Presenting Research Data Repositories: The re3data.org Registration Mechanism

Authors: Heinz Pampel, Paul Vierkant, Frank Scholze, Roland Bertelmann, Maxi Kindling, Jens Klump, Hans-Jürgen Goebelbecker, Jens Gundlach, Peter Schirmbacher, Uwe Dierolf (Author), Translated by Gu Liping

Date: 2016-06-04T00:00:00+00:00

Abstract

In the process of investigating research questions and sharing research data, researchers require infrastructure to ensure maximum accessibility, stability, and usability of data. Such infrastructure can be collectively termed as Research Data Repository (RDR). The re3data.org project, initiated in 2012, primarily focuses on the registration of research data repositories and provides researchers, funding agencies, libraries, and publishers with a comprehensive overview of heterogeneous research data repositories. As of July 2013, 400 research data repositories have been registered with re3data.org, among which 288 have adopted re3data.org' s information icons to assist researchers in selecting appropriate repositories for storing and reusing their data. This paper delineates the landscape of heterogeneous RDRs, articulating institutional, disciplinary, interdisciplinary, and project-specific types of RDRs. It provides an in-depth description of re3data.org' s features and how this registry system helps researchers identify repositories suitable for storing and searching research data.

Full Text

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 14473, Germany
[2]Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Berlin 10099, Germany

[3]Karlsruhe Institute of Technology (KIT), KIT Library, Karlsruhe 76131, Germany

Translated by Gu Liping
National Science Library, Chinese Academy of Sciences, Beijing 100190

Abstract

Researchers require infrastructures that ensure maximum accessibility, stability, and availability of data to facilitate working with and sharing research data. Such infrastructures are collectively referred to as Research Data Repositories (RDR). Since its launch in 2012, the re3data.org project has focused on registering research data repositories and providing researchers, funding organizations, libraries, and publishers with a comprehensive overview of heterogeneous research data repositories. As of July 2013, 400 research data repositories have been registered with re3data.org, of which 288 utilize re3data.org' s information icons to help researchers select appropriate repositories for storing and reusing their data. This paper depicts the landscape of heterogeneous RDRs, describing institutional, disciplinary, interdisciplinary, and project-specific repository types. It provides an in-depth description of re3data.org' s features and explains how this registry system assists researchers in identifying suitable repositories for storing and searching research data.

Keywords: Research data, Research data management, Open access, Institutional repository development, Open science, Research data repositories, Information management

Introduction

The sharing and reuse of research data has gained increasing attention in recent years. As early as 2003, major research organizations worldwide began calling for open access to scientific and humanities knowledge, and their Berlin Declaration (2003) recognized research data as part of scholarly knowledge [1]. In 2007, the Organisation for Economic Co-operation and Development (OECD) published Principles and Guidelines for Access to Research Data from Public Funding, which established "promoting access to and sharing of research data among researchers" as its core mission [2]. These two documents represent early frameworks that emerged from extensive and ongoing debates among various stakeholders in the academic system.

In its 2012 report Science as an Open Enterprise, the Royal Society urged scientists to make their research data accessible and usable through "intelligent openness" : "Scientists should store data in appropriate data repositories to enable use and validation of the data" [3]. EU policy responded to this recommendation by seeking to ensure that member states adopt policies guaranteeing "public access to, use, and reuse of research data from publicly funded research results through electronic infrastructures" [4]. The U.S. government

went further, requiring maximum access to digital research data from national research institutions. The White House Office of Science and Technology Policy (OSTP) specified that "digital research data generated through projects funded wholly or partially by the federal government should be stored and made publicly accessible for search, retrieval, and use" [5]. The EU planned to incorporate similar requirements in its 8th Horizon 2020 framework [6]. Implementation of these policies depends on selecting appropriate research data repositories across different disciplines, which necessitates precise definitions of research data terminology.

Perceptions of research data vary according to research methods and objects across different disciplines. However, the concept of research data, like that of research data repositories, must serve the understanding of different academic communities and their information infrastructure needs. We therefore define research data as digital data that are part of or result from the research process. This process covers all stages of research, from data generation to research output, including empirical studies in science and social science or observations of cultural phenomena. Due to disciplinary and methodological differences, digital research data exhibit various data types, integration levels, and formats. For the purpose of accessing and reusing research data, digital data have no value without metadata and descriptive tools that specify how data archives are created, stored, adapted, and analyzed [7].

Data policies require funded researchers and authors to ensure the accessibility of data produced in publications or projects [8]. This affects how scientists and research institutions hold research data and has even greater implications for how funding agencies and journals recommend and enforce relevant policies. For example, the U.S. National Science Foundation' s (NSF) Data Sharing Policy requires 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 measurement of policy implementation through Data Management Plans [10]. Since 2010, the German Research Foundation (DFG) has included similar principles in its project proposals, requiring German researchers to comply with "existing standards and data repositories" to the best of their ability [11]. Similar conditions appear in academic publishers' editorial policies; for instance, Nature Publishing Group (NPG) requires authors to "make materials, data and associated protocols available to others without restriction" [12]. Therefore, open sharing of research data should be realized through public repositories.

Although scientists recognize the common underlying benefits of data sharing for scientific progress, most remain hesitant in practice [13,14]. Incentives can promote change, such as proper citation of data [15]. Consequently, data sharing has significant potential for integration into scholarly communication.

Research data can implement open access through three publication strategies [16]: (1) Independent publication of research data as standalone information

objects distributed by repositories [17]; (2) Publication of research data alongside textual literature, sometimes called data papers [18]; and (3) Publication of research data attached to publications (to enrich them) as supplementary materials that annotate textual content [19]. A common feature of these strategies is the requirement for information infrastructure that ensures permanent preservation and maximum open sharing of data. Such infrastructure is known by various names, including data archives, data centers, digital libraries, digital collections, and others, which we refer to as Research Data Repositories (RDR).

Until now, a comprehensive overview of these infrastructures and their functions has been lacking. The research data repository registry re3data.org has changed this situation. The project began indexing research data repositories in 2012 and provides researchers, funding agencies, libraries, and publishers with a systematic overview of the heterogeneous RDR landscape. In July 2013, the re3data.org registry listed 400 research data repositories, with 288 employing a special icon system developed by re3data.org to describe themselves in detail. The following sections provide a landscape overview of RDRs (see Section 2) and describe the development of the registry, its features on re3data.org, and how this mechanism helps identify repositories suitable for storing and searching research data (see Section 3).

2. Landscape of Research Data Repositories

The European Commission' s 2009 ICT report concluded that "data repositories across Europe are very heterogeneous, but we have a coherent development policy that provides a solid foundation for overcoming this fragmentation and enabling research communities to better manage, use, store, and preserve data" [20]. The Commission emphasized that a clear landscape overview of information infrastructure appropriately expresses the need for integration and assimilation of research data services.

RDRs and their services are characterized by their stored content and provide storage for access and use of various file types in different contexts. However, compared to storing research data, greater attention has been paid to standardizing repositories for research publications. The Open Archives Initiative (OAI) established standards and networks early on to promote institutional or disciplinary repositories for open access to textual information objects such as research papers (pre- or post-publication) and dissertations [21]. Conversely, the RDR community lacks comparable standards.

To date, only a few studies have investigated the global status of research data repositories, such as Marcial and Hemminger' s 2010 survey of 100 RDRs [22] and Schaaf' s similar study in 2011 [23]. Across disciplines, numerous different RDR projects have emerged. Even within a single discipline, biomedicine offers a substantial number of RDRs that shape today' s research data infrastructure landscape. Moreover, biomedical digital infrastructure can be accessed by other specialized academic communities.

The Nucleic Acids Research journal' s 2013 "Molecular Biology Database" (http://www.oxfordjournals.org/nar/database/a/) presents 1,512 infrastructure facilities for storing biological research data [24]. Among these, 200 infrastructure facilities are managed within the ELIXIR bioinformatics project scope for European life sciences infrastructure, with at least 350 staff from 100 research institutions responsible for operating these repositories. Communities comprising thousands of scientists use such RDRs. The annual direct cost of these 200 RDRs is approximately €30 million [25]. For sustainable operation of biomedical RDRs, ELIXIR joined the European Strategy Forum on Research Infrastructures (ESFRI) to strategically promote and maintain Europe' s position as an international research center within the European Research Area (ERA). Since its launch in 2004, ESFRI has recognized that research infrastructure includes not only physical infrastructure such as research fleets or particle accelerators but also digital information infrastructure like "electronic archive systems for research publications and databases" [26].

2.1 Types of Research Data Repositories

The following typology of RDRs has evolved from analysis of 400 repositories. Based on broad differences between institutional and disciplinary academic literature repositories [27], the authors distinguish institutional, disciplinary, multidisciplinary, and project-specific RDRs [28]. The distinctions among these four repository types are described below. This systematic classification helps provide a comprehensive overview of different concepts and strategies for permanent access and reuse of research data infrastructure.

(1) Institutional Research Data Repositories
Institutional research data repositories are operated by universities or research institutions and function at an interdisciplinary scale within the university level. Edinburgh DataShare (http://datashare.is.ed.ac.uk) is a UK example of an institutional RDR. The University of Edinburgh established a "web-based digital repository for interdisciplinary research data" [29] built on the DSpace software framework between 2007 and 2009 [30]. As of March 2013, the repository stored 61 datasets. Open Data LMU (http://data.ub.uni-muenchen.de) from the University of Munich is another institutional RDR example from Germany. Since 2010, it has used ePrints software as a research data publishing platform for all university members [31]. The repository stored 35 datasets as of March 2013.

(2) Disciplinary Research Data Repositories
Prominent examples in disciplinary RDRs include GenBank and PANGAEA. GenBank' s service (http://www.ncbi.nlm.nih.gov/genbank), launched in 1982, positions itself as a "public database of nucleotide sequences supporting bibliographic data and biological annotation."The U.S. National Center for Biotechnology Information (NCBI) operates this infrastructure, providing information on over 250,000 nucleotide sequences [33]. PANGAEA (http://www.pangaea.de), a data publisher for earth and environmental sciences, positions itself as "an open-access library for archiving, publishing, and distributing georeferenced data from

earth system research" [34]. This RDR is operated by the Alfred Wegener Institute for Polar and Marine Research (AWI) and the Center for Marine Environmental Sciences (MARUM) at the University of Bremen. PANGAEA began as a "paleoclimate data center" funded by the German Federal Ministry of Education and Research (BMBF) between 1994 and 1997 [35]. In 2011, PANGAEA stored approximately 500,000 datasets in the earth sciences domain [36].

(3) Multidisciplinary Research Data Repositories
Beyond institutional and disciplinary approaches, research data repositories can also serve multidisciplinary needs. Figshare (http://figshare.com), one such RDR case, "allows researchers to publish all of their data in a citable, searchable, and shareable manner" [37]. Since 2011, Figshare has been operated by the Digital Science division of Macmillan Publishers [38]. A second example, LabArchives (http://www.labarchives.com), is "web-based electronic notebook software"operated by a private company that allows scientists to "store, organize, and publish their research data" [39].

(4) Project-Specific Research Data Repositories
Special research projects generate particular research data, leading to various RDR configurations. The Scientific Drilling Database (SDDB) (http://www.scientificdrilling.org), operated by the GFZ German Research Centre for Geosciences, serves as a model, providing open reuse of drilling data generated by the International Continental Scientific Drilling Program (ICDP) [40]. The Bern Digital Pantheon project (http://www.digitalpantheon.ch/repository) is another example, offering free access to high-resolution images and visualizations of the Roman Pantheon.

These four types present the overall landscape of heterogeneous RDRs and are practically applied to describe RDR services that meet potential user needs.

2.2 Requirements for Research Data Management Services and Tools

From researchers' perspectives, various barriers currently affect scholars' willingness to actively share research data. Comprehensive studies by Kuipers and Van der Hoeven [13], Tenopir et al. [14], and the ODE project [41] demonstrate that willingness to share data is closely related to supportive research data infrastructure. Repositories embedded in scholarly workflows, along with associated incentives, can promote data sharing. In a survey of over 1,300 scientists, Tenopir et al. concluded that "the majority of respondents in almost all disciplines···would be willing to place at least some of their research data in a central repository with no restrictions." However, one barrier is scholars' lack of comprehensive knowledge about existing RDRs.

To address this, re3data.org has taken action. Today, obtaining a complete overview of existing RDRs is difficult in most disciplines. Although mechanisms like the Directory of Open Access Repositories (OpenDOAR) (http://www.opendoar.org) and the Registry of Open Access Repositories (ROAR) (http://roar.eprints.org) exist, they contain

information on only a small portion (less than 5%) of research data repositories, as these mechanisms focus on repositories for scholarly publications. In recent years, websites such as the Open Access Directory (OAD) (http://oad.simmons.edu/oadwiki/Data_repositories) and DataCite (http://www.datacite.org/repolist) have begun listing RDRs. However, these directories and similar services provide only basic information about RDRs and their services, such as brief descriptions of maintained repositories, disciplines, and URLs. To overcome the barriers identified in user surveys [13,14,41], it is necessary to provide researchers, funding agencies, libraries, and publishers with systematic and user-friendly overviews of RDRs. This means that if a registry is to deliver substantive information or implement open sharing under conditions of research data reuse, it must provide more detailed RDR descriptions than existing directories. Researchers want to know how to store their data, so they need information about RDR characteristics such as: How long has this RDR been online? What kind of funding does it receive? Does this RDR have policies? Who is responsible for its operation? This is essential information for researchers to trust an RDR.

3. The re3data.org Registry Mechanism for Research Data Repositories

Researchers are paying increasing attention to the storage and utilization of RDRs. However, without trustworthy, persistent, and sustainable infrastructure to support scientists in sharing their research data, the technical and policy demands of open science [42,43], including open access to publicly funded research data and results, are doomed to fail. Surveys of RDR operations show uncertainty about financial support for infrastructure beyond a five-year secure period [13,44]. Consequently, current research development strategies and long-term funding plans have many gaps that need active filling.

The European Commission commissioned a study in 2010 that described a vision for research data management in 2030. From this, we learn that researchers need to be able to "discover, access, and process the data they need." Furthermore, when collecting data, researchers will "store their trusted data in reliable repositories" based on international standards [45]. On the other hand, the core challenge for the academic system in realizing this vision is addressing the growth of data sharing and the heterogeneity of RDRs. In response, the registry mechanism re3data.org (http://re3data.org) was developed and operates a directory of RDRs. This initiative aims to index and structurally describe RDRs across all fields based on a registration mechanism. Using information icons to describe each RDR' s basic characteristics creates added value by enabling a rapid and easy-to-use system.

The re3data.org project core members include the Library and Information Services (LIS) at the German Research Centre for Geosciences (GFZ), the Berlin School of Library and Information Science at Humboldt-Universität zu Berlin, and the KIT Library at the Karlsruhe Institute of Technology (KIT). These

three project partners have long-standing cooperative relationships with the German Initiative for Networked Information (DINI). DINI funded a research data policy report published in 2009 [46]. The German Research Foundation (DFG) funded the first project phase from January 2012 to December 2013.

The project' s primary goal is to provide guidance on the heterogeneous RDR landscape for researchers on both the data provider and data user sides, and to serve research funders and infrastructure maintainers such as data centers and academic libraries. Furthermore, re3data.org aims to establish a more centralized and integrated "ecosystem of data repositories" [47]; the registry mechanism depicts worldwide RDR development. This global overview can also assist disciplines where RDRs are not yet well developed.

Initially, the re3data.org list contained only a few RDRs with basic information such as repository name, maintainer, and subject area. By December 2012, the project had collected and documented nearly 400 infrastructure facilities for storing research data. The current Open Access Directory (OAD) list has adopted this landscape overview. The three re3data.org project partners independently examined 20 randomly selected RDRs each. The first analysis confirmed an extremely heterogeneous RDR landscape and formed the basis for creating a draft RDR description framework. Due to the lack of a suitable framework, re3data.org developed a new metadata framework for describing RDRs.

In the second phase, this framework was aligned with similar metadata frameworks, icon elements were adjusted, and basic conditions for RDRs were introduced. The first version of key terms for describing RDRs was published in a July 2012 document [48]. Comments on the icons were gathered via email feedback and the project website, ensuring transparency in the development of the icon system and gaining participation and acceptance from the RDR community. Feedback was overwhelmingly positive, with some cases providing very detailed elaboration. The project received input from reBIND (http://rebind.bgbm.org), DataCite (http://www.datacite.org), OpenAIREplus (http://www.openaire.eu), and other sources. The project team analyzed and discussed all comments and recommended incorporating key terms to guide revisions. The second version of core elements was published in December 2012 [49]. These icon elements (the icon system) include the following aspects (see Figure 1 [Figure 1: see original paper]): (1) General information (e.g., brief RDR description, content types, keywords); (2) Repository community (e.g., responsible institution, content or technical issues); (3) Policies (e.g., RDR policies including their URLs); (4) Legal issues (licenses for databases and datasets); (5) Technical standards (e.g., application programming interfaces, dataset versioning, RDR software); and (6) Quality standards and services (e.g., certificates, audit processes).

Due to heterogeneous needs across academic communities and the common lack of RDR standards, certification conditions and RDR audit procedures have been examined [50-54], with some conditions found not universally applicable to RDRs. Therefore, the barrier to entry for RDRs to join the re3data.org registry

needs to be appropriately lowered. However, for repositories to be indexed in re3data.org, details about research data access and licensing are indispensable. If your RDR meets these basic requirements, it can be indexed and reviewed. The icon set highlights the main characteristics of repositories, as shown in Figure 2 [Figure 2: see original paper]. This icon system helps users select appropriate repositories for storing their data. Researchers can clearly see each RDR' s access and usage conditions and other features on re3data.org.

The website (http://www.re3data.org/faq) explains the icons and their meanings. These icons motivate RDR operators to register on re3data.org. However, the icon system is useful not only for researchers but also for RDR maintainers, helping them compare repository strengths and weaknesses. This makes re3data.org a practical tool that continuously updates and evolves with RDRs.

The simple and straightforward re3data.org search design (see Figure 3 [Figure 3: see original paper]) displays the following for each search result: the RDR name, subject coverage, icons describing the infrastructure (see Figure 4 [Figure 4: see original paper]), and information on whether it has been reviewed and approved by re3data.org. Information filtering refines search results using secondary options such as subject, content type, country, certification, open access, persistent identifiers, and review status to generate search pages. By clicking on the repository name, users can browse each RDR portal as shown in Figure 5 [Figure 5: see original paper].

RDR maintainers can suggest their digital infrastructure for inclusion in re3data.org through a simple application process. The project team lists and examines submitted repositories. When minimum inclusion criteria are met, repositories are indexed, meaning they can provide data access and have clear terms explained on their webpages. In practice, we have considered that overly compartmentalized RDR website structures would consume indexing time, and currently only a few RDRs have service policies, designated communities, and terms of use. Some RDRs require contacting maintainers to obtain this information. Given the frequency of such cases, we have optimized the re3data.org workflow and improved feedback channels for RDR maintainers.

With European Commissioner Kroes, responsible for the Digital Agenda (http://ec.europa.eu/digital-agenda), declaring "we want to open the era of open science," openness has become the paradigm of digital science [42]. This requires developing a permanent information infrastructure that enables scientists to share their research data and allows future generations of researchers to continue accessing and reusing it.

All re3data.org project partners fund the long-term operation of the registry. Based on stakeholder feedback, re3data.org will continue developing new functions and services for research data management. A memorandum of understanding was signed with DataCite in spring 2012 for this arrangement. As one of the initiatives for persistent unique identifiers for research data, DataCite is an outcome of a DFG-funded data publishing project and an alliance member

of re3data.org [17]. Information exchange between the two groups is extremely important in this cooperation model. Discussions are currently underway with related initiatives such as Databib (http://databib.org). To promote the technical and structural development of the registry, re3data.org and its project partners will continue to foster tighter integration and larger-scale coherence of RDRs.

Although re3data.org is still in its early stages, as of July 2013, 400 RDRs have been indexed in re3data.org, with over 288 reviewed and approved. The next project phase focuses on improving usability and implementing new features. Beyond registration development, this project pursues standards and networks for research data repositories. The project strives to make all registered metadata openly available under the Creative Commons CC0 license (public domain dedication). Meanwhile, re3data.org has opened a practical path for open science.

References

[1] Berlin Declaration. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [OL]. [2009-10-27]. http://oa.mpg.de/openaccess-berlin/berlindeclaration.html.
[2] Organisation for Economic Co-operation and Development (OECD). Principles and Guidelines for Access to Research Data from Public Funding [OL]. [2013-05-16]. http://www.oecd.org/dataoecd/9/61/38500813.pdf.
[3] The Royal Society. Science as an Open Enterprise [OL]. [2013-05-16]. http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.
[4] European Commission. Commission Recommendation on Access to and Preservation of Scientific Information [OL]. [2013-05-16]. http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf.
[5] Office of Science and Technology Policy. Increasing Access to the Results of Federally Funded Scientific Research [OL]. [2013-05-16]. http://www.whitehouse.gov/sites/default/files/micros
[6] European Commission. Communication from the Commission to the European Parliament, the European Economic and Social Committee and the Committee of the Regions [OL]. [2013-05-16]. http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf.
[7] Kindling M, Schirmbacher P. "Die Digitale Forschungswelt" als Gegenstand der Forschung [J]. Information - Wissenschaft & Praxis, 2013, 64: 127-136.
[8] Pampel H, Bertelmann R. "Data Policies" im Spannungsfeld Zwischen Empfehlung und Verpflichtung [A]. //Büttner S, Hobohm H C, Müller L, (edt). Handbuch Forschungsdatenmanagement [M]. Bad Honnef: Bock+Herchen, 2011:49-61.
[9] National Science Foundation(NSF). Proposal and Award Policies and Procedures Guide. Chapter VI - Other Post Award Requirements and Considerations

[OL]. [2013-05-16]. http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4.
[10] National Science Foundation(NSF). Proposal and Award Policies and Procedures Guide. Grant Proposal Guide. Chapter II - Proposal Preparation Instructions [OL]. [2013-05-16]. http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp.
[11] Deutsche Forschungsgemeinschaft. Proposal Preparation Instructions. DFG form 54.01. [OL]. [2013-05-16]. http://www.dfg.de/formulare/54_01/54_01_en.pdf.
[12] Nature. Availability of Data and Materials [OL]. [2013-05-16]. http://www.nature.com/authors/policies/availability.html.
[13] Kuipers T, Van der Hoeven J. Insight into Digital Preservation of Research Output in Europe [R/OL]. [2013-05-16]. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.
[14] Tenopir C, Allard S, Douglass K, et al. Data Sharing by Scientists: Practices and Perceptions [J]. PLoS ONE, 2011, 6(6): e21101.
[15] Nature Biotechnology. Credit Where Credit is Overdue [J]. Nature Biotechnology, 2009, 27: 579.
[16] Pampel H, Dallmeier-Tiessen S. Open Research Data - From Vision to Practice [A]. //Bartling S, Friesike S, (edt). Opening Science–The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing [M]. Heidelberg: Springer, 2009.
[17] Klump J, Bertelmann R, Brase J, et al. Data Publication in the Open Access Initiative [J]. Data Science Journal, 2006, 5: 79-83.
[18] Chavan V, Penev L. The Data Paper: A Mechanism to Incentivize Data Publishing in Biodiversity Science [J]. BMC Bioinformatics, 2011, 12: S2. DOI:10.1186/1471-2105-12-S15-S2.
[19] Woutersen-Windhouwer S, Brandsma R, Hogenaar A, et al. Enhanced Publications?: Linking Publications and Research Data in Digital Repositories [OL]. [2013-05-16]. http://dare.uva.nl/aup/nl/record/316849.
[20] European Commission. ICT Infrastructures for e-Science [OL]. [2013-05-16]. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF.
[21] Lagoze C, Sompel H Van De. The Open Archives Initiative: Building a Low-Barrier Interoperability Framework [OL]. [2013-05-16]. http://public.lanl.gov/herbertv/papers/Papers/2001/J
[22] Marcial L H, Hemminger B M. Scientific Data Repositories on the Web: An Initial Survey [J]. Journal of the American Society for Information Science and Technology, 2010, 61: 2029-2048.
[23] Schaaf I. Forschungsdaten im Netz –Untersuchung Von Online Verfügbaren Repositorien. Hochschule der Medien Stuttgart [A]. //Schäfer L, Pampel H, Klump J, et al. Bericht Symposium "Forschungsdaten-Infrastrukturen (FDI 2013)" [C]. 2013. DOI:10.2312/radieschen_003.
[24] Fernández-Suárez X M, Galperin M Y. The 2013 Nucleic Acids Research Database Issue and the Online Molecular Biology Database Collection [OL]. [2013-05-16]. http://www.ncbi.nlm.nih.gov/pubmed/23203983.
[25] ELIXIR. The ELIXIR Strategy for Data Resources [OL]. [2013-05-16]. http://www.elixir-europe.org/prep/sites/elixir-europe.org.prep/files/documents/reports/elixir_strategy_for_
[26] European Strategy Forum on Research Infrastructures. European Roadmap for Research Infrastructures. Report 2006 [OL]. [2013-05-16]. http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_

[27] Suber P. Open Access. Cambridge, Massachusetts [OL]. [2013-03-15]. The MIT Press. http://mitpress.mit.edu/books/open-access.
[28] Pampel H, Goebelbecker H J, Vierkant P. re3data.org: Aufbau eines Verzeichnisses Forschungsdaten-Repositorien. Ein Werkstattbericht [A]. //Mittermaier B, Wissen V. Daten, Menschen, Systeme. WissKom 2012. Jülich: Verlag des Forschungszentrums Jülich, 2012: 61-73.
[29] Edinburgh DataShare. What is Edinburgh DataShare? [OL]. [2013-05-16]. http://datashare.is.ed.ac.uk.
[30] Rice R. DISC-UK DataShare [OL]. [2013-05-16]. http://ie-repository.jisc.ac.uk/336/.
[31] Schallehn V. Open Data LMU. Universitätsbibliothek München [OL]. [2013-05-16]. http://www.ub.uni-muenchen.de/no_cache/aktuelles/einzelnenachricht/article/open-data-lmu/.
[32] Cravedi K. GenBank Celebrates 25 Years of Service with Two-Day Conference [OL]. [2013-05-16]. http://www.nih.gov/news/health/apr2008/nlm-03.htm.
[33] Benson D A, Karsch-Mizrachi I, Clark K, et al. GenBank [J]. Nucleic Acids Research, 2012, 40: D48-D53.
[34] PANGAEA. About/Imprint [OL]. [2013-05-16]. http://www.pangaea.de/about/.
[35] Diepenbroek M, Grobe H, Reinke M, et al. Data Management of Proxy Parameters with PANGAEA [A]. //Fischer G, Wefer G. Use of Proxies in Paleoceanography [M]. New York: Springer-Verlag, 1999:715-727.
[36] Schindler U, Diepenbroek M, Grobe H. PANGAEA-Research Data Enters Scholarly Communication Online Geophysical Research Abstracts, 2012, 14: EGU2012-5344.
[37] Figshare. FAQs [OL]. [2013-05-16]. http://figshare.com/faqs.
[38] Fenner M. Figshare: Interview with Mark Hahnel [OL]. [2013-05-16]. http://blogs.plos.org/mfenner/2012/02/16/figshare-interview-with-mark-hahnel/.
[39] LabArchives. Electronic Notebook Software for Laboratory Notebooks FAQ [OL]. [2013-05-16]. http://www.labarchives.com/faqs.php.
[40] Klump J, Conze R. The Scientific Drilling Database (SDDB)-Data from Deep Earth Monitoring and Sounding [OL]. [2013-05-16]. http://www.iodp.org/images/stories/downloads/sd4_
[41] Dallmeier-Tiessen S, Darby R, Gitmans K, et al. Summary of the Studies, Thematic Publications and Recommendations [OL]. [2013-05-16]. http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=Summaryofthestudies%2Cthematicpublicationsandrecommendations.
[42] Kroes N. Opening Science Through e-Infrastructures [OL]. [2013-05-16]. http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258.
[43] Kroes N. Making Open Access a Reality for Science [OL]. [2013-05-16]. http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/392.
[44] Pfeiffenberger H, Pampel H, Schäfer A, et al. Report and Strategy on Annotation, Reputation and Data Quality [OL]. [2013-05-16]. http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=D26.1ReportandStrategyonAnnotation%2CReputationandDataQuality.
[45] High Level Expert Group on Scientific Data. Riding the Wave. How Europe can Gain from the Rising Tide of Scientific Data [OL]. [2013-05-16].

http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf.
[46] Dallmeier-Tiessen S, Dobratz S, Gradmann S, et al. Positionspapier Forschungsdaten. 1.0 ed [OL]. [2013-05-16]. http://nbn-resolving.de/urn:nbn:de:kobv:11-10098082.
[47] Graaf M, Waaijers L. A Surfboard for Riding the Wave-Towards a Four Country Action Programme on Research Data [OL]. [2013-05-16]. http://www.knowledge-exchange.info/Default.aspx?ID=469.
[48] Vierkant P, Spier S, Rücknagel J, et al. Vocabulary for the Registration and Description of Research Data Repositories [OL]. [2013-05-16]. http://dx.doi.org/10.2312/re3.001.
[49] Vierkant P, Spier S, Rücknagel J, et al. Vocabulary for the Registration and Description of Research Data Repositories (Version 2.0) [OL]. [2013-05-16]. http://dx.doi.org/10.2312/re3.002.
[50] Braun K, Buddenbohm S, Dobratz S, et al. DINI Certificate Document and Publication Services 2010 [OL]. [2013-05-16]. http://nbn-resolving.de/urn:nbn:de:kobv:11-100182800.
[51] Consultative Committee for Space Data Systems (CCSDS). Audit and Certification of Trustworthy Digital Repositories [OL]. [2013-05-16]. http://public.ccsds.org/publications/archive/652x0m1.pdf.
[52] Data Seal of Approval. Data Seal of Approval. Guidelines Version 1 [OL]. [2013-05-16]. http://assessment.datasealofapproval.org/documentation/.
[53] ESF & EUROHORCs (2011) Basic Requirements for Research Infrastructures in Europe [OL]. [2013-05-16]. http://www.dfg.de/download/pdf/foerderung/programme/wgi/basic_requ
[54] ICSU World Data System (2011) Certification of World Data System Members [OL]. [2013-05-16]. http://icsu-wds.org/images/files/Certification_summary_6_Jul_2011.pdf.

(Corresponding author: Gu Liping, E-mail: gulp@mail.las.ac.cn)

Note: Figure translations are in progress. See original paper for figures.