re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Vocabulary for the Registration and Description of Research Data Repositories

Authors: Paul Vierkant [b], Shaked Spier [b], Jessika Rücknagel [b], Jens Gundlach [c], Maxi Kindling [b], Heinz Pampel [a], Stefka Manova [c], Hans-Jürgen Göbelbecker [c], Jens Klump [a], Roland Bertelmann [a], Peter Schirmbacher [b], Frank Scholze [c]

[a] GFZ German Research Centre for Geosciences, Library and Information Services (LIS)

[b] Humboldt-Universität zu Berlin, Berlin School of Library and Information Science

[c] Karlsruhe Institute of Technology (KIT), KIT Library

Contact

info@re3data.org
http:// re3data.org

Table of Contents

1 Introduction

1.1 re3data.org

Research data are valuable and ubiquitous. The permanent access to research data is a challenge for all stakeholders in the scientific community. The long-term preservation and the principle of open access to research data offer broad opportunities for the scientific community. More and more universities and research centres are starting to build research data repositories allowing permanent access to data sets in a trustworthy environment. Due to disciplinary requirements, the landscape of data repositories is very heterogeneous. Thus it is difficult for researchers, funding bodies, publishers and scholarly institutions to select appropriate repositories for storage and search of research data.

The goal of re3data.org is to create a global registry of research data repositories. The registry will cover research data repositories from different academic disciplines. re3data.org will present repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. In the course of this mission re3data.org aims to promote a culture of sharing, increased access and better visibility of research data. In the first phase of the project the following tasks are prioritized:

• the conception and construction of a web-based registry of research data repositories;

• the definition of selection criteria of research data repositories;

• the formulation of a vocabulary to describe research data repositories.

According to the schedule the registry will go live in autumn 2012.

re3data.org is funded by the German Research Foundation DFG[1].

re3data.org is a joint project of the Berlin School of Library and Information Science at the Humboldt-Universität zu Berlin[2], the GFZ German Research Centre for Geosciences[3] and the Karlsruhe Institute of Technology (KIT)[4]. The partners are actively involved in the German Initiative for Network Information (DINI)[5] and current research data management activities.

1.2 The Vocabulary

The re3data.org vocabulary is a list of metadata properties describing a research data repository regarding its general scope, content and infrastructure as well as its compliance with technical, metadata and quality standards. The vocabulary includes required metadata

---

[1] http://www.dfg.de
[2] http://www.ibi.hu-berlin.de
[3] http://bib.gfz-potsdam.de
[4] http://www.bibliothek.kit.edu
[5] http://www.dini.de/english

properties and optional properties providing additional information. The vocabulary is designed to

• recommend a standard for describing a research data repository;

• provide the basis for interoperability between research data repositories and re3data.org;

• be one of the first steps towards the goal of a certificate for research data repositories.

The existing vocabulary was developed out of 20 randomly chosen repositories of a survey of 400 research data repositories. These findings as well as suggestions from project partners have led to the current and preliminary version 1.0. Future developments of the vocabulary will also depend on the feedback of the research data repository community. This open and transparent development process is to ensure a strong basis for a future standard for describing research data repositories supported by and rooted in the community. Taking all responses into account the preliminary version 1.0 will be updated to 1.1 and published in autumn 2012. We appreciate your feedback and therefore would like to invite you to tell us what you think about our work done so far.

1.3   Research Data Repository Registration Procedure

Until the launch of re3data.org in autumn 2012 institutions can suggest research data repositories to be listed in the registry via email (info@re3data.org). After the launch research data repositories can be registered via an online form. The suggested repositories will be checked by the re3data.org team and, if compatible with the re3data.org minimum requirements, published in the registry.

2 re3data.org Metadata Properties

The table below provides a detailed description of the re3data.org properties that describe a research data repository registered in re3data.org. For an example of how the properties are expressed in XML format, please see the XML example provided at the end of this document.

A naming convention has been used for all properties and attributes as follows: properties, attributes and child properties begin with a lower case letter. If the name of the element is a compound of more than one word, subsequent words begin with capital letters. The indicator A/C shows whether the property being described is an attribute (A) or a child (C) of the corresponding property that has preceded it.

The attribute Occurrence explains if a property can have multiple instances, which is indicated by the notation: Occ.1-n, meaning that a property must occur once (1), and may occur multiple times (n). A notation of "Req" indicates that an attribute is required if the corresponding property is applied; "Opt" indicates that it is optional. Where the allowed values specify a "Controlled List," the members of the list are regulated by the re3data.org team.

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
1	identifier	The ID is a unique string, which allows identifying the research data repository assigned by re3data.org.		1	Auto-Value
2	repositoryName	The name of the repository.		1	The format is open.
3	additionalTitle	The alternative title or acronym for the data repository.		0-1	The format is open.
4	repositoryURL	The URL, which gives reference to the repository.		1	URL
5	repositoryIdentifier	A globally unique identifier that refers to the research data repository.		0-n	Examples: DOI[6], URN, VIAF[7], GND[8]
6	description	A textual description containing additional information about the data repository.		0-1	The format is open (max. 1000 letters).

---

[6] http://www.doi.org/
[7] http://www.oclc.org/viaf/
[8] http://en.wikipedia.org/wiki/Universal_Authority_File

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
7	type	The type of the research data repository.		1-n	*Controlled list* Allowed values: institutional disciplinary other
8	size	The number of items contained in the research data repository.		0-1	Number
8.1	sizeUpdate	The date of the last update of the research data repository size.	C	Req	YYYY or YYYY-MM-DD or any other format described in W3CDTF.[9] (ISO-8601)
9	startDate	Releasing date of the research data repository.		0-1	YYYY or YYYY-MM-DD or any other format described in W3CDTF. (ISO-8601)
10	endDate	Date when the research data repository stopped the ingest of new research data (still making the research data available).		0-1	YYYY or YYYY-MM-DD or any other format described in W3CDTF. (ISO-8601)
11	repositoryLanguage	The user interface language.		1-n	*Controlled Vocabulary* Allowed values from: ISO-639-3[10] Examples: eng, deu, fra
12	subject	The subject of the research data repository.		1-n	The format is dependent on scheme.
12.1	subjectScheme	The name and/or URL of the subject scheme or classification code.	A	Req	*Controlled Vocabulary* Allowed values from: DFG Classification of Subject Area, Review Board, Research Area and Scientific Discipline[11] Examples: 11 Humanities, 10101 Prehistory

---

[9] http://www.w3.org/TR/NOTE-datetime
[10] http://www.sil.org/iso639-3/codes.asp
[11] http://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/dfg_fachsystematik_en_08_11.pdf

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
13	missionStatementURL	The URL of a mission statement describing the designated community of the research data repository.		0-1	URL
14	contentType	All types of resources available in the research data repository.		0-n	The format is dependent on scheme.
14.1	contentTypeScheme	The name and/or URL of the Subject scheme or classification code.	A	Req	*Controlled Vocabulary* Allowed values from: PARSE.Insight Survey, Question 4[12]
15	providerType	The type of provider.		1-2	*Controlled list* Allowed values: dataProvider serviceProvider
16	keywords	Keywords describing the subject focus of the research data repository.		0-n	The format is open.
17	institution	All institutions being responsible for funding, creating and/or running the research data repository.		1-n	-
17.1	institutionName	Name of the responsible institution.	C	1	The format is open.
17.2	institutionCountry	Location of the responsible institution.	C	1-n	*Controlled Vocabulary* Allowed values from: ISO-3166-1 alpha-3[13] Examples: DEU, GBR, USA
17.3	responsibilityType	Type of responsibility for each responsible institution.	C	0-n	*Controlled list* Allowed values: general technical metadata funding sponsoring
17.4	institutionType	Type of responsible institution.	C	0-1	*Controlled list* Allowed values: non-profit commercial
17.5	institutionURL	URL of the responsible institution.	C	0-n	URL

---

[12] http://www.parse-insight.eu/downloads/PARSE-insight_survey_questions_research.pdf
[13] http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
17.6	institutionIdentifier	A globally unique identifier that refers to the research data repository (e.g. ORCID, VIAF, GND).	C	0-n	Examples: DOI, URN, VIAF, GND
17.7	responsibilityStartDate	Start date of period of responsibility.	C	0-n	YYYY or YYYY-MM-DD or any other format described in W3CDTF. (ISO-8601)
17.8	responsibilityEndDate	End date of period of responsibility.	C	0-n	YYYY or YYYY-MM-DD or any other format described in W3CDTF. (ISO-8601)
17.9	emailAddress	Email address of the contact.	C	0-n	The format is open.
18	policyType	Any kind of policy (e.g. data policy, etc.) provided by the research data repository to clarify legal aspects.		0-n	The format is open.
18.1	policyURL	URL of the policy.	C	1	URL
19	databaseAccess	Type of access to the research data repository.		1	*Controlled list* Allowed values: open restricted closed
20	databaseLicense	Database license of the research data repository.		0-n	*Controlled list* Allowed values: CC[14] ODBL[15] PDDL[16] other none
20.1	databaseLicenseURL	Database licence URL.	C	Opt	URL
21	dataAccess	Type of access to data.		1-n	*Controlled list* Allowed values: open restricted closed
21.1	restrictionType	All existing restrictions to the research data repository. (Required if restricted is chosen.)	C	Req	*Controlled list* Allowed values: registration, registrationFeeRequired other

---

[14] http://creativecommons.org/licenses
[15] http://opendatacommons.org/licenses/odbl
[16] http://opendatacommons.org/licenses/pddl

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
22	dataLicense	License of the data, existing in the research data repository.		1-n	*Controlled list* Allowed values: CC ODC-BY[17] OGL[18] RL[19] other none
22.1	dataLicenseURL	Data license URL.	C	Opt	URL
23	dataUpload	Type of data upload.		1	*Controlled list* Allowed values: open restricted closed
24	dataUploadLicense	The licence for data upload.		0-1	The format is open.
24.1	dataUploadLicenseURL	Deposit licence URL.	C	Opt	URL
25	software	The software that is used to run the research data repository.		0-1	*Controlled list* Allowed values: DataVerse[20] DigitalCommons[21] dLibra[22] Dspace[23] ePrints[24] eSciDoc[25] Opus[26] other
26	versioning	Versioning		0-1	*Controlled list* Allowed values: yes no
27	api	API		0-n	*Controlled list* Allowed values: api[27] ftp[28] oai-pmh[29] rest[30] soap[31] sparql[32] sword[33] other

---

[17] http://opendatacommons.org/licenses/by

[18] http://www.nationalarchives.gov.uk/doc/open-government-licence

[19] http://www.ausgoal.gov.au/restrictive-licence-template

[20] http://thedata.org

[21] http://digitalcommons.bepress.com

[22] http://dlibra.psnc.pl/index.php?lang=en

[23] http://www.dspace.org

[24] http://www.eprints.org

[25] https://www.escidoc.org

[26] http://www.opus-repository.org

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
27.1	apiURL	API URL	C	1-n	URL
28	pidSystem	The persistent identifier system that is used by the research data repository.		1-n	*Controlled list* Allowed values: ark[34] doi hdl[35] purl[36] urn other none
29	citationExampleURL	The URL of the research data repository providing information on how to cite its research data.		1	URL
30	qualityManagement	Any form of quality management concerning the data or metadata of the research data repository.		1	*Controlled list* Allowed values: yes no
31	certificates	The certificate, seal or standard with which the research data repository complies.		0-n	*Controlled list* Allowed values: DIN31644[37] DINI-Zertifikat[38] DSA[39] ISO16363[40] ISO16919[41] TRAC[42] WDS[43]

---

[27] http://en.wikipedia.org/wiki/Api

[28] http://en.wikipedia.org/wiki/Ftp

[29] http://www.openarchives.org/OAI/openarchivesprotocol.html

[30] http://en.wikipedia.org/wiki/Representational_state_transfer

[31] http://en.wikipedia.org/wiki/SOAP

[32] http://en.wikipedia.org/wiki/SPARQL

[33] http://en.wikipedia.org/wiki/SWORD_%28Protocol%29

[34] http://en.wikipedia.org/wiki/Archival_Resource_Key

[35] http://www.handle.net

[36] http://en.wikipedia.org/wiki/PURL

[37] http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&artid=133059734&committeeid=54738855&bcrumblevel=2&languageid=de

[38] http://www.dini.de/dini-zertifikat/english

[39] http://www.datasealofapproval.org

[40] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510

[41] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57950

[42] http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

[43] http://icsu-wds.org/images/files/Certification_summary_6_Jul_2011.pdf

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
32	syndication	Alerting service provided by the research data repository.		0-n	*Controlled list* Allowed values: RSS ATOMfeed
32.1	syndicationURL	Alerting service URL.	C	1-n	URL
33	remarksIntern	Remarks for re3data internal use.		0-1	The format is open.
34	remarksExtern	Additional remarks that are visible to all users.		0-1	The format is open.
35	entryDateIntern	The date the metadata entry of the research data repository was made.		n	Auto-Value: ISO-8601
36	lastUpdateIntern	The date the metadata of the research data repository was updated.		n	Auto-Value: ISO-8601
37	lastEditorIntern	Last edited by.		n	Given Name, Surname

3 XML Example

This is an XML example illustrating how the re3data.org vocabulary can be used to describe research data repositories.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!--re3data.org Vocabulary for the Registration and Description of Research Data Repositories. Version 1.0, July 2012. doi:10.2312/re3.001-->
<re3data.org xmlns:r3d="http://www.re3data.org">
  - <r3d:repository>
        <r3d:identifier>r3d000001001</r3d:identifier>
        <r3d:repositoryName>PANGAEA</r3d:repositoryName>
        <r3d:additionalTitle>Publishing Network for Geoscientific and Environmental Data</r3d:additionalTitle>
        <r3d:repositoryURL>http://www.pangaea.de</r3d:repositoryURL>
        <r3d:repositoryIdentifier/>
        <r3d: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.</r3d:description>
        <r3d:type>disciplinary</r3d:type>
        <r3d:size/>
        <r3d:sizeUpdate/>
        <r3d:startDate/>
        <r3d:endDate/>
        <r3d:repositoryLanguage>eng</r3d:repositoryLanguage>
        <r3d:subject subjectScheme="DFG Classification of Subject Area">31302
Oceanography</r3d:subject>
        <r3d:subject subjectScheme="DFG Classification of Subject Area">31401 Geology and
            Palaeontology</r3d:subject>
        <r3d:subject subjectScheme="DFG Classification of Subject Area">31501 Geophysics, Geodesy,
Remote Sensing, Geoinformatics</r3d:subject>
        <r3d:subject subjectScheme="DFG Classification of Subject Area">31601 Geochemistry,
            Mineralogy and Crystallography</r3d:subject>
        <r3d:subject subjectScheme="DFG Classification of Subject Area">21 Biology</r3d:subject>
        <r3d:missionStatementURL/>
        <r3d:contentType contentTypeScheme="PARSE.Insight">Standard office documents</r3d:contentType>
        <r3d:contentType contentTypeScheme="PARSE.Insight">Images</r3d:contentType>
        <r3d:contentType contentTypeScheme="PARSE.Insight">Plain text</r3d:contentType>
        <r3d:contentType contentTypeScheme="PARSE.Insight">Archived data</r3d:contentType>
        <r3d:contentType contentTypeScheme="PARSE.Insight">Audiovisual (multimedia)
data</r3d:contentType>
        <r3d:providerType>dataProvider</r3d:providerType>
        <r3d:keywords>Earth Science</r3d:keywords>
        <r3d:keywords>Environmental Science</r3d:keywords>
    - <r3d:responsibleInstitution>
            <r3d:institutionName>Alfred Wegner Institute for Polar and Marine Research
            (AWI)</r3d:institutionName>
            <r3d:institutionCountry>DEU</r3d:institutionCountry>
            <r3d:responsibilityType>general</r3d:responsibilityType>
            <r3d:responsibilityType>technical</r3d:responsibilityType>
            <r3d:institutionType>non-profit</r3d:institutionType>
            <r3d:institutionURL>http://www.awi.de/</r3d:institutionURL>
            <r3d:institutionIdentifier/>
            <r3d:responsibilityStartDate/>
            <r3d:responsibilityEndDate/>
            <r3d:emailAddress>hgrobe@pangaea.de</r3d:emailAddress>
        </r3d:responsibleInstitution>
    - <r3d:responsibleInstitution>
            <r3d:institutionName>Center for Marine Environmental Sciences (MARUM)</r3d:institutionName>
            <r3d:institutionCountry>DEU</r3d:institutionCountry>
            <r3d:responsibilityType>general</r3d:responsibilityType>
            <r3d:institutionType>non-profit</r3d:institutionType>
            <r3d:institutionURL>http://www.marum.de/</r3d:institutionURL>
```

```xml
            <r3d:institutionIdentifier/>
            <r3d:responsibilityStartDate/>
            <r3d:responsibilityEndDate/>
            <r3d:emailAddress>mdiepenbroek@pangaea.de</r3d:emailAddress>
        </r3d:responsibleInstitution>
    - <r3d:policy>
            <r3d:policyType>Data policy of the information system PANGAEA</r3d:policyType>
            <r3d:policyURL>http://www.pangaea.de/curator/files/pangaea-data-
            policy.pdf</r3d:policyURL>
        </r3d:policy>
        <r3d:databaseAccess>open</r3d:databaseAccess>
        <r3d:databaseLicense/>
        <r3d:databaseLicenseURL/>
        <r3d:dataAccess>open</r3d:dataAccess>
        <r3d:restrictionType/>
        <r3d:dataAccess>restricted</r3d:dataAccess>
        <r3d:restrictionType>registration</r3d:restrictionType>
    - <r3d:dataLicenseGroup>
            <r3d:dataLicense>CC</r3d:dataLicense>
            <r3d:dataLicenseOther/>
            <r3d:dataLicenseURL>http://wiki.pangaea.de/wiki/License</r3d:dataLicenseURL>
        </r3d:dataLicenseGroup>
        <r3d:dataUpload>restricted</r3d:dataUpload>
        <r3d:dataUploadLicense>Data Submission</r3d:dataUploadLicense>
        <r3d:dataUploadLicenseURL>http://wiki.pangaea.de/wiki/Data_submission</r3d:dataUploadLicenseURL>
    - <r3d:technicalStandards>
            <r3d:software>other</r3d:software>
            <r3d:softwareOther>Sybase</r3d:softwareOther>
            <r3d:versioning>yes</r3d:versioning>
        </r3d:technicalStandards>
    - <r3d:apiGroup>
            <r3d:api>oai-pmh</r3d:api>
            <r3d:apiOther/>
            <r3d:apiURL>ws.pangaea.de/oai/</r3d:apiURL>
        </r3d:apiGroup>
        <r3d:pidSystem>doi</r3d:pidSystem>
        <r3d:pidSystemOther/>
        <r3d:citationExampleURL>http://wiki.pangaea.de/wiki/Citation</r3d:citationExampleURL>
        <r3d:qualityManagement>yes</r3d:qualityManagement>
        <r3d:certificates>WDS</r3d:certificates>
        <r3d:certificatesOther/>
        <r3d:syndication>RSS</r3d:syndication>
        <r3d:syndicationURL>http://www.pangaea.de/tools/latest-datasets.rss</r3d:syndicationURL>
        <r3d:remarksIntern/>
        <r3d:remarksExtern/>
        <r3d:entryDateIntern>2012-07-16</r3d:entryDateIntern>
        <r3d:lastUpdateIntern>2012-07-16</r3d:lastUpdateIntern>
        <r3d:lastEditorIntern>Jens Gundlach</r3d:lastEditorIntern>
    </r3d:repository>
</re3data.org>
```