re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Schema for the Description of Research Data Repositories - RFC Version

Authors: Paul Vierkant [b], Shaked Spier [b], Jessika Rücknagel , [b] Heinz Pampel [a], Florian Fritze [b], Jens Gundlach [c], David Fichtmüller [d], Maxi Kindling [b], Agnes Kirchhoff [d], Hans-Jürgen Goebelbecker [c], Jens Klump [f], Gabriele Kloska [c], Evelyn Reuter [c], Angelika Semrau [c], Edeltraud Schnepf [c], Michael Skarupianski [c], Roland Bertelmann [a], Peter Schirmbacher [b], Frank Scholze [c], Claudia Kramer [c], Michal Witt [e], Claudio Fuchs [a], Robert Ulrich [c]

[a] GFZ German Research Centre for Geosciences, Library and Information Services (LIS), Germany
[b] Humboldt-Universität zu Berlin, Berlin School of Library and Information Science (BLIS), Germany
[c] Karlsruhe Institute of Technology (KIT), KIT Library, Germany
[d] Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Germany
[e] Purdue University Libraries, West Lafayette, USA
[f] Commonwealth Scientific and Industrial Research Organisation (CSIRO), Mineral Resources Flagship, Kensington WA, Australia

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

re3data.org is a global registry of research data repositories that covers research data repositories from different academic disciplines. It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. re3data.org promotes a culture of sharing, increased access and better visibility of research data. The registry is funded by the German Research Foundation (DFG)[1] and went online in autumn 2012.

Project partners in re3data.org are the Berlin School of Library and Information Science at the Humboldt-Universität zu Berlin[1], the Library and Information Services department (LIS) of the GFZ German Research Centre for Geosciences[2], and the KIT Library at the Karlsruhe Institute of Technology (KIT)[3]. The partners are actively involved in the German Initiative for Network Information (DINI)[4] and current research data management activities. re3data.org cooperates, among others, with BioSharing[5] and OpenAIRE[6].

Some publishers and journals like Copernicus Publications, PeerJ, Springer and Nature's Scientific Data refer to re3data.org in their Editorial Policies as a tool for the easy identification of appropriate data repositories to store research data[7].

The use of re3data.org is also recommended in the European Commission's "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020"[8].

---

[1] http://www.ibi.hu-berlin.de
[2] http://bib.telegrafenberg.de
[3] http://www.bibliothek.kit.edu
[4] http://www.dini.de/english
[5] See: http://www.re3data.org/2013/11/biosharing-and-re3data-cooperation/
[6] See: http://www.re3data.org/2013/10/memorandum-of-understanding-between-openaire-and-re3data-org/
[7] See: http://www.re3data.org/2013/07/re3data-org-is-moving-forward/ and
http://www.re3data.org/2013/09/new-repositories-and-functionalities/
[8] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

In March 2013 re3data.org and Databib announced to merge their two directories into one service that will be managed under the auspices of DataCite by the end of 2015. The aim of merging these two activities is to reduce duplication of effort and to better serve the research community with a single, sustainable registry of research data repositories that incorporates the best features of both projects[9].

Further information on re3data.org can be found in the following publication:

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H.-J., Gundlach, J., Schirmbacher, P., Dierolf, U. (2013). Making Research Data Repositories Visible: The re3data.org Registry. PLOS ONE, 8(11), e78080. doi:10.1371/journal.pone.0078080. http://doi.org/10.1371/journal.pone.0078080

1.2   The Schema

The re3data.org schema 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 schema includes required metadata properties and optional properties providing additional information. The schema 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 first version of the schema (formerly called 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 version 1.0. After its publication in July 2012 we received feedback leading to substantial changes in the structure of the schema. Version 2.0 of the schema synthesizes all the responses as well as current developments in the realm of research data repositories. Version 2.1 and 2.2 have introduced minor changes, as well as a policy passage. Future developments of the schema will still 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. We appreciate your feedback and therefore would like to invite you to tell us what you think about our work done so far.

---

[9] See: http://www.re3data.org/2014/03/datacite-re3data-org-databib-collaboration/

1.3 Research Data Repository Registration Workflow

RDR operators can suggest their infrastructures to be listed in re3data.org via a simple application form providing the name and URL of the RDR.[10] The RDR will automatically be listed in re3data.org with an icon indicating that it is "not yet reviewed by re3data.org." The project team thoroughly analyzes the website of the RDR using a handbook that gives practical information on how to obtain the metadata properties of the re3data.org schema. A repository is indexed when the minimum requirements of the re3data.org policy 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 web sites. Before a new record of a RDR is published in re3data.org all gathered information is reviewed by a second team member (see figure 1).
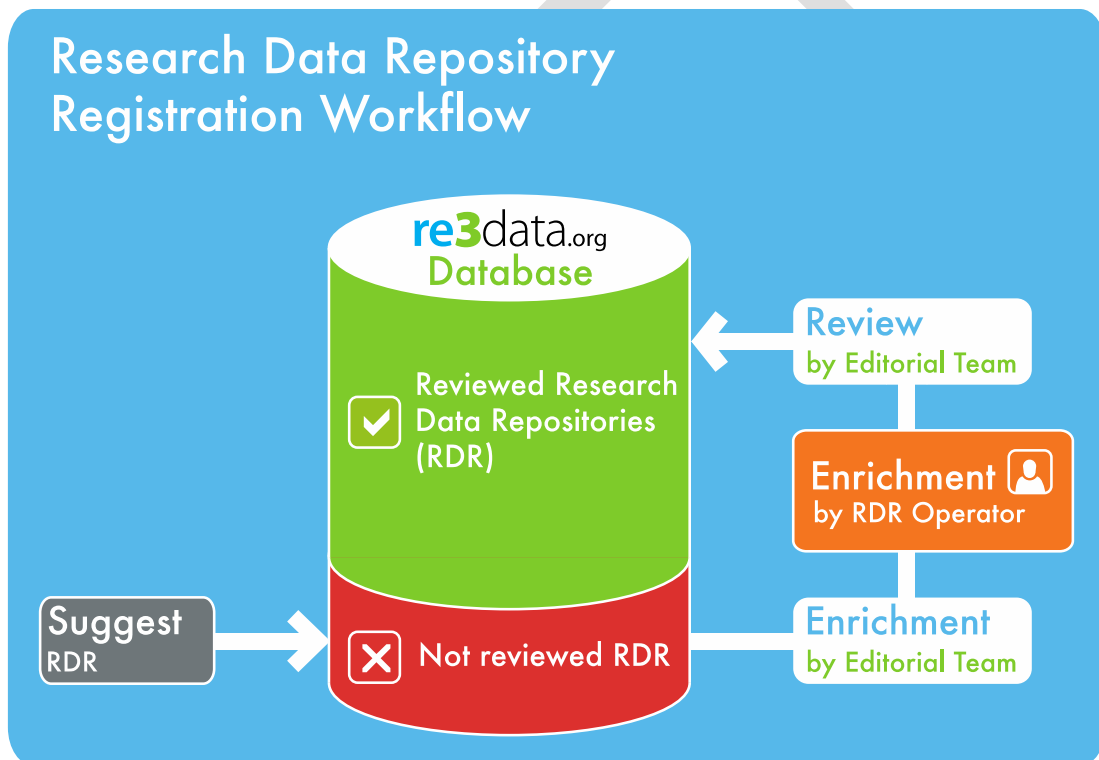

Figure 1: Research Data Repository Registration Workflow of re3data.org

---

[10] http://www.re3data.org/suggest/

1.4 re3data.org Policy

A research data repository is a subtype of a sustainable information infrastructure which provides long-term storage and access to research data that is the basis for a scholarly publication.

Research data means information objects generated by scholarly projects for example through experiments, measurements, surveys or interviews.

To be registered in re3data.org a research data repository must

- be operated by a legal entity with an organisational framework that provides sustainability (e.g. library, university);
- clarify access conditions to the data and repository as well as the terms of use;
- have an English graphical user interface (GUI);
- have focus on research data.

A research data repository listed in re3data.org is either:

- a data provider if it offers research data and its metadata (ideally exposing metadata via interfaces),

or

- a service provider (e.g. a portal) if it harvests the metadata of research data from data providers as a basis for building value-added services.

1.5 Version Update

Version	Date of Publication	Changes to prior Version
1.0	July 2012	Initial Document
2.0	December 2012	Structural changes, including addition of language attributes, addition of definitions of controlled vocabularies, changes in occurrences (See: http://dx.doi.org/10.2312/re3.003)
2.1	December 2013	Addition and deletion of properties, addition of controlled vocabularies, changes in occurrences and definitions, addition of definitions, change of title
2.2	December 2014	Addition of properties, addition of controlled vocabularies, changes and additions of definitions

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 (see section 3).

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 (Occ)* 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). "Req" indicates that an attribute is required if the corresponding property is applied. Where the allowed values specify a "Controlled Vocabulary," the members of the list are regulated by the re3data.org team. The values of the controlled vocabularies are explained and defined in the appendix (4.1).

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
1	re3data.orgIdentifier	A globally unique and persistent identifier that refers to the research data repository entry in re3data.org. The ID is assigned by re3data.org.		1	Auto-Value Example: **http://doi.org/**10.2315/r3d1 00000001
2	repositoryName	The full name of the research data repository.		1	The format is open. Example: Access to Archival Databases
2.1	language	Language of the research data repository name.	A	Req	Controlled vocabulary Allowed values from: ISO-639-3. Examples: eng, deu, fra
3	additionalName	The alternative name or acronym for the research data repository.		0-n	The format is open. Example: AAD

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
3.1	language	Language of the research data repository additional name.	A	Req	Controlled vocabulary Allowed values from: ISO-639-3. Examples: eng, deu, fra
4	repositoryURL	The URL, which gives reference to the research data repository.		1	URL Example: http://cdiac.esd.ornl.gov/home.html
5	repositoryIdentifier	A globally unique identifier (identifier in form of a link) that refers to the research data repository (e.g. DOI, URN, VIAF, GND).		0-n	Example: **http://doi.org/**10.1234/foo
6	description	A textual description containing additional information about the data repository (primary language is English).		0-1	The format is open (max. 1000 characters).
6.1	language	Language of the research data repository description.	A	Req	Controlled vocabulary Allowed values from: ISO-639-3. Examples: eng, deu, fra
7	repositoryContact	Email address of the contact or an URL of an online contact form of the repository.		0-n	The format is open. Example: info@researchdata.com
8	type	The type of the research data repository.		1-n	Controlled vocabulary Allowed values: disciplinary institutional other
9	size	The number of items contained in the research data repository.		0-1	The format is open. Example: 5.000 datasets; 30 studies
9.1	updated	The date of the last update of the research data repository size.	A	Req	YYYY or YYYY-MM-DD or any other format described in W3C-DTF (ISO-8601)
10	startDate	Releasing date of the research data repository.		0-1	YYYY or YYYY-MM-DD or any other format described in W3C-DTF (ISO-8601)

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
11	endDate	Date when the research data repository went offline or stopped ingesting new research data (still making the research data available).		0-1	YYYY or YYYY-MM-DD or any other format described in W3C-DTF (ISO-8601)
12	repositoryLanguage	The user interface language of the research data repository.		1-n	Controlled vocabulary Allowed values from: ISO-639-3. Examples: eng, deu, fra
13	subject	The subject(s) of the research data repository.		1-n	The format is dependent on scheme. Examples: 11 Humanities 10101 Prehistory
13.1	subjectScheme	The subject scheme according to which the subject (see ID 13 subject) of the research data repository described.	A	Req	Controlled vocabulary Allowed values from: DFG
14	missionStatementURL	The URL of a mission statement describing the designated community of the research data repository.		0-1	URL Example: http://cdiac.ornl.gov/aboutc diac.html
15	contentType	All types of resources available in the research data repository.		0-n	The format is dependent on scheme. Examples: Raw data, Images
15.1	contentTypeScheme	The name and/or URL of the subject scheme or classification code (e.g. parse)	A	Req	Controlled vocabulary Allowed values from: parse
16	providerType	The type of provider.		1-2	Controlled vocabulary Allowed values: dataProvider serviceProvider
17	keyword	English keyword(s) describing the subject focus of the research data repository.		0-n	Examples: computer science
18	institution	All institutions being responsible for funding, creating and/or running the research data repository (wrapper element).		1-n	-

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
18.1	institutionName	Name of the responsible institution.	C	1	Example: National Institute of General Medical Sciences, U.S. National Institutes of Health
18.1.1	language	Language of the institution name.	A	Req	Controlled vocabulary Allowed values from: ISO-639-3. Examples: eng, deu, fra
18.2	institutionAdditionalName	The alternative name or acronym for the responsible institution.	C	0-n	Example: NIGMS
18.2.1	language	Language of the institution additional name.	A	Req	Controlled vocabulary Allowed values from: ISO-639-3. Examples: eng, deu, fra
18.3	institutionCountry	Location of the responsible institution.	C	1	Controlled vocabulary Allowed Values from: ISO-3166-1 alpha-3. Examples: DEU, GBR, USA
18.4	responsibilityType	Type of responsibility for each responsible institution.	C	0-n	Controlled vocabulary Allowed values: funding general sponsoring technical
18.5	institutionType	Type of responsible institution.	C	0-1	Controlled vocabulary Allowed values: commercial non-profit
18.6	institutionURL	URL of the responsible institution.	C	0-1	Example: http://www.nigms.nih.gov/
18.7	institutionIdentifier	A globally unique identifier that refers to the institution (e.g. ISNI, VIAF, GND).	C	0-n	Example: isni: 0000000100755874
18.8	responsibilityStartDate	Start date of period of responsibility.	C	0-1	YYYY or YYYY-MM-DD or any other format described in W3C-DTF (ISO-8601)
18.9	responsibilityEndDate	End date of period of responsibility.	C	0-1	YYYY or YYYY-MM-DD or any other format described in W3C-DTF (ISO-8601)
18.10	institutionContact	Email address of the contact or an URL of an online contact form of the institution.	C	0-n	The format is open. Example: john.doe@researchdata.com

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
19	policy	Any kind of policy (e.g. data policy, etc.) provided by the research data repository to clarify legal aspects (wrapper element).		0-n	-
19.1	policyName	Name of the policy.	C	1	The format is open. Example: Data policy of the information system PANGAEA
19.2	policyURL	URL of the policy.	C	1	Example: http://www.pangaea.de/curator/files/pangaea-data-policy.pdf
20	databaseAccess	(wrapper element)		1	-
20.1	databaseAccessType	Type of access to the research data repository.	C	1	Controlled vocabulary Allowed values: open restricted closed
20.2	databaseAccessRestriction	All existing access restrictions to the research data repository. (Required if restricted is chosen).	C	0-n	Controlled vocabulary Allowed values: feeRequired registration other
21	databaseLicense	Database license of the research data repository (wrapper element).		0-n	-
21.1	databaseLicenseName	Name of database license.	C	1	Controlled vocabulary Allowed values: Apache License 2.0 BSD CC CC0 Copyrights ODC Public Domain other
21.2	databaseLicenseURL	Database license URL.	C	1	Example: http://creativecommons.org/licenses/
22	dataAccess	(wrapper element)		1-n	-
22.1	dataAccessType	Type of access to data.	C	1	Controlled vocabulary Allowed values: open embargoed restricted closed

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
22.2	dataAccessRestriction	All existing access restrictions to the research data. (Required if restricted is chosen).	C	0-n	Controlled vocabulary Allowed values: feeRequired institutional membership registration other
23	dataLicense	License of the data, existing in the research data repository. (wrapper element)		1-n	-
23.1	dataLicenseName	Name of the data license.	C	1	Controlled vocabulary Allowed values: Apache License 2.0 BSD CC CC0 Copyrights ODC OGL OGLC Public Domain RL other none
23.2	dataLicenseURL	Data license URL.	C	1	Example: http://creativecommons.org/ licenses/
24	dataUpload	(wrapper element)		1-n	-
24.1	dataUploadType	Type of data upload.	C	1	Controlled vocabulary Allowed values: open restricted closed
24.2	dataUploadRestriction	All existing restrictions to the data upload. (Required if restricted is chosen).	C	0-n	Controlled vocabulary Allowed values: feeRequired institutional membership registration other
25	dataUploadLicense	The license for data upload (wrapper element).		0-n	-
25.1	dataUploadLicenseName	Name of the licence for data upload.	C	1	Example: RAD Submission of Array Data
25.2	dataUploadLicenseURL	Data upload license URL.	C	1	URL Example: http://cbil.upenn.edu/RAD/p hp/RAD-data_sub.php?

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
26	software	The software that is used to run the research data repository (wrapper element).		0-n	-
26.1	softwareName	Name of the research data repository software.	C	1	Controlled vocabulary Allowed values: CKAN DataVerse DigitalCommons dLibra DSpace EPrints eSciDoc Fedora MySQL Nesstar Opus other unknown
27	versioning	The research data repository supports versioning of research data.		0-1	Controlled vocabulary Allowed values: yes no
28	api	API URL		0-n	Example: http://www.datadryad.org/oai/
28.1	apiType	Type of API.	A	Req	Controlled vocabulary Allowed values: FTP NetCDF OAI-PMH OpenDAP REST SOAP SPARQL SWORD other
29	pidSystem	The persistent identifier system that is used by the research data repository.		1-n	Controlled vocabulary Allowed values: ARK DOI HDL PURL URN other none

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
30	citationGuidelineURL	The URL of the research data repository providing information on how to cite its research data. The DataCite citation format is recommended (http://www.datacite.org/whycitedata).		0-1	Example: http://wiki.pangaea.de/wiki/Citation; http://archaeologydataservice.ac.uk/archsearch/record.jsf?titleId=1686608
31	aidSystem	The author identifier system that is used by the research data repository.		0-n	Controlled vocabulary Allowed values: AuthorClaim ISNI ORCID ResearcherID other none
32	enhancedPublication	The research data repository offers the interlinking between publications and data.		0-1	Controlled vocabulary Allowed values: yes no unknown
33	qualityManagement	Any form of quality management concerning the data or metadata of the research data repository.		0-1	Controlled vocabulary Allowed values: yes no unknown
34	certificate	The certificate, seal or standard the research data repository complies with.		0-n	Controlled vocabulary: CLARIN certificate B DIN 31644 DINI Certificate DRAMBORA DSA ISO 16363 ISO 16919 RatSWD TRAC Trusted Digital Repository WDS other
35	metadataStandard	(wrapper element)		0-n	
35.1	metadataStandardName	Metadata standards the research data repository complies with.	C	1	The format is dependent on scheme. Examples: AgMES - Agricultural Metadata Element Set other
35.1.1	metadataStandardScheme	The scheme of the metadata standards.	A	Req	Controlled vocabulary Allowed values from: DCC

ID	re3data.org property	Definition	A/C	Occ	Allowed values, examples, other constraints
35.2	metadataStandardURL	The URL of the metadata standard.	C	1	URL Example: http://www.dcc.ac.uk/resources/metadata-standards/agmes-agricultural-metadata-element-set
36	syndication	URL of the alerting service(s) provided by the research data repository.		0-n	Example: https://www.datadryad.org/feed/atom_1.0/10255/3
36.1	syndicationType	Type of alerting service.	A	Req	Controlled Vocabulary: ATOM RSS other
37	remarks	Additional remarks that are visible to all users.		0-1	The format is open.
38	entryDate	The date the research data repository was indexed in re3data.org.		1	YYYY-MM-DD (ISO-8601)
39	lastUpdate	The date the metadata of the research data repository was updated.		1	YYYY-MM-DD (ISO-8601)

3   XML Example

This is an XML example (of a fictional research data repository) illustrating how the re3data.org schema can be used to describe research data repositories.

```xml
<?xml version="1.0" encoding="utf-8"?>
<!--re3data.org Schema for the Description of Research Data Repositories. RFC Version 2.2, September 2014. doi: 10.5281/zenodo.11748 -->
<r3d:re3data xmlns:r3d="http://www.re3data.org/schema/2-2" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.re3data.org/schema/2-2 http://schema.re3data.org/2-2/re3dataV2.2.xsd ">
  <r3d:repository>
    <r3d:re3data.orgIdentifier>http://doi.org/10.2315/r3d100000001</r3d:re3data.orgIdentifier>
    <r3d:repositoryName language="eng">Global Research Data Repository</r3d:repositoryName>
    <r3d:additionalName language="eng">GReDaR</r3d:additionalName> <!-- optional --> <!-- multiple -->
    <r3d:additionalName language="deu">Globales Forschungsdatenrepositorium</r3d:additionalName> <!-- optional --> <!-- multiple -->
    <r3d:repositoryURL>http://www.globalresearchdatarepository.org</r3d:repositoryURL>
    <r3d:repositoryIdentifier>http://doi.org/0123456789/gredar</r3d:repositoryIdentifier> <!-- optional --> <!-- multiple -->
    <r3d:description language="eng">The Global Research Data Repository (GReDaR) is the place where researchers from all academic disciplines can
put their research data. Research data deposited in GReDaR</r3d:description> <!-- optional -->
    <r3d:repositoryContact>info@greadar.org</r3d:repositoryContact> <!-- optional --> <!-- multiple -->
    <r3d:type>disciplinary</r3d:type> <!-- multiple -->
    <r3d:size updated="2012-11-23">10.000 datasets; 323 studies</r3d:size> <!-- optional -->
    <r3d:startDate>2011-01-01</r3d:startDate> <!-- optional -->
    <r3d:endDate>2013-07-31</r3d:endDate> <!-- optional -->
    <r3d:repositoryLanguage>eng</r3d:repositoryLanguage> <!-- multiple -->
    <r3d:repositoryLanguage>deu</r3d:repositoryLanguage> <!-- multiple -->
    <r3d:repositoryLanguage>fra</r3d:repositoryLanguage> <!-- multiple -->
    <r3d:subject subjectScheme="DFG">31302 Oceanography</r3d:subject> <!-- multiple -->
    <r3d:subject subjectScheme="DFG">111 Social Sciences</r3d:subject> <!-- multiple -->
    <r3d:subject subjectScheme="DFG">20103 Cell Biology</r3d:subject> <!-- multiple -->
    <r3d:missionStatementURL>http://www.globalresearchdatarepository.org/mission_statement</r3d:missionStatementURL> <!-- optional -->
    <r3d:contentType contentTypeScheme="parse">Structured text</r3d:contentType> <!-- optional --> <!-- multiple -->
    <r3d:contentType contentTypeScheme="parse">Images</r3d:contentType> <!-- optional --> <!-- multiple -->
    <r3d:providerType>dataProvider</r3d:providerType> <!-- multiple -->
    <r3d:keyword>Earth Science</r3d:keyword> <!-- optional --> <!-- multiple -->
    <r3d:keyword>Measure Data</r3d:keyword> <!-- optional --> <!-- multiple -->
    <r3d:keyword>Social Science</r3d:keyword> <!-- optional --> <!-- multiple -->
    <r3d:keyword>Maps</r3d:keyword> <!-- optional --> <!-- multiple -->
    <r3d:institution> <!-- multiple -->
      <r3d:institutionName language="eng">Institute for Research Data</r3d:institutionName>
      <r3d:institutionAdditionalName language="eng">IRD</r3d:institutionAdditionalName> <!-- optional --> <!-- multiple -->
      <r3d:institutionAdditionalName language="deu">Institut für Forschungsdaten</r3d:institutionAdditionalName> <!-- optional --> <!-- multiple --
>
      <r3d:institutionCountry>DEU</r3d:institutionCountry>
      <r3d:responsibilityType>general</r3d:responsibilityType> <!-- optional --> <!-- multiple -->
      <r3d:responsibilityType>technical</r3d:responsibilityType> <!-- optional --> <!-- multiple -->
      <r3d:institutionType>non-profit</r3d:institutionType> <!-- optional -->
      <r3d:institutionURL>http://www.ird.org</r3d:institutionURL> <!-- optional -->
      <r3d:institutionIdentifier>urn:47110815/ird</r3d:institutionIdentifier> <!-- optional --> <!-- multiple -->
      <r3d:responsibilityStartDate>2011-01-01</r3d:responsibilityStartDate> <!-- optional -->
```

```xml
          <r3d:responsibilityEndDate>2014-07-31</r3d:responsibilityEndDate> <!-- optional -->
          <r3d:institutionContact>John Doe, Institut für Forschungsdaten, Unter den Linden 6, 10099 Berlin, Germany</r3d:institutionContact> <!--
optional --> <!-- multiple -->
          <r3d:institutionContact>johndoe@globalresearchdatarepository.org</r3d:institutionContact> <!-- optional --> <!-- multiple -->
          <r3d:institutionContact>http://www.globalresearchdatarepository.org/contactform/</r3d:institutionContact> <!-- optional --> <!-- multiple -->
      </r3d:institution>
      <r3d:policy> <!-- optional --> <!-- multiple -->
          <r3d:policyName>Data Management Policy</r3d:policyName>
          <r3d:policyURL>http://www.globalresearchdatarepository.org/policy.html</r3d:policyURL>
      </r3d:policy>
      <r3d:databaseAccess>
          <r3d:databaseAccessType>restricted</r3d:databaseAccessType>
          <r3d:databaseAccessRestriction>registration</r3d:databaseAccessRestriction> <!-- optional --> <!-- multiple -->
      </r3d:databaseAccess>
      <r3d:databaseLicense> <!-- optional --> <!-- multiple -->
          <r3d:databaseLicenseName>CC0</r3d:databaseLicenseName>
          <r3d:databaseLicenseURL>http://creativecommons.org/publicdomain/zero/1.0/</r3d:databaseLicenseURL>
      </r3d:databaseLicense>
      <r3d:dataAccess> <!-- multiple -->
          <r3d:dataAccessType>embargoed</r3d:dataAccessType>
      </r3d:dataAccess>
      <r3d:dataAccess> <!-- multiple -->
          <r3d:dataAccessType>restricted</r3d:dataAccessType>
          <r3d:dataAccessRestriction>registration</r3d:dataAccessRestriction> <!-- optional --> <!-- multiple -->
      </r3d:dataAccess>
      <r3d:dataAccess> <!-- multiple -->
          <r3d:dataAccessType>closed</r3d:dataAccessType>
      </r3d:dataAccess>
      <r3d:dataLicense> <!-- multiple -->
          <r3d:dataLicenseName>CC</r3d:dataLicenseName>
          <r3d:dataLicenseURL>http://creativecommons.org/licenses/</r3d:dataLicenseURL>
      </r3d:dataLicense>
      <r3d:dataLicense> <!-- multiple -->
          <r3d:dataLicenseName>Copyrights</r3d:dataLicenseName>
          <r3d:dataLicenseURL>http://www.globalresearchdatarepository.org/copyrights/dataset1234</r3d:dataLicenseURL>
      </r3d:dataLicense>
      <r3d:dataUpload>
          <r3d:dataUploadType>restricted</r3d:dataUploadType>
          <r3d:dataUploadRestriction>registration</r3d:dataUploadRestriction> <!-- optional --> <!-- multiple -->
      </r3d:dataUpload>
      <r3d:dataUploadLicense> <!-- optional --> <!-- multiple -->
          <r3d:dataUploadLicenseName>GReDaR Data Deposit License</r3d:dataUploadLicenseName>
          <r3d:dataUploadLicenseURL>http://www.globalresearchdatarepository.org/deposit_license/</r3d:dataUploadLicenseURL>
      </r3d:dataUploadLicense>
      <r3d:software> <!-- optional --> <!-- multiple -->
          <r3d:softwareName>EPrints</r3d:softwareName>
      </r3d:software>
      <r3d:versioning>yes</r3d:versioning> <!-- optional -->
      <r3d:api apiType="OAI-PMH">http://www.globalresearchdatarepository.org/oai</r3d:api> <!-- optional --> <!-- multiple -->
      <r3d:api apiType="REST">http://www.globalresearchdatarepository.org/rest</r3d:api> <!-- optional --> <!-- multiple -->
      <r3d:pidSystem>DOI</r3d:pidSystem> <!-- multiple -->
```

```xml
      <r3d:citationGuidelineURL>http://www.globalresearchdatarepository.org/how-to-cite.html</r3d:citationGuidelineURL> <!-- optional -->
      <r3d:aidSystem>ORCID</r3d:aidSystem> <!-- optional -->
      <r3d:enhancedPublication>yes</r3d:enhancedPublication> <!-- optional -->
      <r3d:qualityManagement>yes</r3d:qualityManagement> <!-- optional -->
      <r3d:certificate>DSA</r3d:certificate> <!-- optional --> <!-- multiple -->
      <r3d:certificate>ISO16919</r3d:certificate> <!-- optional --> <!-- multiple -->
      <r3d:metadataStandard metadataStandardScheme="DCC"> <!-- multiple -->
        <r3d:metadataStandardName>DataCite Metadata Schema</r3d:metadataStandardName>
        <r3d:metadataStandardURL>http://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema</r3d:metadataStandardURL>
      </r3d:metadataStandard> <!--optional-->
      <r3d:syndication syndicationType="RSS">http://www.globalresearchdatarepository.org/news.rss</r3d:syndication> <!-- optional --> <!-- multiple -
->
      <r3d:remarks>This is additional information on the research data repository visible to all users.</r3d:remarks> <!-- optional -->
      <r3d:entryDate>2012-11-15</r3d:entryDate>
      <r3d:lastUpdate>2012-12-21</r3d:lastUpdate>
  </r3d:repository>
</r3d:re3data>
```

4 Appendix

4.1 Attribute Values and Controlled Vocabularies

The following tables describe and define the allowed attribute values as well as the controlled vocabularies of the re3data.org metadata properties (see section 2).

2.1; 3.1; 6.1; 17.1; 18.1.1; 18.2.1   language

Value	Definition
Examples: eng, deu, fra	ISO-639-3; http://www.iso.org/iso/catalogue_detail?csnumber=39534

8 type

Value	Definition
disciplinary	Disciplinary repository (Subject repository): "This is a collection of research outputs with a common link to a particular subject discipline. Subject repositories are likely to cover one broad-based discipline, with contributors from many different institutions supported by a variety of funders; the repositories themselves are likely to be funded from one or more sources within the subject community. Although for some subject repositories the funding may be fragile, if they are of enough importance to the community then funding crises are usually weathered."[11]
institutional	Institutional Repository: "This is a collection of research outputs with a common link to a particular institution, usually by authorship. These repositories are likely to cover more than one research discipline, to have funders in many if not all the Research Councils and support communities who have different approaches to research dissemination. Whether deposit of content is mandatory is a decision that will be made by each institution. The institutions may have many requirements for the content of the repository, from open access dissemination, through metrics, marketing to strategic planning. It is likely that many of these processes in the past were undertaken through collection of bibliographic information."[12]

---

[11] Jones, C et al. "Report of the Subject and Institutional Repositories Interactions Study." (2008): pag. 5. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.140.7316&rep=rep1&type=pdf retrieved 28 Aug. 2013.
[12] Jones, C et al. "Report of the Subject and Institutional Repositories Interactions Study." (2008): pag. 5

Value	Definition
other	Other (Repository Type): A research data repository that is neither institutional nor disciplinary, e.g. a funder repository: "This is a collection of research outputs with a common link to one (or more) funder(s). These are likely to cover the funder's remit, which is usually subject-based but can become indistinct at subject boundaries, and will have authors from many institutions. Deposit of content is usually mandatory and can include project-related material, such as completion project reports. The funders will have requirements for the content of repository, from metrics through to strategic planning."[13]

9.1 updated

Value	Definition
YYYY or YYYY-MM-DD	ISO-8601; http://www.w3.org/TR/NOTE-datetime

10 startDate

Value	Definition
YYYY or YYYY-MM-DD	ISO-8601; http://www.w3.org/TR/NOTE-datetime

11 endDate

Value	Definition
YYYY or YYYY-MM-DD	ISO-8601; http://www.w3.org/TR/NOTE-datetime

12 repositoryLanguage

Value	Definition
Examples: eng, deu, fra	ISO-639-3; http://www.iso.org/iso/catalogue_detail?csnumber=39534

13.1 subjectScheme

Value	Definition
DFG	DFG Classification of Subject Area, Review Board, Research Area and Scientific Discipline (using Level 1-4). http://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/index.jsp (retrieved 29 Aug. 2014)

15 contentType

PARSE.Insight type	Examples of File Formats
Standard office documents	text documents, spreadsheets, presentations
Networkbased data	web sites, email, chat history, etc.
Databases	DBASE, MS Access, Oracle, MySQL, etc.
Images	JPEG, JPEG2000, GIF, TIF, PNG, SVG, etc.
Structured graphics	CAD, CAM, 3D, VRML, etc.

---

[13] Jones, C et al. "Report of the Subject and Institutional Repositories Interactions Study." (2008): pag. 5

PARSE.Insight type	Examples of File Formats
Audiovisual data	WAVE, MP3, MP4, Flash, etc.
Scientific and statistical data formats	SPSS, FITS, GIS, etc.
Raw data	device specific output
Plain text	TXT in various encodings
Structured text	XML, SGML, etc.
Archived data	ZIP, RAR, JAR, etc.
Software applications	modelling tools, editors, IDE, compilers, etc.
Source code	scripting, Java, C, C++, Fortran, etc.
Configuration data	parameter settings, logs, library files
Other	-

15.1 contentTypeScheme

Value	Definition
parse	For the values of parse see the PARSE.Insight types in contentType 14. For the source see page 5 in http://www.parse-insight.eu/downloads/PARSE-insight_survey_questions_research.pdf

16 providerType

Value	Definition
dataProvider	A research data repository is a data provider if it offers research data and its metadata (ideally exposing metadata via interfaces).
serviceProvider	A research data repository is a service provider if it harvests the metadata of research data from data providers as a basis for building value-added services.

18.3 institutionCountry

Value	Definition
Examples: DEU, GBR, USA	ISO-3166-1 Alpha-3 code; http://en.wikipedia.org/wiki/ISO_3166-1#Current_codes
AAA	international institutions
EEC	European Union

18.4 responsibilityType

Value	Definition
funding	Funding responsibility refers to the institution that funds / funded the setup or running of the research data repository.
general	General responsibility refers to the institution that is generally responsible for the providing and running of the research data repository.
sponsoring	Sponsoring responsibility (in contrast to funding responsibility) refers to financial support of a research data repository in return for e.g. advertising on the website.

technical	General responsibility refers to the institution that is responsible for the technical matters of the research data repository, e.g. hosting.

18.5 institutionType

Value	Definition
commercial	A commercial institution is an institution that distributes surplus revenues as profit or dividends.
non-profit	The institution does not aim at economic profits but at non-profit (e.g. social, cultural and academic) goals of its members.

18.8 responsibilityStartDate

Value	Definition
YYYY or YYYY-MM-DD	ISO-8601; http://www.w3.org/TR/NOTE-datetime

18.9 responsibilityEndDate

Value	Definition
YYYY or YYYY-MM-DD	ISO-8601; http://www.w3.org/TR/NOTE-datetime

20.1 databaseAccessType

Value	Definition
open	There are no access barriers.
restricted	External users can overcome access barriers.
closed	External users cannot overcome access barriers.

20.2 databaseRestriction

Value	Definition
feeRequired	A single or regular payment is required.
registration	A free registration is required.
other	Other restrictions beside registration or payment such as a written consent are required.

21.1 databaseLicenseName

Value	Definition
Apache License 2.0	The research data is licensed under an Apache License. http://www.apache.org/licenses/
BSD	The research data is licensed under Berkeley Software Distribution License. http://opensource.org/licenses/bsd-license.php
CC	The database is licensed under a Creative Commons license. URL: http://creativecommons.org/licenses/
CC0	The repository operator dedicated the database to the public domain by waiving all rights. URL: http://creativecommons.org/publicdomain/zero/1.0/

Value	Definition
Copyrights	The database is copyright-protected. Any public or private use of this data is subject to prevailing copyright laws. Please contact the content provider of this data for permission requests.[14]
ODC	The database is made available under an Open Data Commons license. URL: http://opendatacommons.org/licenses/
Public Domain	The database is in the public domain. URL: http://en.wikipedia.org/wiki/Public_domain
other	Licenses that are not explicitly listed.

22.1 dataAccessType

Value	Definition
open	There are no access barriers.
embargoed	External users cannot overcome access barriers until the data is released for open access.
restricted	External users can overcome access barriers.
closed	External users cannot overcome access barriers.

22.2 dataAccessRestriction

Value	Definition
feeRequired	A single or regular payment is required.
institutional membership	Only people associated to a member institution can access the data.
registration	A free registration is required.
other	Other restrictions beside registration or payment such as a written consent are required.

23.1 dataLicenseName

Value	Definition
Apache License 2.0	The research data is licensed under an Apache License. http://www.apache.org/licenses/
BSD	The research data is licensed under Berkeley Software Distribution License. http://opensource.org/licenses/bsd-license.php
CC	The data is licensed under a Creative Commons license. URL: http://creativecommons.org/licenses/
CC0	The researcher dedicated the research data to the public domain by waiving all of his or her rights. URL: http://creativecommons.org/publicdomain/zero/1.0/
Copyrights	The data(base) is copyright-protected. Any public or private use of this data is subject to prevailing copyright laws. Please contact the content provider of this data for permission requests.[15]

---

[14] Based on: "Copyrighted. Use Requirements to Be Specified Upon Request of the Copyright Holder." *The Cell: An Image Library - License.* http://www.cellimagelibrary.org/pages/license retrieved 28 Aug. 2013.
[15] Based on: "Copyrighted. Use Requirements to Be Specified Upon Request of the Copyright Holder." *The Cell: An Image Library - License.* http://www.cellimagelibrary.org/pages/license retrieved 28 Aug. 2013.

Value	Definition
ODC	The data is made available under an Open Data Commons license. URL: http://opendatacommons.org/licenses/
OGL	The research data is licensed under an Open Government Licence. URL: http://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/re-use-and-licensing/ukglf/
OGLC	The research data is licensed under an Open Government License-Canada. URL: http://data.gc.ca/eng/open-government-licence-canada
Public Domain	The research data is in the public domain. URL: http://en.wikipedia.org/wiki/Public_domain
RL	The research data is licensed under a Restrictive License. URL: http://www.ausgoal.gov.au/restrictive-licence-template
other	Licenses that are not explicitly listed.
none	No license exists or was found.

24.1 dataUploadType

Value	Definition
open	There are no access barriers.
restricted	External users can overcome access barriers.
closed	External users cannot overcome access barriers.

24.2 dataUploadRestrictionType

Value	Definition
feeRequired	A single or regular payment is required.
institutional membership	Only people associated to a member institution can access the data.
registration	A free registration is required.
other	Other restrictions beside registration or payment such as a written consent are required.

26.1 softwareName

Value	Definition
CKAN	http://ckan.org
DataVerse	http://thedata.org/DVN-software
Digital Commons	http://digitalcommons.bepress.com
dLibra	http://dlibra.psnc.pl
DSpace	http://www.dspace.org
EPrints	http://www.eprints.org
eSciDoc	https://www.escidoc.org
Fedora	http://fedorarepository.org/software
MySQL	http://www.mysql.com/
Nesstar	http://www.nesstar.com/
OPUS	http://www.kobv.de/opus4

Value	Definition
other	The repository is using software, which is not listed in the controlled vocabulary.
unknown	The software of the research data repository is unknown.

28.1 apiType

Value	Definition
FTP	http://en.wikipedia.org/wiki/FTP
NetCDF	http://de.wikipedia.org/wiki/NetCDF
OAI-PMH	http://en.wikipedia.org/wiki/OAI-PMH
OpenDAP	http://en.wikipedia.org/wiki/Opendap
REST	http://en.wikipedia.org/wiki/REST
SOAP	http://en.wikipedia.org/wiki/SOAP
SPARQL	http://en.wikipedia.org/wiki/Sparql
SWORD	http://swordapp.org/about/a-brief-history/
other	An API that is not explicitly listed.

29 pidSystem

Value	Definition
ARK	https://wiki.ucop.edu/display/Curation/ARK
DOI	http://en.wikipedia.org/wiki/Digital_object_identifier
hdl	http://www.handle.net/
PURL	http://en.wikipedia.org/wiki/PURL
URN	http://en.wikipedia.org/wiki/Uniform_Resource_Name
other	A persistent identifier system that is not explicitly listed.
none	No persistent identifier is used to identify a research data set.

31 aidSystem

Value	Definition
AuthorClaim	http://authorclaim.org/
ISNI	http://www.isni.org/
ORCID	http://orcid.org/
ResearcherID	http://www.researcherid.com/Home.action
other	An author identifier system that is not explicitly listed.

34 certificate

Value	Definition
CLARIN certificate B	http://www.clarin.eu/node/3767
DIN 31644	http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=133059734&languageid=en&bcrumblevel=
DINI Certificate	http://nbn-resolving.de/urn:nbn:de:kobv:11-100182800
DRAMBORA	http://www.repositoryaudit.eu/

DSA	http://www.datasealofapproval.org/
ISO 16363	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56510
ISO 16919	http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57950
RatSWD	http://ratswd.de/dl/doc/RatSWD_FDZCriteria.pdf
TRAC	http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories
Trusted Digital Repository	http://www.trusteddigitalrepository.eu/Site/Trusted%20Digital%20Repository.html
WDS	http://www.icsu-wds.org/
other	A certificate that is not explicitly listed.

35.1.1 metadataStandardScheme

Value	Definition
DCC	http://www.dcc.ac.uk/resources/metadata-standards/list
other	A standard that is not explicitly listed.

36.1 syndicationType

Value	Definition
ATOM	http://en.wikipedia.org/wiki/Atom_%28standard%29
RSS	http://en.wikipedia.org/wiki/Rss

38 entryDate

Value	Definition
YYYY or YYYY-MM-DD	ISO-8601; http://en.wikipedia.org/wiki/ISO_8601

39 lastUpdate

Value	Definition
YYYY or YYYY-MM-DD	ISO-8601; http://en.wikipedia.org/wiki/ISO_8601

4.2   Access To Research Data Repositories Explained

Beside accessibility of research data the terms of use are probably the most important issue of a research data repository. To define the general accessibility of a research data repository the term access has to be split into three categories representing different levels of access. First there is the access to the research data repository defining whether users can access the database in general. Secondly there is the access to the research data sets in the research data repository in particular. Thirdly there is the access to upload research data sets to the research data repository. The access to each level can be open, restricted and/or closed. Open means that there are no access barriers. Restricted means that external users can overcome access barriers. Closed means external users cannot overcome access barriers. Embargoed access applies only to the level of access to the research data sets and means that external users cannot overcome access barriers until the data is released for open or restricted access.

As shown in the matrix below the access to the research data repository is the basic level to define the general framework of access to the actual research data sets, e.g. a research data set in a research data repository that is restricted through registration cannot be open but just restricted or even closed; a research data set in a closed research data repository which underlies access restrictions that cannot be overcome by external users is neither open nor restricted but only closed. Thus to define the general accessibility of a research data repository the values highlighted in orange are needed (values in braces can occur).

Access (property)	Open Access		Restricted Access		Closed Access
Access to Repository (19.1 databaseAccessType)	open		open or restricted		closed
Access to Data (21.1 dataAccessType)	open (embargoed, restricted, closed)		restricted (embargoed, closed)		closed
Data Upload (23.1 dataUploadType)	open or restricted	closed	open or restricted	closed	-