RDR landscape. In July 2013 re3data.org lists 400 research data repositories. 288 of these are described in detail by a special vocabulary, which was developed in the re3data.org project. In the following we give an overview on the RDR landscape (section 2). Further, we describe the development of the registry, the features of the re3data.org registry and explain how the registry helps to identify appropriate repositories for storage and search of research data (section 3).

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" [20]. 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-
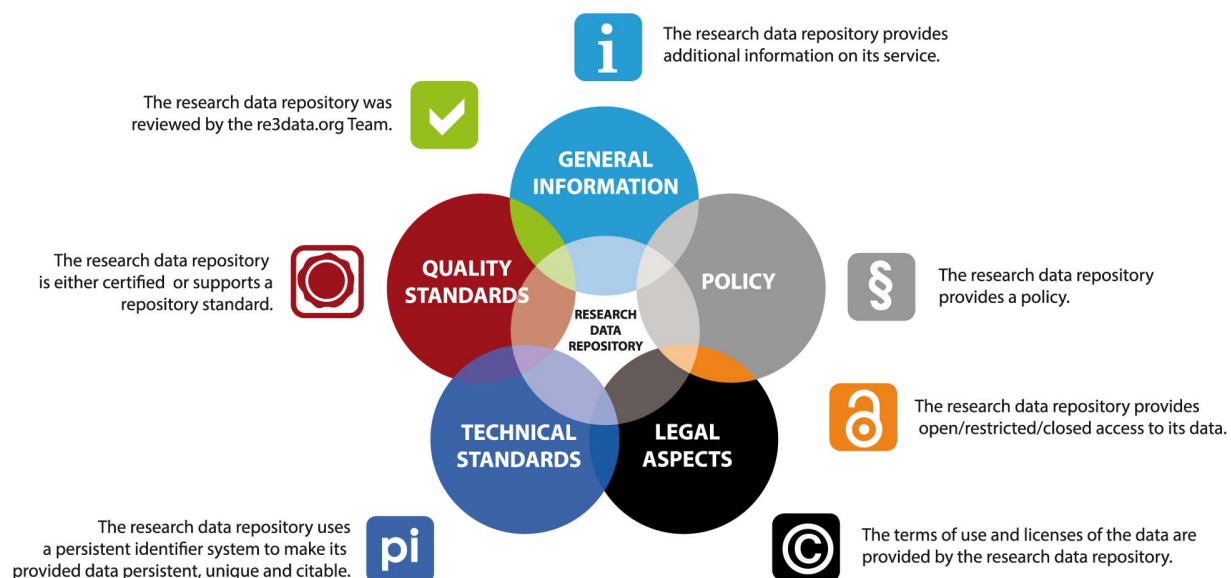
Figure 1. Aspects of a Research Data Repository with the corresponding icons used in re3data.org.
doi:10.1371/journal.pone.0078080.g001
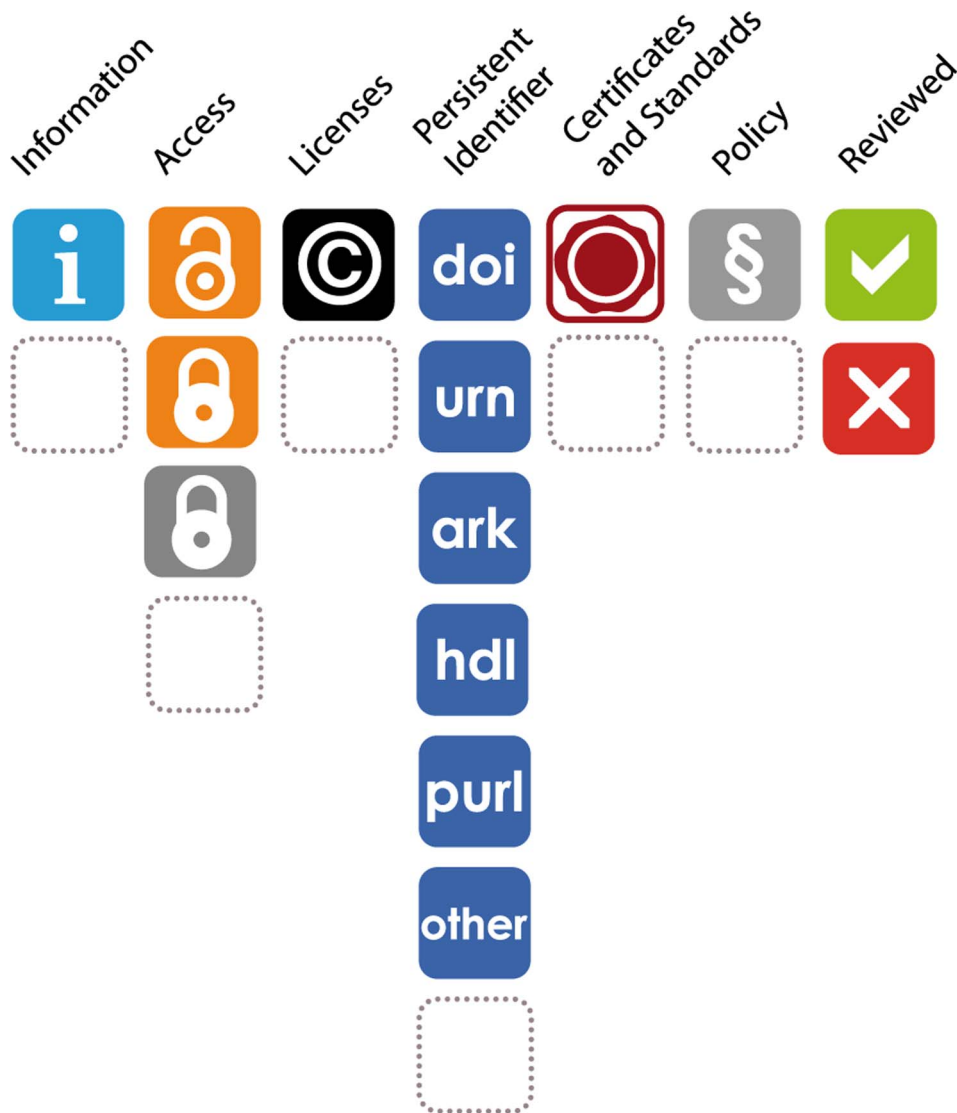
Figure 2. The re3data.org icon system depicting all possible values for each icon.
doi:10.1371/journal.pone.0078080.g002

prints), theses and dissertations [21]. 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 [22]. Schaaf made a similar attempt in 2011 [23]. 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 [24]. 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 [25]. 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" [26].

2.1 Typology of Research Data Repositories

In the following, we present a typology of RDR. This systematization was developed on the basis of a first analysis of 400 RDR (see section 3). Based on the widespread distinction between institutional and disciplinary repository for scholarly literature [27] we differentiate between institutional, disciplinary,

Figure 3. The home page of the re3data.org search.
doi:10.1371/journal.pone.0078080.g003

multidisciplinary, and project specific RDR [28]. The following description of RDR outlines the differences between the four repository types. This systematization helps to get an overview of the varying concepts and strategies of infrastructures for the permanent access and re-use of research data.

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 multidisciplinary research data sets produced at the University of Edinburgh" [29] is based on the DSpace software framework and was developed in the period 2007 to 2009 [30]. 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 [31]. 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 [32] 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

Figure 4. The hit list of a search.
doi:10.1371/journal.pone.0078080.g004

species [33]. 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" [34]. 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 [35]. In 2011, PANGAEA stored "about half a million data sets from all fields of geosciences" [36].

Figure 5. A detailed description of a Research Data Repository.
doi:10.1371/journal.pone.0078080.g005

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." [37] Figshare started in 2011 and is operated by Digital Science, a Macmillan Publishers company [38]. 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" [39].

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 [40]. 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.

Due to the heterogeneous RDR landscape, which is outlined here, it is often difficult for scholars to identify appropriate repositories for the storage of and access to research data.

2.2 Need for Research Data Management Services and Tools

Studies on the researchers perspective showing a broad variety of obstacles that affect scholars' willingness to share their own data. The comprehensive studies by Kuipers & Van der Hoeven [13],

Tenopir et al. [14] and the analysis of the ODE project [41] show that the willingness to share data is often strong related to the need of a supportive research data infrastructure. Repositories, which are embedded in scholarly workflows and associated with incentive systems can help to promote data sharing. Tenopir et al., for example, conclude that in a survey of more than 1300 scientists, that "a majority of respondents in almost all disciplines […] would be willing to place at least some of their data into a central data repository with no restrictions". One of the obstacle to this willingness is the lack of knowledge by scholars on already existing RDR.

At this point re3data.org comes into action. Today, in most academic disciplines it is difficult to gain a comparative overview of existing RDR. Already existing registries like the OpenDOAR – Directory of Open Access Repositories (http://www.opendoar.org) and the ROAR – Registry of Open Access Repositories (http://roar.eprints.org) only contain a small share of research data repositories (<5%) since both registries focus on repositories for scholarly publications. 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, all of the mentioned directories only provide rudimentary information about a RDR and their services, such as a short description containing the repository operator, discipline and URL. To overcome the obstacles mentioned in the user surveys [13,14,41], it is necessary to offer researchers, funding organizations, libraries and publishers a systematic and easy-to-use overview of RDR. This means, that in contrast to the already existing directories a more detailed description of a RDR is needed if a registry wants to deliver essential information on access to or conditions of reuse of research data. This description would take into account the disciplinary focus of researchers looking for a place to deposit their research data. It would also take into account the researcher's provisions by providing information on for how long this RDR has been online, how it is funded, does the RDR have a policy and finally who is responsible for the RDR. All of this information is necessary to illustrate the trustworthiness of a RDR.

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 [42,43], 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 [13,44]. 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 [45]. 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 [46]. 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" [47]. The registry describes 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 [48]. 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 [49]. The vocabulary covers the following aspects:

- 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. there 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),

The current requirements for certification and auditing procedures for RDR were also examined [50–54]. 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 by the project team. A detailed description of the requirements can be found in the re3data.org vocabulary [49].

A set of icons has been developed to show the main characteristics of a repository (Figure 1). The record of an RDR in re3data.org shows up to seven icons (with the respective value) according to the information gathered (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 in figure 1 and 2 and also on the website (http://www.re3data.org/faq/).

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. Furthermore the icons create an incentive for RDR operating institutions to keep the information about there repository up to date on re3data.org. Our first experiences have shown, that the operators are very interested in a correct and detailed representation of there repositories in re3data.org. Especially the clear presentation of the repositories features via the icon system seems to be an important reason for the RDR operators to keep the information on there infrastructure up to date.

re3data.org offers two search possibilities: On the one hand a free text search via a simple search box. In this free text search any terms can be searched in the following data base fields: title, additional title, description, keyword, subject, institution, remarks extern (Figure 3). On the other hand the following filters can be used to narrow down the search: subjects, content type, country, certification, open access, persistent identifier and review status. (Figure 4), In the list of results each record includes the name of the RDR, the subjects covered, a brief description of the content and a set of icons visualizing key issues as well as the review status. A comprehensive view of the respective RDR record 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 amount of metadata to be collected and 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. Before a new record of a RDR is published in re3data.org all gathered information is reviewed by a second team member. This second step takes at least half as much time as the indexing, however in terms of quality assurance it is absolutely necessary. Based on these first lessons learned, the workflows will be optimized and ways of obtaining feedback from RDR operators improved. Technically a proper workflow management system will cover all steps from ingest to publication of a RDR record (including a persistent identifier).

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 [42]. 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 [17]. 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.

Although re3data.org is still in its starting phase, as of July 2013, 400 RDR have already been indexed in re3data.org out of which 288 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.

Acknowledgments

Author Contributions

Wrote the paper: HP PV FS RB MK JK HJG JG PS UD. Conceived and implemented the registry: HP PV FS RB MK JK HJG JG PS UD.

References

1. Berlin Declaration (2003) Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Available: http://oa.mpg.de/openaccess-berlin/berlindeclaration.html. Accessed 27 October 2009.
2. Organisation for Economic Co-operation and Development (OECD) (2007) Principles and Guidelines for Access to Research Data from Public Funding.

Paris. Available: http://www.oecd.org/dataoecd/9/61/38500813.pdf. Accessed 16 May 2013.
3. The Royal Society (2012) Science as an open enterprise. The Royal Society Science Policy Centre report 02/12. Available: http://royalsociety.org/

uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf. Accessed 16 May 2013.

4. European Commission (2012) Commission Recommendation on access to and preservation of scientific information. C(2012) 4890 final. Available: http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf. Accessed 16 May 2013.

5. Office of Science and Technology Policy (2013) Increasing Access to the Results of Federally Funded Scientific Research. Available: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf. Accessed 16 May 2013.

6. European Commission (2012) Communication from the Commission to the European Parliament, 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. COM(2012). Available: http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf. Accessed 16 May 2013.

7. Kindling M, Schirmbacher P (2013) "Die digitale Forschungswelt" als Gegenstand der Forschung. Information - Wissenschaft & Praxis 64: 127–136. Available: http://dx.doi.org/10.1515/iwp-2013-0017.

8. Pampel H, Bertelmann R (2011) "Data Policies" im Spannungsfeld zwischen Empfehlung und Verpflichtung. In: Büttner S, Hobohm H-C, Müller L, editors. Handbuch Forschungsdatenmanagement. Bad Honnef: Bock+Herchen. 49–61. Available: http://nbn-resolving.de/urn:nbn:de:kobv:525-opus-2287. Accessed 16 May 2013.

9. National Science Foundation (2011) Proposal and Award Policies and Procedures Guide. Chapter VI - Other Post Award Requirements and Considerations. Available: http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4. Accessed 18 February 2011. Accessed 16 May 2013.

10. National Science Foundation (2011) Proposal and Award Policies and Procedures Guide. Grant Proposal Guide. Chapter II - Proposal Preparation Instructions. Available: http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp. Accessed 18 February 2011.

11. Deutsche Forschungsgemeinschaft (2012) Proposal Preparation Instructions. DFG form 54.01. Available: http://www.dfg.de/formulare/54_01/54_01_en.pdf. Accessed 16 May 2013.

12. Nature (2013) Availability of data and materials. Available: http://www.nature.com/authors/policies/availability.html. Accessed 20 Mai 2013.

13. Kuipers T, Van der Hoeven J (2009) Insight into digital preservation of research output in Europe. Survey Report. Available: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf. Accessed 16 May 2013.

14. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) Data sharing by scientists: practices and perceptions. PLoS ONE 6: e21101. Available: http://dx.doi.org/10.1371/journal.pone.0021101.

15. Nature Biotechnology (2009) Credit where credit is overdue. Nature Biotechnology 27: 579. Available: http://dx.doi.org/10.1038/nbt0709-579.

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

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

18. Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12: S2. Available: http://www.biomedcentral.com/1471-2105/12/S15/S2.

19. Woutersen-Windhouwer S, Brandsma R, Hogenaar A, Hoogerwerf M, Doorenbosch P, et al. (2009) Enhanced Publications?: Linking Publications and Research Data in Digital Repositories. Vernooy-Gerritsen M, editor Amsterdam: Amsterdam University Press. Available: http://dare.uva.nl/aup/nl/record/316849. Accessed 29 August 2012.

20. European Commission: ICT Infrastructures for e-Science. COM (2009) 108 final. 05.03.2009 (n.d.). Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF. Accessed 16 May 2013.

21. 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). Roanoke, Virginia. 54–62. Available: http://public.lanl.gov/herbertv/papers/Papers/2001/JCDLlagoze.pdf. Accessed 16 May 2013.

22. Marcial LH, Hemminger BM (2010) Scientific data repositories on the Web: An initial survey. Journal of the American Society for Information Science and Technology 61: 2029–2048. Available: http://www.ils.unc.edu/bmh/pubs/ScientificDataRepositories-JASIST-2010.pdf. Accessed 15 March 2013.

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

24. Fernández-Suárez XM, Galperin MY (2012) The 2013 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. Nucleic acids research: 1–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/23203983. Accessed 4 December 2012.

25. ELIXIR (n.d.) The ELIXIR Strategy for Data Resources. Draft Report from Workpackage 2. The ELIXIR Preparatory Phase. Available: http://www.elixir-europe.org/prep/sites/elixir-europe.org.prep/files/documents/reports/elixir_strategy_for_data_resources_report.pdf. Accessed 15 March 2013.

26. European Strategy Forum on Research Infrastructures (2006) European Roadmap for Research Infrastructures. Report 2006. Available: http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_en.pdf Accessed 15 March 2013.

27. Suber P (2012) Open Access. Cambridge, Massachusetts: The MIT Press. Available: http://mitpress.mit.edu/books/open-access. Accessed 15 March 2013.

28. Pampel H, Goebelbecker H-J, Vierkant P (2012) re3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositorien. Ein Werkstattbericht. In: Mittermaier B, editor. Vernetztes Wissen – Daten, Menschen, Systeme. WissKom 2012. Jülich: Verlag des Forschungszentrums Jülich. 61–73. Available: http://hdl.handle.net/2128/4699. Accessed 15 March 2013.

29. Edinburgh DataShare (n.d.) What is Edinburgh DataShare? Available: http://datashare.is.ed.ac.uk. Accessed 7 June 2012.

30. Rice R (2009) DISC-UK DataShare. Available: http://ie-repository.jisc.ac.uk/336/. Accessed 15 March 2013.

31. Schallehn V (2010) Open Data LMU. Universitätsbibliothek München. Available: http://www.ub.uni-muenchen.de/no_cache/aktuelles/einzelne-nachricht/article/open-data-lmu/. Accessed 2 August 2011.

32. 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. Available: http://www.nih.gov/news/health/apr2008/nlm-03.htm. Accessed 29 February 2012.

33. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, et al. (2012) GenBank. Nucleic Acids Research 40: D48–D53. Available: http://dx.doi.org/10.1093/nar/gkr1202. Accessed 29 February 2012.

34. PANGAEA (n.d.) About/Imprint. Available: http://www.pangaea.de/about/. Accessed 29 July 2011.

35. Diepenbroek M, Grobe H, Reinke M, Schlitzer R, Sieger R (1999) Data management of proxy parameters with PANGAEA. In: Fischer G, Wefer G, editors. Use of Proxies in Paleoceanography. Examples from the South Atlantic. Berlin, Heidelberg & New York: Springer-Verlag. 715–727. Available: http://hdl.handle.net/10013/epic.11224. Accessed 15 March 2013.

36. Schindler U, Diepenbroek M, Grobe H (2012) PANGAEA - Research Data enters Scholarly Communication. Geophysical Research Abstracts 14: EGU2012–13378–1. Available: http://meetingorganizer.copernicus.org/EGU2012/EGU2012-13378-1.pdf. Accessed 15 March 2013.

37. Figshare (n.d.) FAQs. Available: http://figshare.com/faqs. Accessed 21 April 2013.

38. Fenner M (2012) Figshare: Interview with Mark Hahnel. Gobbledygook. Available: http://blogs.plos.org/mfenner/2012/02/16/figshare-interview-with-mark-hahnel/. Accessed 15 March 2013.

39. LabArchives (n.d.) Electronic Notebook Software for Laboratory Notebooks FAQ's. Available: http://www.labarchives.com/faqs.php. Accessed 15 March 2013.

40. Klump J, Conze R (2007) The Scientific Drilling Database (SDDB) - Data from Deep Earth Monitoring and Sounding. Scientific Drilling: 30–31. Available: http://www.iodp.org/images/stories/downloads/sd4_07.pdf#page=32. Accessed 18 May 2012.

41. Dallmeier-Tiessen S, Darby R, Gitmans K, Herterich P, Lambert S, et al. (2012) Summary of the studies, thematic publications and recommendations. Available: http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=Summary+of+the+studies%2C+thematic+publications+and+recommendations. Accessed 15 March 2013.

42. Kroes N (2012) Opening Science Through e-Infrastructures. SPEECH/12/258. Available: http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258. Accessed 15 March 2013.

43. Kroes N (2012) Making Open Access a reality for Science. SPEECH/12/392. Available: http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/392. Accessed 15 March 2013.

44. Pfeiffenberger H, Pampel H, Schäfer A, Guidetti V, Bruch C, et al. (2012) Report and Strategy on Annotation, Reputation and Data Quality. Available: http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=D26.1+Report+and+Strategy+on+Annotation%2C+Reputation+and+Data+Quality. Accessed 15 March 2013.

45. High Level Expert Group on Scientific Data (2010) Riding the wave. How Europe can gain from the rising tide of scientific data. Available: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf. Accessed 15 March 2013.

46. Dallmeier-Tiessen S, Dobratz S, Gradmann S, Horstmann W, Kleiner E, et al. (2009) Positionspapier Forschungsdaten. 1.0 ed. Available: http://nbn-resolving.de/urn:nbn:de:kobv:11-10098082. Accessed 15 March 2013.

47. Graaf M van der, Waaijers L (2011) A surfboard for Riding the Wave. Towards a four country action programme on research data. Available: http://www.knowledge-exchange.info/Default.aspx?ID=469. Accessed 16 November 2011.

48. Vierkant P, Spier S, Rücknagel J, Gundlach J, Kindling M, et al. (2012) Vocabulary for the Registration and Description of Research Data Repositories. Version 1.0. Available: http://dx.doi.org/10.2312/re3.001.

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

50. Braun K, Buddenbohm S, Dobratz S, Herb U, Müller U, et al. (2011) DINI Certificate Document and Publication Services 2010. 3.0 ed. Göttingen.

Available: http://nbn-resolving.de/urn:nbn:de:kobv:11-100182800. Accessed 15 March 2013.

51. Consultative Committee for Space Data Systems (CCSDS) (2011) Audit and Certification of Trustworthy Digital Repositories. Recommended Practice. CCSDS 652.0-M-1. Magenta Book. Available: http://public.ccsds.org/publications/archive/652x0m1.pdf. Accessed 15 March 2013.

52. Data Seal of Approval (2010) Data Seal of Approval. Guidelines version 1. Available: http://assessment.datasealofapproval.org/documentation/. Accessed 15 March 2013.

53. ESF & EUROHORCs (2011) Basic Requirements for Research Infrastructures in Europe. Available: http://www.dfg.de/download/pdf/foerderung/programme/wgi/basic_requirements_research_infrastructures.pdf. Accessed 15 March 2013.

54. ICSU World Data System (WDS) (2011) Certification of World Data System Members Introduction. Available: http://icsu-wds.org/images/files/Certification_summary_6_Jul_2011.pdf. Accessed 03 January 2012.