Supplement A: common data model

This document describes the common data 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 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 common data model shared by all the registries as a base for their interoperability. In the definition of the common data 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 common data model.

Metadata properties

The following table lists the field of the proposed common data model. The column *Property* specifies the name of the field. The column *Definition* briefly describes the field. The column *A/C* is to specify if the field is an attribute or a child of another field. If 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: Common data model.

ID	Property	Definition	A/C	Occ	Notes
1	identifier	A globally unique identifier of the resource.		0-n	recordURI in the crosswalk
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	nameLanguage	The language of the repository name.	A	O	A code in the reference vocabulary, e.g., eng
3.2	languageVocabulary	The vocabulary of the language.	A	O	e.g., ISO-639-3. Desiderata
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	O	deu. Desiderata
4.2	languageVocabulary	The vocabulary of the language.	A	O	e.g., ISO-639-3. Desiderata
5	url	The home page of the repository.		1	e.g., https://dgd.ids-mannheim.de/ - repositoryURL in the crosswalk
					Continued on next page

Table 1 – continued from previous page

ID	Property	Definition	A/C	Occ	Notes
6	type	The type of the repository.		1-n	Controlled vocabulary, e.g., [institutional, disciplinary, governmental, multidisciplinary, project-related, etc.] - repositoryType in the crosswalk
7	description	The description of the repository.		0-1	Free text - repositoryDescription in the crosswalk
7.1	language	The language of the description.	A	O	e.g., deu. Desiderata
7.2	languageVocabulary	The vocabulary of the language.	A	O	e.g., ISO-639-3. Desiderata
8	content	The types of items in the repository.		0-n	Dependent on the schema or free text - repositoryContent in the crosswalk
8.1	scheme	The scheme of the content	A	O	Desiderata
9	recordCount	The number of resources in the repository		0-1	
10	subject	The subject of the repository		0-n	
10.1	code	The code of the subject	A	O	Dependent from the scheme. Desiderata

| | | | | | Continued on next page |

3

Table 1 – continued from previous page

ID	Property	Definition	A/C	Occ	Notes
10.2	scheme	The scheme of the subject name	A	O	Desiderata
10.3	uri	The URI of the subject	A	O	Desiderata
11	keyword	Keyword(s) and labels describing the subject focus of repository.		0-n	
12	organization	The organization(s) related to the repository		1-n	
12.1	name	The name of the organization	C		
12.1.1	nameLanguage	The language of the organization name	A	O	Desiderata
12.1.2	languageVocabulary	The vocabulary of the language	A	O	Desiderata
12.2	acronym	The acronym of the organization	C	0-1	
12.3	id	The id of the organization	C	0-n	e.g., 021nxhr62
12.3.1	type	The type of the id of the organization	A	O	e.g., ROR
12.4	country	The country of the organization	C	1	
12.4.1	code	The code of the country.	A	O	Desiderata
12.4.2	vocabulary	The vocabulary of the country code.	A	O	Desiderata
12.5	organizationUrl	The home page of the organization	C	1	

4

Table 1 – continued from previous page

ID	Property	Definition	A/C	Occ	Notes
12.6	type	The type of the organization	C	0..1	e.g. funding, Desiderata
12.7	role	The role of the organization	C	0..1	e.g. maintain, Desiderata
13	softwareName			1	
14	versioning	Is data versioning allowed?		1	True, false
15	apiUrl	The URL of the access point to the repository		0-1	
15.1	type	The type of the API	A	O	e.g., OAI-PMH
15.2	documentationUrl	URL to the API documentation	A		O
16	policyUrl	The URL of the policy		0-n	
17	access	The kind of access to the resource		0-n	e.g, dataAccess, dataUpload, /ldots
17.1	type	The type o f access	A	Req	e.g., open
17.2	restrictions	The type of restriction if access is not open	A	O	e.g., fee
18	licence	The kind of licence on the resource		0-n	e.g., dataLicence, databaseLicence, . . .
18.1	type	The type	A	Req	The type of licence, e.g., CC-BY
18.2	URL	The URL to the licence	A	Req	
19	startDate	The repository birth		0-1	
19.1	format	The format of the date	A	O	Desiderata
20	updateDate	The date of the last update of the repository		0-1	

Table 1 – continued from previous page

ID	Property	Definition	A/C	Occ	Notes
20.1	format	The format of the date	A	O	Desiderata

Supplement B: XML example

The following example documents the interoperability profile of a generic repository according to the proposed common data model

```xml
<registryEntries>
    <registryEntry>
        <identifier type="DOI">10.</identifier>
        <internalIdentifier>r3d100000004</internalIdentifier>
        <name language="eng" languageVocabulary="ISO-639-3">
                Database for Spoken German
        </name>
        <additionalName language="deu" languageVocabulary="ISO-639-3">
                Datenbank Gesprochenes Deutsch
        </additionalName>
        <URL>https://dgd.ids-mannheim.de/</repositoryURL>
        <type>disciplinary</type>
        <description language="eng" languageVocabulary="ISO-639-3">
            The "Database for Spoken German (DGD)" is a...
        </description>
        <content scheme="parse">Audiovisual data</content>
        <content scheme="parse">Standard office documents</content>
        <content scheme="parse">Structured text</content>
        <recordCount></recordCount>
        <subject>Humanities and Social Sciences</subject>
        <subject code="104" scheme="DFG">Linguistics</subject>
        <subject code="11" scheme="DFG">Humanities</subject>
        <keyword>Folk</keyword>
        <keyword>Australian German</keyword>
        <organization>
            <name language="deu" languageVocabulary="ISO-693-3">
                Institut für Deutsche Sprache, Archiv für Gesprochenes Deutsch
            </name>
            <acronym>AGD</acronym>
            <id type='ROR'>...</id>
            <country vocabulary="">DEU</country>
            <url>http://agd.ids-mannheim.de/index.shtml</url>
        </organization>
        <softwareName>other</software>
        <versioning>true</versioning>
        <apiUrl type="oai-pmh" documentationURL="https://www.openarchives.org/pmh/">https://..</api>
        <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>
        <updateDate format="YYYY-MM-DD">yyyy-MM-dd</updateDate>
    </registryEntry>
</registryEntries>
```