

# Introduction to the CESSDA Vocabulary Service

---



Taina Jääskeläinen, Finnish Social  
Science Data Archive

**SSHOC Online Information Session**  
30 September, 2020



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782



# Contents

- Overview of CESSDA Vocabulary Service
- Technical infrastructure
- Tool functionalities and purpose
- Access to the tool for other domains
- Some screenshots

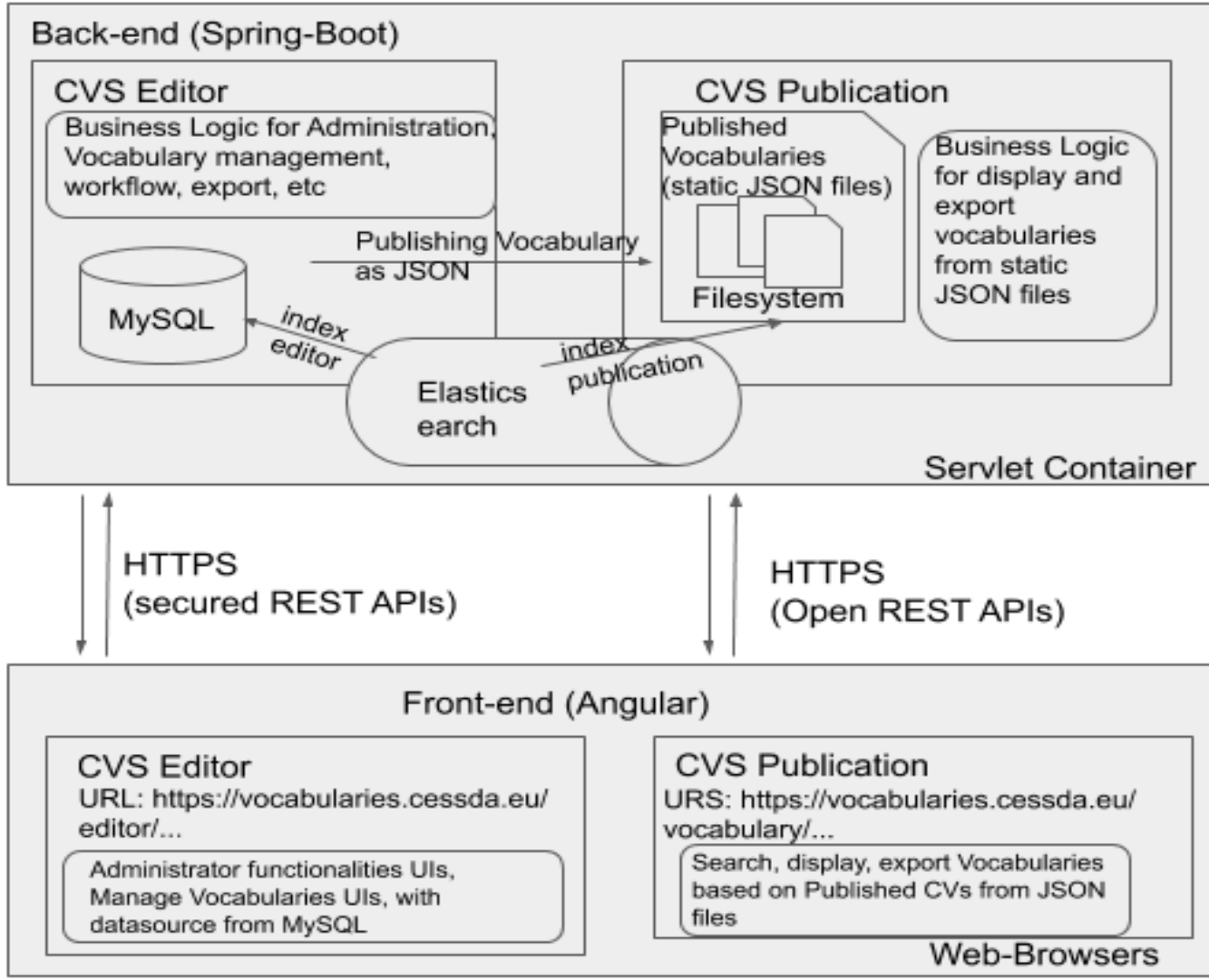


# CESSDA Vocabulary Service: Overview

- Funded by Consortium of Social Sciences Data Archives (CESSDA)
- Developed by GESIS
- User requirements: UK Data Service, Finnish Social Science Data Archive (FSD) and DDI Alliance
- User management FSD
- Currently 27 vocabularies created by CESSDA and DDI Alliance, with up to eleven language versions
- <https://vocabularies.cessda.eu>

# CVS2 Technical Infrastructure

- CVS1 had mixed front-end and back-end => testing and code maintenance difficult.
- New architecture for CVS2
  - Back-end (Java Spring-Boot)
  - Front-end (Angular 9)
- Improvements
  - Improved code quality
  - Better usability, flexibility/testability, security
  - Better scalability, reliability, extensibility and portability



# Browsing interface

- Browse and search published vocabularies in all language versions
- Download/export in **SKOS/RDF, HTML and PDF**
- **Concept-level information:** code value, descriptive term and definition
- Human-readable **version history** for all language versions
- **Comparison table** between previous and current version
- **Open source Rest APIs**
- **Challenge:** resolution of URLs

# Editor for registered users

- **Vocabulary management:** creation, translation and versioning
- **User management** and agency management
- **Access controlled by roles** which are specified by language and agency
- One user can have two or more roles
- Online access from different locations, simultaneous users allowed, no installations needed
- No knowledge of SKOS required

# Editor for registered users

- Clones previous versions for new version drafts, in all languages
- **Version history** produced by the system but editing human readable version history possible
  - change logs of vocabularies
  - comparison table between previous and current version
- Checks that mandatory elements are there before publication allowed
- Code values are the same across language versions, so are not translated. Descriptive terms must be translated (mandatory) but definition translations are not enforced (recommended).



# Editor for registered users

- Download/export of draft vocabularies can be exported for **Review**
- At publication, license, URN, citation and copyright produced by the system from info entered elsewhere in the Editor



# Main goals

- Tool for **controlled vocabularies**, not for thesauri!
- Broader and narrower terms, no related terms or synonyms
- Primary use cases were controlled vocabularies created for standardising metadata across datasets and organisations, particularly methodological
- **Harmonised metadata** needed for cross-national, multilingual data catalogues → the importance of version history and concept descriptions
- Change in a concept may mean changes in other concepts in metadata, so the **CV in table format** is useful for the users

# Access for other domains

- **Host own vocabularies** within the tool as an agency, CESSDA provides access
- CESSDA provides the platform also for **other domains**
- Software can run on you own infrastructure as **docker container**
- Software can be downloaded from **Bitbucket** to be adopted to own requirements
  - Example: GESIS TheSoz Thesaurus Manager

27 results found

RELEVANCE

A-Z

Agency

Language

**Aggregation Method** [AggregationMethod](#)

Identifies the type of aggregation used to combine related categories, usually within a common branch of a hierarchy, to provide information at a broader level than the level at which detailed observations are taken. (From: The OECD Glossary of Statistical Terms)

Version: 1.1 [Download](#)

EN

DA

**Analysis Unit** [AnalysisUnit](#)

Describes the entity being observed or analyzed in the study or in the variable.

Version: 2.0 [Download](#)

EN

DA

FI

FR

IT

NO

PT

SR

SV

**CDC Publisher Names** [CdcPublisherNames](#)

Names of organisations providing metadata for CESSDA Data Catalogue.

Version: 3.0 [Download](#)

EN

**CESSDA Persistent Identifier Types** [CessdaPersistentIdentifierTypes](#)

Lists the types of persistent identifiers that CESSDA accepts as study level PIDs in its data catalogue.

5 Vocabularies

Results per page

30



Agency



Language



Search languages

**Méthode temporelle** TimeMethod

Décrit la dimension temporelle de la collecte des données

Version: 1.2.2 [Download](#)Longitudinale : tendance / transversale répétée : Exemples : **Enquête** sociale européenne (ESS), des **enquêtes** nationales longitudinales sur la criminalité ...

EN

DA

FI

FR

IT

NO

PT

SR

SL

SV

**Procédure d'échantillonnage** SamplingProcedure

Une typologie des méthodes d'échantillonnage

Version: 1.1.3 [Download](#)**Probabiliste : stratifié** : Par exemple, dans une **enquête** nationale d'opinion publique la population entière est divisée en deux ...**Probabiliste : aléatoire systématique** : Par exemple, une **enquête** dans une entreprise porte sur un échantillon de 1 000 employés sur un total ...**Probabiliste : stratifié : disproportionné** : d'unités échantillonnées à partir de chaque strate peut être égal, optimal, ou peut refléter la raison de **l'enquête** ...

EN

DA

FI

FR

IT

PT

SR

SL

SV

**Unité d'analyse** AnalysisUnit

Décrit l'entité analysée dans l'étude ou dans la variable.



## CV definition

The procedure, technique, or mode of inquiry used to attain the data.

## Details

<b>CV short name:</b>	ModeOfCollection
<b>CV name:</b>	Mode Of Collection
<b>CV notes:</b>	This vocabulary was first published by the DDI Alliance. Please see: <a href="https://ddialliance.org/controlled-vocabularies/all">https://ddialliance.org/controlled-vocabularies/all</a> .
<b>Language:</b>	English (en)
<b>Version:</b>	3.0
<b>Version notes:</b>	
<b>Version changes:</b>	
<b>Canonical URI:</b>	<a href="urn:ddi:int.ddi.cv:ModeOfCollection:3.0">urn:ddi:int.ddi.cv:ModeOfCollection:3.0</a>
<b>Agency:</b>	<a href="#">DDI Alliance</a>

## Code List

Code value	Code descriptive term	Code definition
Interview	Interview	A pre-planned communication between two (or more) people - the interviewer(s) and the interviewee(s) - in which information is obtained by the interviewer(s) from the interviewee(s). If group interaction is part of the method, use "Focus group".
Interview.FaceToFace	Face-to-face interview	Data collection method in which a live interviewer conducts a personal interview, presenting questions and entering the responses. Use this broader term if not CAPI or PAPI, or if not known whether CAPI/PAPI or not.
Interview.Telephone	Telephone interview	Interview administered on the telephone. Use this broader term if not CATI, or if not known whether CATI or not.
Interview.Email	E-mail interview	Interviews conducted via e-mail, usually consisting of several e-mail messages that allow the discussion to continue beyond the first set of questions and answers, or the first e-mail exchange.
Interview.WebBased	Web-based interview	An interview conducted via the Internet. For example, interviews conducted within online forums or using web-based audio-visual technology that enables the interviewer(s) and interviewee(s) to communicate in real time.





▼ **Compare previous version en-1.0 with version en-2.0**

Only Show Lines with Differences (57)

Side by Side Line by Line

1	Cv Name: Analysis Unit
2	- Cv Def: Describes the entity being analyzed in the study or in the variable.
3	- Cv Notes: null
4	
5	
6	Code: Individual
7	Code Term: Individual
8	- Code Def: Any individual person, irrespective of demographic characteristics, professional, social or legal status, or affiliation
9	
10	- Code: Organization
11	- Code Term: Organization
12	- Code Def: Any kind of formal administrative and functional structure - includes associations, institutions, agencies, business organizations, etc.
13	
14	Code: Family
15	Code Term: Family
16	- Code Def: Two or more people related by blood, marriage (including step-relations), adoption or fostering and who may live together

1	Cv Name: Analysis Unit
2	+ Cv Def: Describes the entity being observed or analyzed in the study or in the variable.
3	+ Cv Notes: This vocabulary was first published by the DDI Alliance. Please see: <a href="https://ddialliance.org/controlled-vocabulary">https://ddialliance.org/controlled-vocabulary</a>
4	
5	
6	Code: Individual
7	Code Term: Individual
8	+ Code Def: Any individual person.
9	
10	+ Code: OrganizationOrInstitution
11	+ Code Term: Organization/Institution
12	+ Code Def: Any kind of formal administrative and functional structure - includes associations, institutions, agencies, business organizations, etc.
13	+
14	
15	Code: Family
16	Code Term: Family
17	+ Code Def: Two or more people related by blood, marriage (including step-relations), co-habitation or registered partnership
18	+





## DDI Alliance Controlled Vocabulary for Analysis Unit

CV name	Analysis Unit
CV short name	AnalysisUnit
CV definition	Describes the entity being analyzed in the study or variable. This vocabulary can also be used to describe data collected. The unit of observation can be the same as, or different from the unit of analysis.
CV notes	This vocabulary was first published by the DDI Alliance. Please see: <a href="https://ddialliance.org/controlled-vocabulary">https://ddialliance.org/controlled-vocabulary</a>
Language	English (en) <span style="float: right;">Version 3.0</span>

EN

Details

Versions

Identity and general

Usage

License and Citation

Export/Download

### ▼ en: 3.0 Published: 2020-08-19

#### Version changes

CV definition rephrased.

Code added: MediaUnit.Text.Book

Code added: Graphics

Code deleted: MediaUnit.Video

Code value changed: Object changed into Objects

Code descriptive term rephrased: Objects

Code definition rephrased: Objects

Code definition rephrased:HousingUnit

► Compare previous version en-2.0 with version en-3.0





# CESSDA Vocabulary Service: Translation Use Case

---



Taina Jääskeläinen, Finnish Social  
Science Data Archive

**SSHOC Online Information Session**  
30 September, 2020



This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782



# Translation needs

- **CESSDA Metadata Model** includes a number of controlled vocabularies
- Aim is to **standardise metadata** across data collections and languages, for tools like CESSDA Data Catalogue and European Question Bank
- Many CVs are **categorisations of research methods** used to create the data. Often provided by DDI Alliance.
- Metadata providers need these vocabularies in their metadata languages. CVS offers a tool for maintaining such **multilingual vocabularies**

Edit Finnish (fi) code translation for **RegistersRecordsAccounts.VotingResults**



**Code (en)** RegistersRecordsAccounts.VotingResults **Language** Finnish (fi)

**Descriptive term (en)** Registers/Records/Accounts: Voting results

**Descriptive term (fi) \***

**Definition (en)** Details of votes cast for candidates in an election (also known as 'election returns/results'), parliamentary or organizational votes on policies, motions or proposals, public referendums, etc.

**Definition (fi)**

Change notes

**Type \***

**Description**

- Code descriptive term rephrased
- Code definition amended with meaning change
- Code definition rephrased

## CV definition

Framgangsmåten, teknikken, eller metoden som er anvendt for å samle inn data.

## Details

<b>CV short name:</b>	ModeOfCollection
<b>CV name:</b>	Innsamlingsmetode
<b>Language:</b>	Norwegian (no)
<b>Version:</b>	3.0.1
<b>Version notes:</b>	
<b>Version changes:</b>	
<b>Canonical URI:</b>	<a href="urn:ddi:int.ddi.cv:ModeOfCollection:3.0">urn:ddi:int.ddi.cv:ModeOfCollection:3.0</a>
<b>Agency:</b>	<a href="#">DDI Alliance</a>
<b>Translator Agency:</b>	<a href="#">NSD - Norwegian Centre for Research Data</a>

## Code List

Code value	Code descriptive term	Code definition
Interview	Intervju	En forhåndsbestemt kommunikasjon mellom to (eller flere) personer – intervjueren(e) og intervjuobjekt(et/ene) – hvor intervjueren(e) henter informasjon fra intervjuobjekt(et/ene). Dersom gruppeinteraksjoner er del av metoden, bruk 'Fokusgruppe'.
Interview.FaceToFace	Personlig intervju	Datainnsamlingsmetode hvor en tilstedeværende intervjuer gjennomfører et personlig intervju, presenterer spørsmål og registrerer svarene. Bruk dette bredere begrepet dersom det ikke er CATI eller PAPI, eller dersom det er uvisst om det er CATI/PAPI.
Interview.Telephone	Telefonintervju	Intervju som gjennomføres over telefon. Bruk dette bredere begrepet dersom det ikke er CATI, eller dersom det er uvisst om det er CATI.
Interview.Email	E-post intervju	Intervjuer gjennomført over e-post, vanligvis bestående av en rekke e-poster som gjør at diskusjonen kan fortsette utover de første settene med spørsmål og svar, eller den første e-postutvekslingen.
Interview.WebBased	Webbasert intervju	Intervju gjennomført over internett. For eksempel, intervjuer gjennomført på internettforum eller ved bruk av internettbasert audiovisuell teknologi som muliggjør fortløpende kommunikasjon mellom intervjuer(e) og Intervjuobjekt(er).
Interview.FaceToFace.CAPIorCAMI	Personlig intervju: CAPI/CAMI	Dataassistert personlig intervjuing (CAPI). Datainnsamlingsmetode hvor intervjueren leser opp spørsmål for respondenten fra skjermen til en datamaskin, bærbar datamaskin, smarttelefon eller nettbrett, og registrerer svarene på det samme apparatet. Administreringen av intervjuet er styrt av et spesialdesignet program eller en applikasjon.
Interview.FaceToFace.PAPI	Personlig intervju: PAPI	Penn og papir intervju. Intervjueren bruker et tradisjonelt papirspørreskjema for å lese opp spørsmålene og til å registrere svarene.
Interview.Telephone.CATI	Telefonintervju: CATI	Dataassistert telefonintervju. Intervjueren stiller spørsmål ifølge instruksjoner fra en datamaskin, registrerer svarene direkte på datamaskinen og administreringen

# FSD workflow for dataset descriptions

- No metadata editor, **xml templates are used**
- All dataset descriptions are produced both in Finnish and in English.
- First Finnish, then English based on the Finnish one
- Bit of code to **automatically translate vocabulary terms** into English at publication, based on the vocabularies maintained in CVS



# FSD workflow

- Using vocabularies in this way in **metadata**:
  - Reduces effort required to produce description in English
  - Prevents mistakes in translation
- **Small language**: most users of CESSDA Data Catalogue would find FSD data through the English description



## FSD3388 manually produced xml description in Finnish

```
<dataColl>
  -<timeMeth>
    Pitkittäisaineisto: trendi/toistuva poikkileikkausaineisto
    <concept>Longitudinal.TrendRepeatedCrossSection</concept>
  </timeMeth>
  <collMode>Itsetäytettävä lomake: verkkolomake</collMode>
  <resInstru>Strukturoitu lomake</resInstru>
```

## FSD3388 system-produced description in English

```
<dataColl>
  -<timeMeth>
    Longitudinal: Trend/Repeated cross-section
    <concept>Longitudinal.TrendRepeatedCrossSection</concept>
  </timeMeth>
  <collMode>Self-administered questionnaire: Web-based (CAWI)</collMode>
  <resInstru>Structured questionnaire</resInstru>
```



# Thank you for your attention!

Join our community



<https://www.sshopencloud.eu>



@SSHOpenCloud



[info@sshopencloud.eu](mailto:info@sshopencloud.eu)



/in/sshopencloud

