

Registry interoperability core model

This document is the description of the core model proposed in the paper “Towards the Interoperability of Scholarly Repository Registries” where a study on the interoperability among repository registries was undertaken. We studied four different registries - FAIRsharing, re3data, OpenDOAR, and ROAR - and analysed their information model and content in order to isolate the similarities and the differences. For the fields shared by at least two registries, we produced a crosswalk. For each field in the crosswalk we computed its coverage, to be aware of the actual number of results presenting a specific attribute value. Only fields having a coverage greater than zero in more than two registries are used in the definition of a proposal of a minimal core model shared by all the registries as base for their interoperability. In the definition of the model, we also included fields not considered in the crosswalk, but which we believe are important when searching for a reference repository. This document presents the scheme of the proposed Minimal Core Model.

Metadata Properties

The following table lists the field of the proposed minimal core model. The column *Property* specifies the name of the field. The column *Definition* provides a brief description of the field. The column *A/C* is to specify if the field is an attribute or a child of another field. In case it is an attribute the column *Occ* is to specify if it is required or optional, otherwise it is to specify the number of allowed occurrences. The Notes are for examples or comments.

Table 1: Core Model.

ID	Property	Definition	A/C	Occ	Notes
1	identifier	A globally unique identifier of the resource.		0-n	recordURI in the cross walk
1.1	identifierType	The type of the identifier.	A	Req.	type: [DOI, RRID, ...]
2	internalIdentifier	The internal identifier of the repository.		1	
3	name	The name of the repository.		1	
3.1	language	The language of the repository name.	A	Req	A code in the reference vocabulary, e.g., eng
3.2	languageVocabulary	The vocabulary of the language.	A	Req	e.g., ISO-639-3
4	additionalName	Other names for the repository.		0-n	Free text. e.g., Datenbank Gesprochenes Deutsch
4.1	language	The language of the additional name.	A	Req	deu
4.2	languageVocabulary	The vocabulary of the language.	A	Req	e.g., ISO-639-3
5	repositoryURL	The home page of the repository.		1	e.g., https://dgd.ids-mannheim.de/
6	repositoryType	The type of the repository.		1-n	Controlled vocabulary e.g., [institutional, disciplinary, governmental, multi-disciplinary, project-related, etc.]

Continued on next page

Table 1 – continued from previous page

ID	Property	Definition	A/C	Occ	Notes
7	repositoryDescription	The description of the repository.		0-1	Free text
7.1	language	The language of the description .	A	Req	e.g., deu
7.2	languageVocabulary	The vocabulary of the language.	A	Req	e.g., ISO-639-3
8	repositoryContent	The types of the items in the repository.		0-n	Dependent on the schema or free text
8.1	scheme	The scheme of the content	A		
9	recordCount	The number of resources in the repository		0-1	
9	subject	The subject of the repository		0-n	
9.1	code	The code of the subject	A		Dependent from the scheme (desiderata)
9.2	scheme	The scheme of the subject name	A		Desired
9.3	uri	The URI of the subject	A		Desired
10	keyword	Keyword(s) and labels describing the subject focus of repository.		0-n	
11	organization	The organization(s) related to the repository		1-n	
11.1	name	The name of the organization	C		

Continued on next page

Table 1 – continued from previous page

ID	Property	Definition	A/C	Occ	Notes
11.1.1	language	The language of the organization name	A	Req	
11.1.2	languageVocabulary	The vocabulary of the language	A	Req	
11.2	acronym	The acronym of the organization	C	0-1	
11.3	id	The id of the organization	C	0-n	e.g., funding
11.4	country	The country of the organization	C	1	
11.4.1	code	The code of the country	A	Req	
11.4.2	vocabulary	The vocabulary of the country code	A	Req	
11.7	url	The home page of the organization	C	1	
12	softwareName			1	
13	versioning	Is data versioning allowed?		1	True, false
14	apiUrl	The URL of the access point to the repository		0-1	
14.1	type	The type of the api	A	Req	e.g OAI-PMH
14.2	documentationUrl	Url to the api documentation	A		Required if the API does not have a known protocol
15	policyURL	The URL of the policy		0-n	
16	access	The kind of access on the resource		0-n	e.g dataAccess, dataUpload, ...
					Continued on next page

Table 1 – continued from previous page

ID	Property	Definition	A/C	Occ	Notes
16.1	type	The type of access	A	Req	e.g., open
16.2	restrictions	The type of restriction if access is not open	A		e.g., fee
17	licence	The kind of licence on the resource		0-n	e.g., dataLicence, databaseLicence,...
17.1	type	The type	A	Req	The type of licence, e.g., CC-BY
17.2	url	The URL to the licence	A	Req	
18	startDate	The repository birth		0-1	
18.1	format	The format of the date	A	Req	
19	lastUpdate	The date of the last update of the repository		0-1	
19.1	format	The format of the date	A	Req	

XML example

The following is an XML example of the generic repository information according to the proposed model

```
<registryEntries>
  <registryEntry>
    <identifier type="DOI">10.</identifier>
    <internalIdentifier></internalIdentifier>
    <name language="eng" languageVocabulary="ISO-639-3">
      Database for Spoken German
    </name>
    <additionalName language="deu" languageVocabulary="ISO-639-3">
      Datenbank Gesprochenes Deutsch
    </additionalName>
    <repositoryURL>https://dgd.ids-mannheim.de/</repositoryURL>
    <repositoryType>disciplinary</repositoryType>
    <repositoryDescription language="eng" languageVocabulary="ISO-639-3">
      The "Database for Spoken German (DGD)" is a
      corpus management system in ...
    </repositoryDescription>
    <repositoryContent scheme="parse">Audiovisual data</repositoryContent>
    <repositoryContent scheme="parse">Standard office documents</repositoryContent>
    <repositoryContent scheme="parse">Structured text</repositoryContent>
    <recordCount></recordCount>
    <repositorySubject>Humanities and Social Sciences</subject>
    <rsubject code="104" scheme="DFG">Linguistics</subject>
    <subject code="11" scheme="DFG">Humanities</subject>
    <keyword>Folk</keyword>
    <rkeyword>Australian German</keyword>
    <organization>
      <name language="deu" languageVocabulary="ISO-639-3">
        Institut für Deutsche Sprache, Archiv für Gesprochenes Deutsch
      </name>
      <acronym>AGD</acronym>
      <id>ror...</id>
      <country vocabulary="">DEU</country>
      <url>http://agd.ids-mannheim.de/index.shtml</url>
    </organization>
    <software>other</software>
    <versioning>true</versioning>
    <apiUrl type="oai-pmh" documentationURL="https://www.openarchives.org/pmh/">https://...</apiUrl>
    <policyURL>https://</policyURL>
    <access type="open">dataAccess</access>
    <access type="restricted" restriction="fee">dataUpload</access>
    <licence type="databaseLicence" url="https://">CC-BY</licence>
    <licence type="dataLicence" url="https://">CC-BY</licence>
    <startDate format="YYYY-MM-DD">yyyy-MM-dd</startDate>
    <lastUpdate format="YYYY-MM-DD">yyyy-MM-dd</lastUpdate>
  </registryEntry>
</registryEntries>
```

</registryEntries>