

CESSDA Work Plan 2020  
CESSDA Metadata Office Task 2

## D3: Report on creating a CDC Publisher Names Vocabulary and investigating Data Access Interoperability

### Document info

Dissemination Level	PU
Due Date of Deliverable	30/09/20
Actual Submission Date	02/11/20
Type	Report
Approval Status	Approved by CESSDA Tools & Services Working Group leader Mari Kleemola and CESSDA Technical Working Group leader John Shepherdson
Version	V2.0
Number of Pages	p.1 – p.17
DOI	10.5281/zenodo.4524429

## Version history

Version	Date	Comment	Revised by
<b>0.1</b>	04.09.2020	First draft produced by UKDS and FSD partners in MDO Task 2 team	UKDS/FSD
<b>0.2</b>	05.09.2020	Comments received from FSD	FSD
<b>1.0</b>	16.09.2020	First version produced by UKDS and FSD partners in CESSDA MDO Task 2 team	UKDS/FSD
<b>1.1</b>		Review by CESSDA MO	MO
<b>2.0</b>		Second version, edits made after MO and WGL review	UKDS/FSD

## Author List

Organisation	Name	Contact information
<b>UK Data Service</b>	Sharon Bolton	sharonb@essex.ac.uk
<b>Finnish Social Science Data Archive</b>	Taina Jääskeläinen	taina.jaaskelainen@tuni.fi

## Peer-review

Organisation	Name	Contact information
<b>UK Data Service</b>	Jeannine Beeken	jeannine.beeken@essex.ac.uk
<b>UK Data Service</b>	Lorna Balkan	balka@essex.ac.uk



## Contents

1. The CESSDA Data Catalogue	5
2. Creating the CDC Service Provider/Publisher Names vocabulary	5
3. Constructing a Data Access Interoperability CV	6
3.1 Background	6
3.2 Starting the modelling work	6
3.2.1 Data Access at the UK Data Service	7
3.2.2 Data Access at GESIS	9
3.2.3 Comparing data access at UKDS and GESIS	12
3.2.4 Researching data access criteria at other CESSDA SPs	14
3.3 Moving the model forward	14
3.4 Future plans	15
Appendix 1: CESSDA Controlled Vocabulary for CdcPublisherNames	16
CESSDA Controlled Vocabulary for CdcPublisherNames	16
CV definition	16
Details	16
Code List	17
Usage: Copyright and License	18
References	18

## Executive Summary

This report has been created for CESSDA ERIC 2020 Work Plan Task Metadata Office (MDO) Task 2 Deliverable 3. The Deliverable text is as follows: "Report on the creation of Service Provider CV in English in the CVS with appropriate supporting documentation. Work will also commence on a data access interoperability CV, delivering a short intermediate report on progress in identifying the different statuses and problems". Therefore, the Deliverable, and this report, covers two discrete tasks, both related to the CESSDA Data Catalogue (CDC). The first task, the creation of the Service Provider CV is complete, and its planning and execution is described in the report. The second task is more complex than the first, due to the diverse nature of data access conditions across CESSDA Service Providers. The CESSDA Data Access Policy notes that "Access conditions to data shall, by 2022, be fully interoperable" (Principle 11) (Woollard, L'Hours and Beedham, 2016). The work to be done towards this goal by MDO Task 2 during 2020 was to start the move towards interoperability by researching and identifying likely access conditions and bringing them together in broad tiers or classes of access type. The report describes this first stage of the work and indicates the next steps to be taken. Eventually, a Data Access Interoperability CV encompassing all CESSDA SP records will be constructed that will enable users to filter CDC search results by common data access criteria.

## Abbreviations and Acronyms

<b>ADP</b>	Slovenian Social Science Data Archives
<b>CDC</b>	CESSDA Data Catalogue
<b>CESSDA</b>	Consortium of European Social Science Data Archives
<b>CMV</b>	CESSDA Metadata Validator
<b>CV</b>	Controlled Vocabulary
<b>CVS</b>	CESSDA Vocabulary Service
<b>DDI</b>	Data Documentation Initiative
<b>DDI-L</b>	DDI (see above) Lifecycle
<b>FE</b>	Further Education
<b>FSD</b>	Finnish Social Science Data Archive
<b>GESIS</b>	GESIS- Leibniz-Institute for the Social Sciences
<b>HE</b>	Higher Education
<b>NSD</b>	Norwegian Centre for Research Data
<b>OGL</b>	Open Government Licence
<b>SND</b>	Swedish National Data Service
<b>SP</b>	CESSDA Service Provider (an organisation, usually a national data archive, which is responsible for providing relevant services in the framework of CESSDA).
<b>UK</b>	United Kingdom

## 1. The CESSDA Data Catalogue

The CESSDA Data Catalogue<sup>1</sup> (CDC) contains the metadata of data holdings of CESSDA's Service Providers and associated organisations. The catalogue allows search and discovery, enabling effective access to European social science research data.

Users of the catalogue are able to restrict their searches to certain chosen elements if the catalogue has relevant filters. To enable filtering, the catalogue relies on standardised metadata and the use of controlled vocabularies (CVs).

## 2. Creating the CDC Service Provider/Publisher Names vocabulary

A search filter that covers the names of organisations whose records are included in the CDC allows users to subset results by publisher. In the context of the CDC, 'Publisher' therefore refers to any organisation providing records for the catalogue. The CV must cover all publishing organisations, but not all may be CESSDA members; associated organisations may also be included. During the course of the work, the conscious decision was taken to change the name of this CV from the original 'Service Provider Names' stipulated in the Deliverable text, to 'CDC Publisher Names', in order to ensure all publishers could be covered now and, in the future, as more organisations contribute records to the CDC.

The filter is also needed because publisher names appear in many languages and in many variants across the metadata records included in the CDC. Organisation names may also have changed over time. Because of this variance, the CDC publisher filter needs to be based on a Controlled Vocabulary that can encompass all permutations and link each one to a specific publisher. To make the filter work in practice, the CDC system will attach each SP's endpoint to a publisher name in the vocabulary. The filter will become visible in the next release of the CDC software (planned for 2021).

The CDC User Group decided that the user interface of the data catalogue will only be available in English, though the metadata records within it may be in any language. This means that for the filter, and consequently also for the CV, publisher names are only required in English.

The CDC Publisher Names CV is now published in the CESSDA Vocabulary Service<sup>2</sup> (CVS) (it is also included in the Appendix of this document). Each organisation has approved the

---

<sup>1</sup> CESSDA Data Catalogue: <https://datacatalogue.cessda.eu/> (accessed 21 October 2020)

<sup>2</sup> CESSDA Vocabulary Service: <https://vocabularies.cessda.eu/#!discover> (accessed 21 October 2020)



name to be used. New organisations are added whenever a new publisher endpoint becomes available for the CDC.

### 3. Constructing a Data Access Interoperability CV

#### 3.1 Background

To meet the needs of CESSDA and its data user community by enabling easy access to data records across countries, the CESSDA Data Catalogue (CDC) has to be a useful and flexible tool. A range of search filters for data records is needed to ensure that the many thousands of records in the CDC are easily navigable and browsable. One desirable way of filtering records is by data access criteria, so that users can see easily whether they are likely to be able to gain access to the data in order to plan their research. Some data have very restrictive access conditions, making them beyond the reach of many researchers. The filter will also enable users to find data across CESSDA countries with similar or comparable access conditions. This can facilitate cross-national research; for example, enabling students to find comparable open use datasets that they can access quickly and easily for their coursework.

In order to construct a robust data access filter, some preparatory work on interoperability was needed. CESSDA Service Provider (SP) organisations vary in size and nature across countries and may have very different data access criteria. To include every record in the CDC in the filter, the controlled vocabulary must cover the full spectrum of each SP's access criteria. The following access categories were proposed:

- open data with no restrictions on access;
- data with some restrictions on access;
- data accessible only in certain very restricted conditions and/or secure locations.

The next step was to investigate the full spectrum of SP access conditions over time to see whether the three proposed categories identified for the data access filter model would work in practice.

#### 3.2 Starting the modelling work

To define the access spectrum, we started by comparing access conditions between two CESSDA SP data archives, the UK Data Service (UKDS) and GESIS – Leibniz-Institute for the Social Sciences (GESIS), based in Germany. These organisations were chosen because both of their large and varied collections include data across the full spectrum of likely access conditions. To start the work, both organisations provided details of their access conditions and these were compiled into tables for comparison.

### 3.2.1 Data Access at the UK Data Service

The UK Data Service has implemented three access tiers for data, governed by a Data Access Policy. The three tiers are: open, safeguarded and controlled. These categories are generic because they combine modes of access and conditions of use, and there are some overlaps and differences between the three major categories. The tiers are guided by legislation defining 'personal' data, including the UK Statistics and Registration Services Act, the Data Protection Act and the Digital Economy Act. According to the legislation, open data and safeguarded data are legally 'not personal'. The difference between 'open' and 'safeguarded' is that safeguarded data may have a residual risk of disclosure, and open data do not. Controlled data can be defined as personal and so their use is highly restricted. The broad criteria and the kind of data and requirements included for access are detailed in the tables below. Further information about the categories, the UKDS End User Licence<sup>3</sup> and Data Access Policy<sup>4</sup> is available on the UKDS website. See Table 1 for the UKDS data access table.

**Table 1: UK Data Service Access Levels**

Data Type	Access conditions
<b>Open data</b> <i>Teaching datasets, less detailed data</i>	Open access. Licencing usually Creative Commons (CC-BY) <sup>5</sup> or Open Government Licence (OGL) <sup>6</sup> .
<b>Safeguarded data</b>  <b>1. End User Licence (Safeguarded)</b>  <b>2. End User Licence (Safeguarded) with additional conditions</b>	<p>Requires UKDS registration (agreement to the UKDS End User Licence) and username and password.  UK academics and students login using their institutional username and password, UKDA registration available for non-UK users.</p> <p>Requires UKDS registration as above and username and password.  UK academics and students login using their institutional username and password.</p>

<sup>3</sup> UKDS End User Licence: <https://www.ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx> (accessed 21 October 2020)

<sup>4</sup> UKDS Data Access Policy: <https://www.ukdataservice.ac.uk/get-data/data-access-policy.aspx> (accessed 21 October 2020)

<sup>5</sup> Creative Commons Licences: <https://creativecommons.org/about/cclicenses/> (accessed 21 October 2020)

<sup>6</sup> Open Government Licence: <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/> (accessed 21 October 2020)

<b>3. Special Licence (Safeguarded)</b>	<p>Extra conditions must be met, e.g. access may be limited to academic users, or UK or EEA users, or output drafts submitted to data owner before publication.</p> <p>Requires UKDS registration as above and username and password. UK academics and students login using their institutional username and password. Express permission also required from the data owner for each user, guidance provided on safe data handling. May require a number of extra conditions to be met.</p>
<b>Controlled data</b> <i>Secure access only</i>	<p>Requires UKDS registration as above. Bespoke user credentials provided for secure data access after successful training completion. Access restricted to trained 'Accredited Researchers' under the Statistics and Registration Act/Digital Economy Act/GDPR. Access only available through a physical or virtual environment (remote Secure Lab service, or on-site safe room access)</p>

In addition, data access within the above categories at UKDS may depend on the type of user. Some users are not allowed to access certain data types, e.g. undergraduate students are not eligible to access Controlled data and commercial users are generally only allowed to access safeguarded data with express permission from the data owner.

**Table 2: UK Data Service User Types**

User type and location	<ul style="list-style-type: none"> <li>• Higher Education (HE)/Further Education (FE)/other</li> <li>• Non-HE/FE</li> <li>• UK</li> <li>• non-UK</li> </ul>
Data access conditions	<ul style="list-style-type: none"> <li>• End User Licence</li> <li>• Special Conditions</li> <li>• Special Licence</li> <li>• Secure Lab/Safe Room access only (Accredited Researcher)</li> </ul>
Project type	<ul style="list-style-type: none"> <li>• Commercial</li> <li>• Non-commercial</li> <li>• Teaching</li> </ul>



While the UKDS data access model is complex, it includes data across the access spectrum, meaning that it forms a good basis for modelling on which to create more simplified and broader categories for the data access filter.

### 3.2.2 Data Access at GESIS

The access spectrum at GESIS is broadly similar to UKDS, tiered within three major categories, each with two or three sub-categories that can accommodate legal issues (for example whether commercial use is permitted or not, the conditions of grant approvals, etc.). As at UKDS, the GESIS data acquisition team works with the data depositor to set up the correct access option for their data.

GESIS provide information on their current access categories and Usage Regulations<sup>7</sup> on their website. At present, they are set as follows:

- Category 0 - Data and documents are released for everybody;
- Category A - Data and documents are released for academic research and teaching;
- Category B - Data and documents are released for academic research and teaching, if the results are not published. If any publication or further processing of the results is planned, permission must be obtained by the Data Archive;
- Category C - Data and documents are only released for academic research and teaching after the data depositor's written authorization. For this purpose, the Data Archive obtains written permission with specification of the user and the analysis intention.

Category 0 broadly corresponds to open access with no restrictions; category A restricts use of the materials by user status; category B builds on category A's restriction by imposing further conditions; and category C restricts access still further by user type and bespoke permission for each case.

These categories may change in the future, as GESIS is running an exercise to map their currently used access options to Creative Commons (CC) licences (Bishop, 2019). Table 3 below shows proposed GESIS data access tiers, that may not yet be in active use. While the current categories noted above focus on type of user, the new tiers in the table are organised by type of licence, though they describe similar access restriction by credentials.

While the terminology is different and the GESIS exercise is still a work in progress, the important principle in terms of data access interoperability across the two organisations is that similarities can clearly be seen with the UKDS access levels and user type described in Tables 1 and 2 above.

---

<sup>7</sup> GESIS Usage Regulations: [https://www.gesis.org/fileadmin/user\\_upload/Usage\\_regulations.pdf](https://www.gesis.org/fileadmin/user_upload/Usage_regulations.pdf) (accessed 21 October 2020)

**Table 3: GESIS data access levels**

(Adapted from a table compiled by Libby Bishop of GESIS)

Category	Description	Access credentials	Sub category	Registration	Licence(s)
A	Open	Free download	A1	No	CC-BY
		Free download and click-through terms of use	A2	Yes	Attribution Open Access
B	Accountable	Secure download and signed user contract	B1	Yes	Attribution Commercial use allowed
			B2	Yes	Attribution Commercial use not allowed
C	Restricted	Off-site use and signed user contract	C1	Yes	Attribution Commercial use not allowed
		Remote access and signed user contract	C2	Yes	Attribution Commercial use not allowed
		On-site use and signed user contract	C3	Yes	Attribution Commercial use not allowed



### **3.2.3 Comparing data access at UKDS and GESIS**

The next step in the process was to compare the GESIS and UKDS information to identify common types of access, licensing and restrictions. The overall access levels and types are similar, though there are of course some differences between the countries. For example, some open access studies in the UKDS collection are covered by the UK Open Government Licence (OGL) rather than Creative Commons, though the nature of the OGL is similar to CC-BY. However, matching across the categories is extensive enough to provide a good basis for the construction of a prototype data access interoperability model that could work across CESSDA SPs. The combined access levels and their features are provided in Table 4 below.

**Table 4: Data Access Categories Comparison**

<b>Category</b>	<b>Access credentials</b>	<b>Registration</b>	<b>Licence(s)</b> <i>(non-commercial use always allowed)</i>
<b>Open</b>	Free download	No	Open (e.g. Creative Commons) Commercial use allowed
	Free download and click-through terms of use	No	Open Commercial use allowed
	Free download and click-through terms of use	Yes	Open Commercial use allowed
<b>Safeguarded/ Accountable</b>	Secure download and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use allowed
	Secure download and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use not allowed
	Secure download and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use allowed Additional conditions
	Secure download and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use not allowed Additional conditions
<b>Restricted</b>	Secure remote access and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use not allowed Additional conditions
	Secure remote access and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use allowed Additional conditions
	On-site use only and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use not allowed Additional conditions
	On-site use only and signed user contract/licence (registration or application)	Yes	Signed user contract/licence Commercial use allowed Additional conditions

### 3.2.4 Researching data access criteria at other CESSDA SPs

Once the comparison of UKDS and GESIS was complete and the combined table drafted, the next step was to research access categories and levels at some of the other CESSDA SP archives via their websites, to ensure that our approach was not restrictive and that we could better anticipate whether the combination of UKDS and GESIS categories would really work in practice. To do this we surveyed the websites of other SPs such as the Norwegian Centre for Research Data (NSD)<sup>8</sup>, the Swedish National Data Service (SND)<sup>9</sup>, the Finnish Social Science Data Archive (FSD)<sup>10</sup> and the Slovenian Social Science Data Archives (ADP)<sup>11</sup>. Between them, these organisations hold many different datasets across the social sciences, including register and official data as well as social surveys. Some organisations hold a larger proportion of open data than others; most of those surveyed so far require a form of registration (or supply data on user request/application) and differentiate between categories of user.

From findings so far, CESSDA SPs have similar criteria across the access spectrum, holding some open data and other data subject to tighter controls and restrictions. The terminology may be different, but our research so far strongly suggests that data access across all SPs do at least fall into the three broad categories identified by comparing UKDS and GESIS: Open; Safeguarded/Accountable; and Restricted. While some SPs do not have a registration system but supply on user request/application, and some may not have licences, or do not make a distinction between commercial and non-commercial use, the broad categories can still accommodate this.

Therefore, we are optimistic that the Data Access Categories Comparison table we have drafted from the work done so far does indeed provide a useful base model for data access interoperability without further modification at this stage. It encompasses all needs we have identified across the data access spectrum. It can be updated and revisited as work progresses.

## 3.3 Moving the model forward

The next stage of the work (and the real test of the model's viability) will be to share the Data Access Categories Comparison table with all CESSDA SPs to see if they can successfully map their existing data access levels to the broad main categories, and hopefully some of the subcategories. We plan to start this work in November 2020. This process will provide information on any further access conditions we may not have considered that need to be added to the model.

---

<sup>8</sup> Norwegian Centre for Research Data: <https://nsd.no/nsd/english/index.html> (accessed 21 October 2020)

<sup>9</sup> Swedish National Data Service: <https://snd.gu.se/en> (accessed 21 October 2020)

<sup>10</sup> Finnish Social Science Data Archive: <https://www.fsd.tuni.fi/en/> (accessed 21 October 2020)

<sup>11</sup> Slovenian Social Science Data Archives: <https://www.adp.fdv.uni-lj.si/eng/> (accessed 21 October 2020)



It is important to say at this point that the creation of a Data Access Interoperability CV is **not** intended to make any SP change their access conditions, but to make sure that all organisations can map their categories to the CV. The successful conclusion of this work and production of the completed model will allow SPs to undertake the technical work needed to include their records in the CDC data access filter.

### 3.4 Future plans

The final goal is to add the completed Data Access Interoperability filter to the CESSDA Data Catalogue. However, this cannot be done immediately, as some elements needed to make this possible are not yet available. For example, the inclusion of the filter depends on finding a matching DDI element for the CESSDA Metadata Validator (CMV). At present there does not seem to be a suitable element in DDI 2.5, but this requirement has been taken forward in the development process for DDI 2.6<sup>12</sup>. Similarly, a suitable element will need to be found in the DDI-L schema, although it may be possible to use the substitute `dc:AccessRights` or `dc:Rights` elements for Dublin Core (`dc`)<sup>13</sup> terms. All this of course depends on the work outlined above; the successful development of a Data Access Interoperability model that suits CESSDA. In addition, as organisations outside CESSDA begin to add their metadata records to the CDC, we will have to revisit and expand the data access interoperability model, adding extra categories as needed.

---

<sup>12</sup> 'Data access vocabulary element?' <https://ddi-alliance.atlassian.net/browse/DDICODE-72> (accessed 21 October 2020)

<sup>13</sup> Dublin Core Metadata Initiative: <http://purl.org/dc/terms/accessRights> (accessed 21 October 2020)



## Appendix 1: CESSDA Controlled Vocabulary for CdcPublisherNames

(Note: style slightly modified to suit document format)

### CESSDA Controlled Vocabulary for CdcPublisherNames

---

#### CV definition

Names of organisations providing metadata for CESSDA Data Catalogue.

#### Details

**CV short name:**

CdcPublisherNames

**CV name:**

CDC Publisher Names

**Language:**

English (en)

**CV notes:**

Used for the Publisher filter in CDC. No need for other language versions.

**Version:**

3.0

**Version notes:**

Organisation official name change.

**Version changes:**

Code value changed: SODA changed into SODHA

Code descriptive term rephrased: SODA changed into Social Sciences and Digital Humanities Archive (SODHA)

**Canonical URI:**

<urn:ddi:int.cessda.cv:CdcPublisherNames>

**Canonical URI of this version:**

<urn:ddi:int.cessda.cv:CdcPublisherNames:3.0>

**Agency:**

[CESSDA](#)

## Code List

Code value	Code descriptive term	Code definition
AUSSDA	Austrian Social Science Data Archive (AUSSDA)	
CSDA	Czech Social Science Data Archive (ČSDA)	
DNA	Danish National Archives (DNA)	
DANS	DANS-KNAW	Netherlands
FORS	FORS - Swiss Centre of Expertise in the Social Sciences	
FSD	Finnish Social Science Data Archive (FSD)	
GESIS	GESIS - Leibniz Institute for the Social Sciences	Germany
NSD	NSD - Norwegian Centre for Research Data	
APIS	Portuguese Archive of Social Information (APIS)	
ProgedoCDSP	PROGEDO: Center for Socio-Political Data (CDSP)	France
ProgedoCNRS	PROGEDO: CNRS	France
ProgedoSciencesPo	PROGEDO: Sciences Po	France
SASD	Slovak Archive of Social Data (SASD)	
ADP	Slovenian Social Science Data Archives (ADP)	
SODHA	Social Sciences and Digital Humanities Archive (SODHA)	Belgium
SoDaNet	SoDaNet - Greek Research Infrastructure for Social Science	
SND	Swedish National Data Service (SND)	
UKDS	UK Data Service	
UniData	UniData - Bicocca Data Archive	Italy

## Usage: Copyright and License

Copyright © [CESSDA](#) 2020.

Citation: CESSDA. (2020). *CDC Publisher Names* (Version 3.0) [Controlled vocabulary].

urn:ddi:int.cessda.cv:CdcPublisherNames:3.0. Available from:

<https://vocabularies.cessda.eu/urn:urn:ddi:int.cessda.cv:CdcPublisherNames:3.0>





## References

Bishop, L., *Reforming Data Access Categories at GESIS*, presented at the CESSDA Expert Seminar 2019 (CES2019), UK Data Service, University of Essex, 26 September 2019.

Woollard, Matthew, L'Hours, Hervé, & Beedham, Hilary. (2016, June 30). *CESSDA Data Access Policy* (Version 1). Zenodo. <http://doi.org/10.5281/zenodo.4054793>