CESSDA ERIC Agenda 21-24, Tasks 21-22

Tools Pillar: Data access information upgrade

Data access information upgrade

D4-D6 Guidelines and updated specifications for CESSDA Service Provides to make data access interoperable

Document info

Dissemination Level	PU
Due Date of Deliverable	31/08/2022
Actual Submission Date	15/09/2022
Type	RE
Approval Status	Approved by Tools Working Group Leader Mari Kleemola
Version	V1.0
Number of Pages	p.1 – p.15
DOI	10.5281/zenodo.5554488

Version history

Version	Date	Comment	Revised by
1.0	15.09.2022	For review	Morten Jakobsen

Author List

Organisation	Name	Contact information
Sikt	Morten Jakobsen	morten.jakobsen@sikt.no

Peer-review

Organisation	Name	Contact information
Sikt	Hilde Orten	hilde.orten@sikt.no

Contents

Executive Summary

Abbreviations and Acronyms

CDC	CESSDA Data Catalogue
CESSDA	Consortium of European Social Science Data Archives
CMM	CESSDA Metadata Model
CMV	CESSDA Metadata Validator
CV	Controlled vocabulary
DDI	Data Documentation Initiative
DDI-C	DDI (see above) Codebook
DDI-L	DDI (see above) Lifecycle
MDO	CESSDA Metadata Office
SP	CESSDA Service Provider (an organisation, usually a national data archive, which is responsible for providing relevant services in the framework of CESSDA)

1. Introduction

An important part of improving both the user-friendliness of the CESSDA Data Catalogue[1] and the quality of metadata records harvested from Service Providers is to make data access information interoperable. As described by CESSDA Metadata Office 2020 Task 2[2] there is a nuanced variation of access conditions to research data among the SPs. These nuances can be attributed to the different legislations in the various countries and different praxis among the SPs. As shown by the report, it is possible to generalize the various access conditions implemented by SPs.

The main objective of this report is to give guidelines to SPs on how they can implement interoperable access conditions based on the specifications presented last year[3]. But after the publication of the report, discussions arose around the previously proposed metadata specifications which have identified problematic issues. Therefore, this report also reviews the proposed specifications and redefines them such that the initial issues are resolved. This means that the specifications given are no longer valid. SPs seeking to make their access

---

[1] CESSDA Data Catalogue, https://datacatalogue.cessda.eu/, [accessed:12.09.2022]

[2] Bolton, Sharon, & Jääskeläinen, Taina. (2021). Report on creating a CDC Publisher Names Vocabulary and investigating Data Access Interoperability (2.0). Zenodo. https://doi.org/10.5281/zenodo.4524429

[3] Bolton, Sharon, Jakobsen, Morten, & Storviken, Silje. (2021). Specification for interoperable access conditions in CDC. https://doi.org/10.5281/zenodo.5554430

conditions interoperable should instead implement the specifications, as they are presented in this report.

In addition to revising the specifications, the initial plan of creating a CESSDA specific data access controlled vocabulary is no longer deemed necessary. Instead, the suggestion is to implement an already existing and established CV for categorizing openness to digital resources.

2. Revised Specifications

The main issue with the proposed solution in "Specification for interoperable access conditions in CDC"[4] is that relevant metadata is placed outside of the metadata record. This impacts the validity and authority of the information in the record. The reason for placing the information outside the record was the lack of possibility to express the necessary CV within the Data Documentation Initiative framework for the relevant element. Implementing data access information through a controlled vocabulary is crucial to make data access interoperable. The difficulty of doing this within the DDI framework still exists, but new discussions have proposed a more feasible solution to the problem than what originally was thought. Below follows the specifications for how to make data access interoperable for CESSDA metadata records using DDI.

2.1 DDI Lifecycle

In DDI-Lifecycle it is possible to implement a CV in the newest version, 3.3. An issue with specifying it using version 3.3 is that the CESSDA Metadata Model[5] is specified in version 3.2. It is not optimal to have metadata specifications spanning across different versions of the standard, but in reviewing the structure of both 3.2 and 3.3, no major conflict for the relevant element has been identified.

In CMM two sub-elements of *Access*[6] is used to define data access. 1) *Description*[7], which is recommended to be used to express the possibilities of accessing data using free text and 2)

---

[4] Bolton, Sharon, Jakobsen, Morten, & Storviken, Silje. (2021). Specification for interoperable access conditions in CDC. https://doi.org/10.5281/zenodo.5554430

[5] Akdeniz, Esra, Borschewski, Kerrin, Moilanen, Katja, Jääskeläinen, Taina, Jakobsen, Morten, Storviken, Silje, Beeken, Jeannine, Bell, Darren, & Bolton, Sharon. (2021). CMM CESSDA Metadata Model (2.0). Zenodo. https://doi.org/10.5281/zenodo.4751455

[6] DDI Alliance, Access: https://ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/schemas/archive_xsd/elements/Access.html [accessed:12.09.2022]

[7] DDI Alliance, Description: https://ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/schemas/reusable_xsd/elements/Description.html [accessed:12.09.2022]

AccessConditions[8] which is recommended to include any information on restrictions of usage of the data, information here is also registered using free text.

DDI Lifecycle 3.3 allows the use of values from a controlled vocabulary, through the sub-element *TypeOfAccess*[9], alongside the elements presented in CMM. This then allows SPs to keep their archive specific access conditions in the elements specified in CMM and to categorize them into the new data access CV using the *TypeOfAccess* element.

In the event where SPs already use the *TypeOfAccess* element for an internal, or archive specific, CV alongside the *Description* element, the top element *Access* must be repeated to avoid violation of DDI 3.3's structure. It is not allowed within DDI 3.3 to have multiple occurrences of *TypeOfAccess* within an *Access* element. The recommendation is therefore that in the event where *TypeOfAccess* already is in use by SPs, a secondary top element for *Access* should be created that holds the relevant code value from the CV.

As *Description*/*Content* already has specified use in the CMM, these are excluded in specifying interoperable access conditions to avoid any conflicts and ease the implementation for SPs.

Below is an overview of the new elements to be implemented in DDI 3.3. For full x-paths, see Appendix 1.

Element	Type of Access
Element type	Content element
Description	A classification of the restrictions on data access.
Mandatory/Recommended/Optional	Recommended
Occurrence	0-1
Controlled vocabulary	Access Rights
Usage notes	Code value from the Controlled Vocabulary "Access Rights".

[8] DDI Alliance, Access conditions: https://ddialliance.org/Specification/DDI-Lifecycle/3.2/XMLSchema/FieldLevelDocumentation/schemas/archive_xsd/elements/AccessConditions.html [accessed:12.09.2022]

[9] DDI Alliance, Type of Access: https://ddialliance.org/Specification/DDI-Lifecycle/3.3/XMLSchema/FieldLevelDocumentation/schemas/archive_xsd/elements/TypeOfAccess.html [accessed:12.09.2022]

Element	Code list name
Element type	Attribute
Description	The name of the vocabulary used for Type of Access.
Mandatory/Recommended/Optional	Recommended
Occurrence	0-1
Controlled vocabulary	[Access Rights](#)
Usage notes	Attribute of element Type of Access. Use the string "Access Rights" regardless of language.

Element	Code list URN
Element type	Attribute
Description	Persistent identifier of the vocabulary for Type of Access.
Mandatory/Recommended/Optional	Optional
Occurrence	0-1
Controlled vocabulary	-
Usage notes	Attribute of element Type of Access.

2.2 DDI Codebook

As no suitable element for a CV for data access exists within DDI Codebook 2.5, a workaround needs to be implemented. Due to existing mappings from the CDC to other data

catalogues, the workaround needs to be placed within the element *restrctn*[10]*,* which is used for registering information on data access in the CDC DDI Codebook 2.5 profile[11].

The specifications for registering data access in the CDC DDI Codebook 2.5 profile is as follows: It is **recommended** for SPs to register free text information about data access in the *restrctn* element. The element can be repeated as needed. If information is registered, it is **mandatory** to include information on language using the xml:lang attribute. Either at the element level for multilingual records, or at top level for monolingual records.

In specifying a code value from the data access CV in the *restrctn* element, three considerations need to be made. 1) It must not interfere with existing metadata that SPs have registered in the element, 2) It must not violate the existing rules of the DDI profile, and 3) information needs to be registered in such a way that machines can distinguish the free text from the CV code value.

Satisfying 1) is done by repeating the element, this keeps the CV separated from the free text information. Repetition of the element is permitted within the DDI 2.5 specifications and the CDC DDI 2.5 profile.

In order to ensure 2) the xml:lang attribute needs to be included even though it does not provide any relevant information for the CV. If this is excluded from the metadata record the CESSDA Metadata Validator will not validate the information.

To enable the differentiation required from 3) the chosen solution is to introduce a new attribute to the *restrctn* element. This new attribute will work as a tag for consumers of metadata to indicate that the content of the element is a code value, instead of a free text value. In review of available attributes, *elementVersion* has been identified as the least intrusive and most viable for the purpose.

Below follows an overview of elements to be used for DDI 2.5 to ensure interoperable access conditions. Full x-paths can be found in appendix 2.

Element	Restrctn
Element type	Content element

---

[10] DDI Alliance, restrctn element:
https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation_files/schemas/codebook_xsd/elements/restrctn.html [accessed:12.09.2022]
[11]CESSDA Data Catalogue (CDC) DDI2.5 Profile:
https://cmv.cessda.eu/profiles/cdc/ddi-2.5/1.0.4/profile.xml [accessed:12.09.2022]

Description	Refers to any information relating to the access and use conditions of data that the SP wants to send.
Mandatory/Recommended/Optional	Recommended
Occurrence	0-n
Controlled vocabulary	[Access Rights](#)*
Usage notes	SPs need to concatenate the information for 'restrctn' the way they want it to appear in the CDC, if they use data from different elements e.g. specPerm, restrctn avlStatus etc.

*Element is to be repeated for categorization of data using the Access Rights CV. |

Element	Xml:lang
Element type	Attribute
Description	Language of the Access statement
Mandatory/Recommended/Optional	Mandatory
Occurrence	1
Controlled vocabulary	-
Usage notes	Attribute of element Restrctn. ISO 639-1 codes are strongly encouraged to be used. When registering code value information in 'restrctn', element needs to be populated.

Element	elementVersion
Element type	Attribute
Description	Indicator for CV information in Restrctn

Mandatory/Recommended/Optional	Recommended
Occurrence	0-1
Controlled vocabulary	-
Usage notes	Attribute of element Restrctn. To be populated with "Access Rights" when the content of the element Restrctn is a code value from the Access Rights vocabulary.

3. Data Access controlled vocabulary

As mentioned previously, an essential aspect of making data access interoperable within the CDC is through the creation of a CV for data access. One of the main objectives of the task was to define such a CV, this has been done[12]. Recent discussions amongst representatives from MDO, CESSDA MO, CDC and Tools pillar have reevaluated the need for a CESSDA specific data access CV. In creating a CESSDA specific CV additional steps in metadata transmissions are introduced. This is especially relevant for CDC's objective of providing metadata records from CESSDA SPs to other catalogues. To simplify this process the new proposal is to instead implement an already established CV. More concrete, the data access CV to be implemented to ensure interoperable data access information is the Access Rights CV[13] in use by OpenAIRE[14].

The change to the Access Rights CV has not been communicated to SPs until now. And the feedback and information given on the previous proposed CV will be taken into consideration with how SPs should implement the Access Rights CV in their metadata records. This is especially relevant for the proposed mapping of SPs access conditions to the CV, and how SPs are to categorize their research data. The main objective is to be able to differentiate between openly available data and data which requires applicants to satisfy some sort of criteria.

Table 1 gives an overview of the mapping between the previously proposed CV and the Access Rights CV from OpenAIRE. The initial sub-categories are removed, and from the

---

[12] Bolton, Sharon, Jakobsen, Morten, & Storviken, Silje. (2021). Specification for interoperable access conditions in CDC. https://doi.org/10.5281/zenodo.5554430
[13] Surfnet, Access rights:
https://wiki.surfnet.nl/display/standards/info-eu-repo#infoeurepo-AccessRights
[accessed:12.09.2022]
[14] OpenAIRE Guidelines for Data Archives » 16. Rights (MA)
https://guidelines.openaire.eu/en/latest/data/field_rights.html [accessed 13.10.2022]

Access Rights CV only relevant terms are included. This is done to simplify the implementation of interoperable access conditions.

Table 1: Data access category mapping

Code value Access Rights CV	Category from CESSDA CV	CESSDA CV definition	Access Rights CV definition
info:eu-repo/semantics/openAccess	Open	Free download with or without click-through terms of use and/or without registration	Open Access, this refers to access without restrictions, and without financial incentives. Access to the resource is gained directly, without any obstacles.
info:eu-repo/semantics/restrictedAccess	Restricted	Secure download, remote access or on-site use only, signed user contract/licence (registration or application)	Open Access, but with restrictions. Often the Author wants to control the people who have access to for example a dataset. Access is provided in exchange for an e-mail address, to keep in touch with other people who are interested

In a head-to-head examination of the CVs, the main difference is how restrictedAccess/Restricted is defined. In the Access Rights CV, the main means of getting access to the resource is through the provision of an email so that the author of the resource can contact end-users. The CESSDA CV definition includes this but expands with quite more comprehensive processes. Even though this distinction exists, the recommendation is that for the implementation of the Access Rights CV by SPs they should use the restrictedAccess code value as they would use the Restricted category from the previously communicated CV.

4. Implementation and the future

With new metadata specifications that keep the relevant information in the metadata record, and agreement on a controlled vocabulary that is already implemented by other data catalogues. SPs now have the necessary means of making their access conditions interoperable.

At the time this report was written, the new version of DDI 2.6 was published for public review. In moving forward, the specifications from the report here should be adopted into the new standard when SPs are ready to make the transition.

5. Appendix

The x-paths and profiles presented here are meant as examples to illustrate the DDI structure for relevant elements. The profile versions presented in the CMV documentation[15] are always the authoritative sources.

5.1 Appendix 1 DDI 3.3 profile addendum

```
<pr:Used

xpath="/ddi:DDIInstance/s:StudyUnit/a:Archive/a:ArchiveSpecific/a:Item/a:Access/r:TypeOf
Access"

                isRequired="false">
                <r:Description>
                        <r:Content>Required: Recommended</r:Content>
                        <r:Content>ElementType: Content element</r:Content>
                        <r:Content>Usage: Code value of the CV "Access Rights". Describes
the restrictions
                        on data access.</r:Content>
                </r:Description>
                <pr:Instructions>
                        <r:Content>
                        <![CDATA[
                        <Constraints>
                                <RecommendedNodeConstraint/>
                        </Constraints>
                        ]]></r:Content>
                </pr:Instructions>
        </pr:Used>
        <pr:Used
```

xpath="/ddi:DDIInstance/s:StudyUnit/a:Archive/a:ArchiveSpecific/a:Item/a:Access/r:TypeOf
Access/@CodeListName"

        isRequired="false">

        <r:Description>

            <r:Content>Required: Recommended</r:Content>

            <r:Content>ElementType: Attribute</r:Content>

            <r:Content>Usage: Name of the CV being used. Use the string
"Access Rights" regardless of language</r:Content>

        </r:Description>

        <pr:Instructions>

            <r:Content>

            <![CDATA[

            <Constraints>

                <RecommendedNodeConstraint/>

            </Constraints>

            ]]></r:Content>

        </pr:Instructions>

    </pr:Used>

    <pr:Used

xpath="/ddi:DDIInstance/s:StudyUnit/a:Archive/a:ArchiveSpecific/a:Item/a:Access/r:TypeOf
Access/@CodeListURN"

        isRequired="false">

        <r:Description>

            <r:Content>Required: Optional</r:Content>

            <r:Content>ElementType: Attribute</r:Content>

            <r:Content>Usage: The Canonical URL of the CV version that is being
used e.g.

https://wiki.surfnet.nl/display/standards/info-eu-repo#infoeurepo-AccessRights</r:Content>

```
            </r:Description>

            <pr:Instructions>

                    <r:Content>

                    <![CDATA[

                    <Constraints>

                            <OptionalNodeConstraint/>

                    </Constraints>

                    ]]></r:Content>

            </pr:Instructions>

    </pr:Used>
```

5.2 Appendix 2 DDI 2.5 profile addendum

```
        <pr:Used xpath="/codeBook/stdyDscr/dataAccs/useStmt/restrctn/@elementVersion"
isRequired="false">

            <r:Description>

                    <r:Content>Required: Recommended</r:Content>

                    <r:Content>ElementType: Attribute</r:Content>

                    <r:Content>Usage: To be populated with "Access Rights" to indicate
that the content of the

                    element 'restrctn' is a Code Value from the controlled vocabulary
Access Rights</r:Content>

            </r:Description>

            <pr:Instructions>

                    <r:Content>

                            <![CDATA[

                            <Constraints>

                                    <RecommendedNodeConstraint/>

                            </Constraints>
```

```
                    ]]>
            </r:Content>
        </pr:Instructions>
    </pr:Used>
```

6. References

Bolton, Sharon, & Jääskeläinen, Taina. (2021). Report on creating a CDC Publisher Names Vocabulary and investigating Data Access Interoperability (2.0). Zenodo. https://doi.org/10.5281/zenodo.4524429

Bolton, Sharon, Jakobsen, Morten, & Storviken, Silje. (2021). Specification for interoperable access conditions in CDC. https://doi.org/10.5281/zenodo.5554430

Akdeniz, Esra, Borschewski, Kerrin, Moilanen, Katja, Jääskeläinen, Taina, Jakobsen, Morten, Storviken, Silje, Beeken, Jeannine, Bell, Darren, & Bolton, Sharon. (2021). CMM CESSDA Metadata Model (2.0). Zenodo. https://doi.org/10.5281/zenodo.4751455