

Data Description and Metadata

What it takes to produce a good one?

CESSDA Training Local Event

FSD and DNA

10.5281/zenodo.5807607

 [cessda.eu](https://www.cessda.eu)

 [@CESSDA_Data](https://twitter.com/CESSDA_Data)



Licence: CC-BY 4.0

Practical information

TIME: 2½ hours reserved – but we finish when ready!

QUESTIONS: short questions after each presentation, discussion at the end (raise hand & unmute when asked to)

Feel free to use our shared notes document (link in the chat)

Agenda

- Introduction to CESSDA and some tools and services
 - Tuomas J. Alaterä
- The significance of data description quality for cross-national data catalogues
 - Taina Jääskeläinen
- Describing data at FSD
 - Emilia Hakkola
- Metadata in reuse: harvesting, licensing, repurposing and FAIR
 - Tuomas J. Alaterä
- Q & A



Taina Jääskeläinen
Emilia Hakkola
Tuomas J. Alaterä

Finnish Social Science Data Archive

What is CESSDA?

Consortium of European Social Science Data Archives

- Long tradition of cooperation even before becoming an ERIC

European Research Infrastructure Consortium

- Legal form for facilitating establishment and operation of RI's within European interest

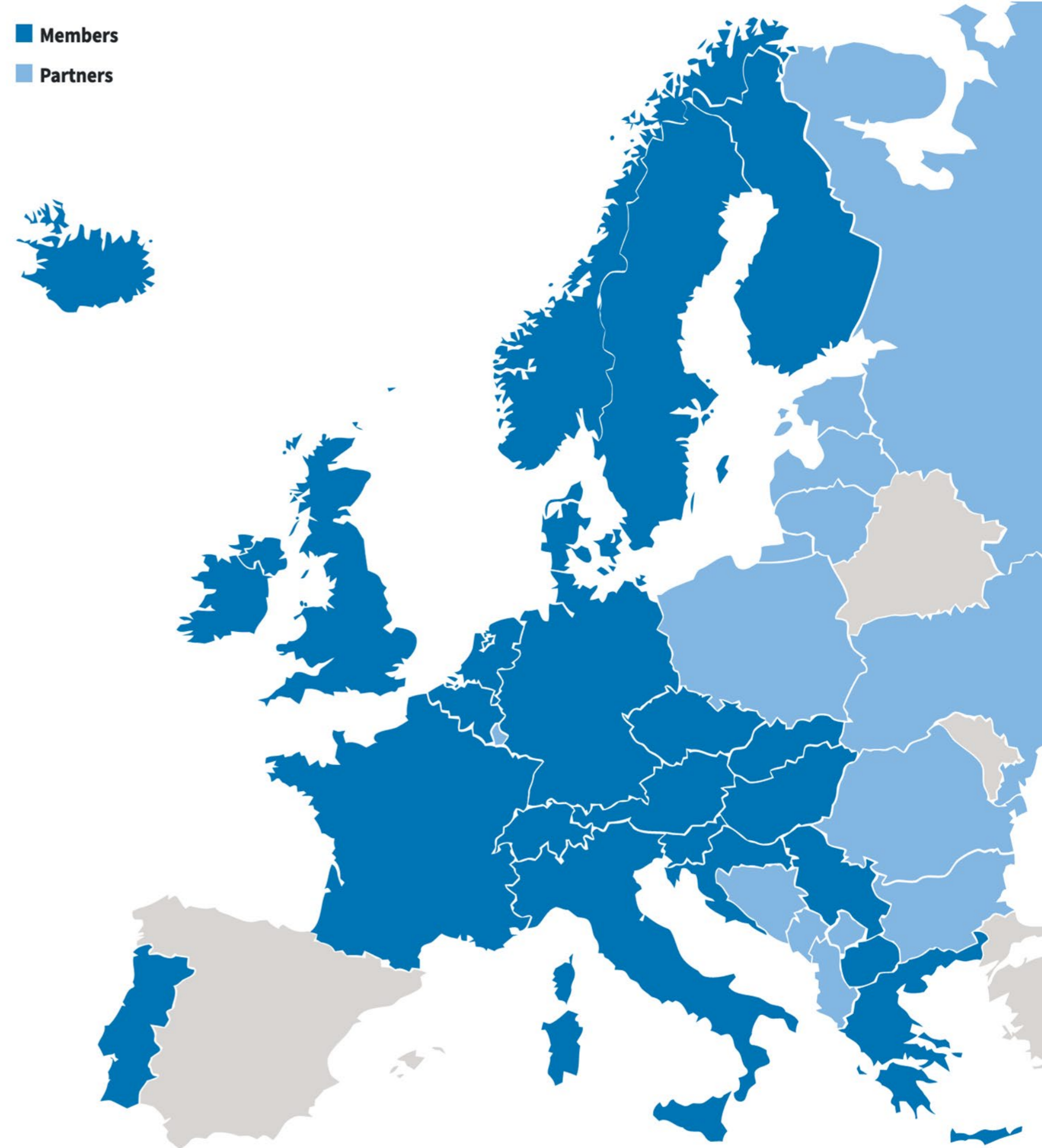
CESSDA provides and develops tools and services for both data **producers** and data **users** in social sciences and related fields

Coordinates participation in SSH fields for many EC funded projects

Service providers located in each member country

Main Office in Bergen, Norway

See: cessda.eu



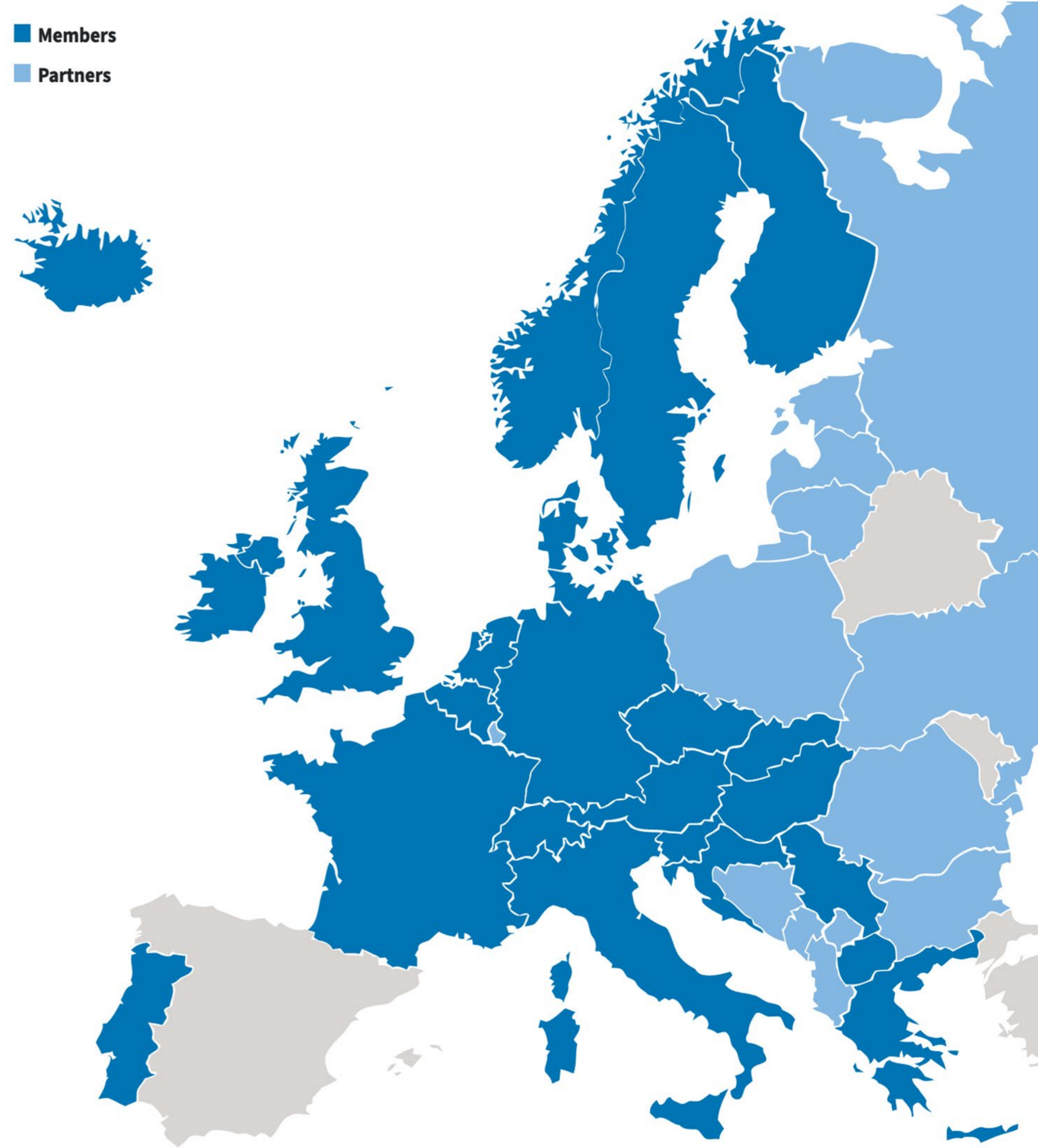
Our vision is that the provision of **access to social science data** and metadata is vital – for both science and society.

For this we must offer **services to data producers** to easily describe and store their data – if needed in a secured environment.

We will adhere to the **FAIR** (Findable, Accessible, Interoperable, Reusable) data principles to make data findable and provide information about the data, where they are, how they can be accessed.

We will also focus on providing **training** and enabling the transfer of expertise and sharing of knowledge on data, as well as relevant rules and regulations.

■ Members
■ Partners





Mission of CESSDA

- Provide a distributed and sustainable **research infrastructure**
 - **enabling** the research community to conduct high-quality research in the social sciences,
 - contributing to the production of effective solutions to the **major challenges facing society today.**
- Facilitate **teaching and learning** in the social sciences.

Tools & services





Data Management Expert Guide

This guide is designed by European experts to help social science researchers make their research data Findable, Accessible, Interoperable and Reusable (FAIR).

You will be guided by different European experts who are - on a daily basis - busy ensuring long-term access to valuable social science datasets, available for discovery and reuse at one of the [CESSDA social science data archives](#).

You can [download](#) the full DMEG for your personal study offline (DOI: [10.5281/zenodo.3820473](https://doi.org/10.5281/zenodo.3820473)). PDFs for every [single chapter](#) are also available for being printed as handouts for training.

Search this guide

Search

Data Management Expert Guide

- 1. Plan
- 2. Organise & Document
 - Designing a data file structure
 - Organisation of variables
 - File naming and folder structure
 - Documentation and metadata**
 - Adapt your DMP: part 2
 - Sources and further reading
- 3. Process
- 4. Store
- 5. Protect
- 6. Archive & Publish
- 7. Discover
- 8. Contributors

Documentation and metadata

I have never documented my data before. I have both qualitative and quantitative data and I work on a collaborative project. Where do I start?



⊖ How to start?

- 1. Do not panic. Much documentation is simply good research practice, so you are probably already doing much of it.*
- 2. Start early! Careful planning of your documentation at the beginning of your project helps you save time and effort. Do not leave the documentation for the very end of your project. Remember to include procedures for documentation in your data management planning.*
- 3. Think about the information that is needed in order to understand the data. What will other researchers and re-users need in order to understand your data?*
- 4. Create a separate documentation file for the data that includes the basic information about the data. You can also create similar files for each data set. Remember to organise your files so that there is a connection between the documentation file and the data sets.*
- 5. Plan where to deposit the data after the completion of the project. The repository probably follows a specific metadata standard that you can adopt.*
- 6. Document consistently throughout the project. Data documentation gives contextual information about your dataset(s). It specifies the aims and objectives of the original project and harbours explanatory material*

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>

Significance of data description quality for cross-national data catalogues

Taina Jääskeläinen, Content contact for CESSDA Data Catalogue and CESSDA Vocabulary Service / Finnish Social Science Data Archive

Data Description and Metadata – What does it take to produce a good one?

8 December 2021

 cessda.eu

 @CESSDA_Data



Licence: CC-BY 4.0

Overview

- ◇ Introduction to CESSDA Data Catalogue
- ◇ Cross-national catalogue functionalities
- ◇ Benefits of standardised metadata and controlled vocabularies for discovery
- ◇ Short introduction to CESSDA Vocabulary Service and ELSST thesaurus

CESSDA Data Catalogue

- ◊ Main goal: make European social science research data findable
- ◊ At the moment:
 - ◊ Descriptions of over 33,000 research datasets
 - ◊ Provided by CESSDA national service providers from 15 countries
 - ◊ Mostly social sciences but also data from health sciences and humanities
 - ◊ Quantitative, qualitative and mixed modes data
 - ◊ Study level metadata, no variable level information

[Reset filters](#)[Clear search](#)

43 studies found in English from a total of 33767

[User Guide](#)[About](#)

▶ Topic



▶ Collection years



▶ Country



▶ Publisher



Results per page

30

Sort by

Date of publication (newest)



1

2



Trend Questions Corona (Week 20/2021)

Presse- und Informationsamt der Bundesregierung, Berlin

On behalf of the Press and Information Office of the Federal Government, the opinion research institute forsa has regularly conducted representative population surveys on the subject of the 'Corona crisis' (COVID-19) from calendar week 12/2020. The individual question areas were adapted according to the survey period. Concerns about infection with the corona virus (self and family members or friends); concerns about infecting others; other concerns related to the corona virus (open question); ...

[Read more](#)

Study description available in:

DE

EN

[Access data](#)

Parent Ping: Daily Parent Survey, 2020-2021

Education Intelligence

Abstract copyright UK Data Service and data collection copyright owner. The Parent Ping research project began in July 2020 and ended in September 2021. The project aimed to document parental life, particular during the pandemic. Participants were recruited via opportunity sampling or via their children's school. Participants downloaded a mobile phone app which asked them roughly three closed survey questions each day. Question topics included: lockdown, home learning (due to school closures)...

[Read more](#)

Study description available in:

EN

[Access data](#)

Topic-specific Information Behaviour on the COVID-19 Pandemic (April 2021)

Presse- und Informationsamt der Bundesregierung, Berlin

Topic ?

Collection years ?

Country ?

× Belgium (1)

× Denmark (1)

Publisher ?

Results per page

30 ▾

Sort by

Date of publication (newest) ▾

<

1

>

Flash Eurobarometer 494 (Attitudes on **Vaccination** against Covid-19)

European Commission, Brussels; Directorate General Communication, COMMA.3 'Media monitoring and Eurobarometer'

Attitudes on **vaccination** against **Covid-19**. Topics: preferred time for getting **vaccinated**; importance of each of the following issues with regard to getting **vaccinated**: **vaccine** will help to end the pandemic, **vaccine** will protect respondent from getting **Covid-19**, **vaccine** will protect relatives and others from getting **Covid-19**, **vaccine** will make it possible to resume a more normal professional life, **vaccine** will make it possible to travel, **vaccine** will make it possible to meet family and friends, **vaccine** will make it possible to go to restaurants, cinemas etc.; importance of each of the following issues with regard to not getting **vaccinated**: pandemic will be over soon, personal risk of being infected is very low, risk posed by **Covid-19** in general is exaggerated, worries about side effects of **Covid-19 vaccines**, **vaccines** have not been sufficiently tested yet, **vaccines** are ineffective, against **vaccines** in general; factors to increase personal willingness of getting **vaccinated**: more people around doing it, more people have already been **vaccinated** and we see that there are no major side-effects, people that recommend the **vaccines** are **vaccinated** themselves, doctor recommends respondent to do so, **vaccines** are developed in the European Union, full clarity on how **vaccines**

Topic



Collection years



1900

2011 - 2021 Go



Country



- × Belgium (30)
- × Denmark (31)

Results per page

30

Sort by

Relevance



1

2



Second European Union Minorities and *Discrimination* Survey (EU-MIDIS II), 2016

European Union Agency for Fundamental Rights (FRA), Vienna, Austria

The second European Union Minorities and Discrimination Survey (EU-MIDIS II) was conducted in 2016 by the European Union Agency for Fundamental Rights (FRA) as a follow-up to the first survey on minorities' and immigrants' experiences of discrimination and criminal victimisation conducted by the Agency in 2008. The EU-MIDIS II survey collected information from 25,515 respondents from different ethnic minority and migrant backgrounds, including Roma, in all EU Member States (2016: EU-28 includ...

Read more

Study description available in: DE EN Access data

Survey on *Discrimination* and Hate Crime against Jews in the EU 2018

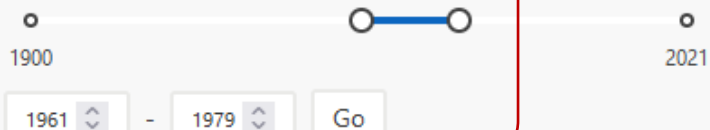
European Union Agency for Fundamental Rights (FRA), Vienna, Austria

The survey provides comparable data on the perceived extent and nature of antisemitism across a number of selected EU Member States, whether it is manifested as hate crime, hate speech, discrimination or in any other form that undermines Jewish people's feelings of safety and security. The survey was commissioned by the European Union Agency for Fundamental Rights (FRA). It follows up on the agency's first survey, conducted in seven countries in 2012. The overall objectives of FRA's second s...

Topic

- Leisure, tourism and sport (56)
- Leisure, tourism and sport - Society and culture (0)

Collection years



Country

Publisher

Results per page

30

Sort by

Relevance



1 2



Vacation Trips and Vacation Intents

EMNID, Bielefeld

Survey on vacation trips in the time interval from 1959 to 1963 and vacation intents for 1964. Topics: Vacation areas; number of trips; vacation length; vacation month; vacation at one or several places; utilization of travel agency; coupling vacation trips with stays at health resort and business trips; preferred means of transport for vacation; satisfaction with vacation; desire for 'recuperation' or 'adventure' on vacation. Demography: sex; occupation; employment; school education; voc...

Read more

Study description available in: DE EN Access data

Vacation Apartments

DIVO, Frankfurt

Vacation behavior as well as possession and intents to purchase property in domestic or foreign vacation areas. Topics: Vacation stays on Mediterranean islands; detailed description of these vacation trips; trip duration; number of persons traveling along; type of accommodations; activities at place of vacation; vacation expenditures. In the case of ownership of land or house in a vacation or recuperation area the following supplemental questions were posed: detailed description of possess...

Read more

Study description available in: DE EN Access data

Vacation Trips (January 1970)

Topic ?

Collection years ?

1900 2021

1961 - 1979 Go

Country ?

Publisher ?

Results per page 30

Sort by Relevance

1 2 3 4 ...

Vacation Trips and Vacation Intents

EMNID, Bielefeld

Survey on vacation trips in the time interval from 1959 to 1963 and vacation intents for 1964. Topics: Vacation areas; number of trips; vacation length; vacation month; vacation at one or several places; utilization of travel agency; coupling vacation trips with stays at health resort and business trips; preferred means of transport for vacation; satisfaction with vacation; desire for 'recuperation' or 'adventure' on vacation. Demography: sex; occupation; employment; school education; voc...

Read more Study description available in: DE EN Access data

Vacation Apartments

DIVO, Frankfurt

Vacation behavior as well as possession and intents to purchase property in domestic or foreign vacation areas. Topics: Vacation stays on Mediterranean islands; detailed description of these vacation trips; trip duration; number of persons traveling along; type of accomodations; activities at place of vacation; vacation expenditures. In the case of ownership of land or house in a vacation or recuperation area the following supplemental questions were posed: detailed description of possess...

Read more Study description available in: DE EN Access data

Vacation Trips (January 1970)

MARBLAN, Frankfurt

[Reset filters](#)[Clear search](#)[User Guide](#)[About](#)**Similar results**[Vacation Trips \(January 1970\)](#)[Vacation Trips \(August 1961\)](#)[Vacation Trips \(March 1971\)](#)[Vacation Trips \(September 1971\)](#)[Vacation Trips \(February 1972\)](#)[Access data](#)[View JSON](#)[Back](#)**Summary information****Study title**

Vacation Trips and Vacation Intents

Creator

EMNID, Bielefeld

Study number / PID

ZA0097, Version 1.0.0 (GESIS)

10.4232/1.0097 (DOI)

Abstract

Survey on vacation trips in the time interval from 1959 to 1963 and vacation intents for 1964. Topics: Vacation areas; number of trips; vacation length; vacation month; vacation at one or several places; utilization of travel agency; coupling vacation trips with stays at health resort and business trips; preferred means of transport for vacation; satisfaction with vacation; desire for 'recuperation' or 'adventure' on vacation. Demography: sex; occupation; employment; school education; vocational training; religious denomination; marital status; self-assessment of social class; household income; size of household.

Methodology**Data collection period**

Why use controlled vocabularies?

- ◊ Machine actionability
- ◊ Technical and semantic interoperability
- ◊ Precision in search - allows users to narrow down their searches in discovery tools
- ◊ Content consistency, comparability and understandability
- ◊ Facilitates metadata creation, organising and sharing
- ◊ Crosses language barriers

- ◊ Use those vocabularies that are generally used in the scientific domain



Fiction Books

BROWSE FICTION

Bestsellers

New Releases

Coming Soon

Fiction Home

FICTION SUBJECTS

General Fiction

Graphic Novels & Comics

Historical Fiction

Horror

Literature

Manga

Mystery & Crime

Poetry

Romance

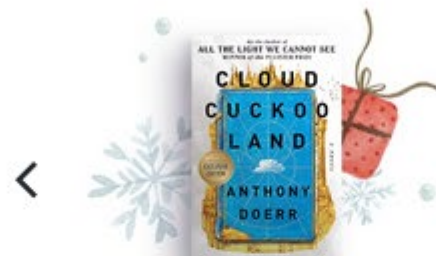
Science Fiction & Fantasy

Thrillers

Westerns

Browse All Subjects

Find the Perfect Gift



LITERARY FICTION FAVORITES



EDGE OF YOUR SEAT MYSTERIES & THRILLERS



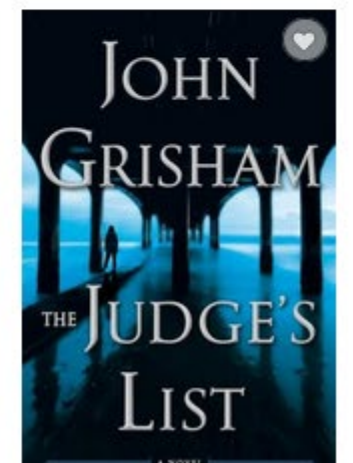
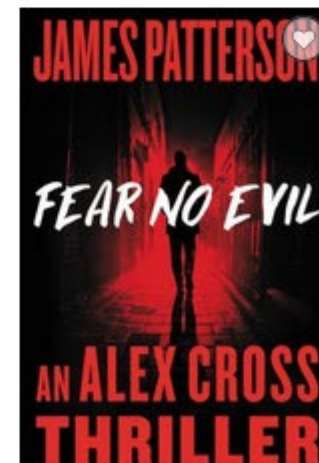
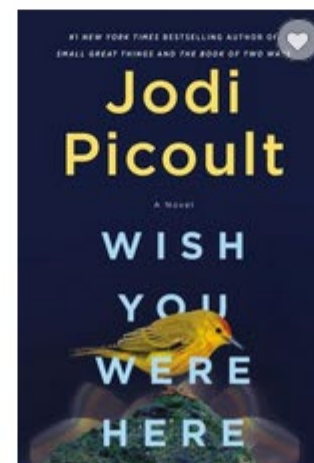
OUT OF THIS WORLD SCI-FI & FANTASY



MANGA

The Biggest Books of the Season

[See All](#)



Filters in CESSDA catalogue 1

- ◊ Study description language
 - ◊ Metadata should contain ISO language code in the agreed element(s)
 - ◊ Agreed element means in this context: in the element required by the cross-national catalogue harvester, that is, in the OAI-PMH record sent by the service provider.
 - ◊ In the internal record, the metadata may be somewhere else but the organisation constructs its OAI-PMH to send it in the required element.
- ◊ Country
 - ◊ Metadata contain the standardised ISO country code in the agreed element.
 - ◊ E.g. "FI" for Finland, regardless of metadata language
 - ◊ Why country names as such cannot be used?
 - ◊ Too many variations: The U.K., The UK, The United Kingdom, United Kingdom, GB, Great Britain, Britain, United Kingdom of Great Britain and Northern Ireland
 - ◊ Codes are language independent.
- ◊ Data collection year(s); Date of publication
 - ◊ Metadata contain the date(s) in the agreed ISO 8601 format in the agreed elements.

Filters in CESSDA catalogue 2

- ◊ Topic filter
 - ◊ Metadata contain a term from *CESSDA Topic Classification* in topic element
 - ◊ Useful to have the vocabulary name as well
- ◊ Without the vocabulary, people would use different terms for the same topic

Vocabularies in CESSDA

- ◊ Topic and subject classifications
 - ◊ CESSDA Topic Classification, 12 languages
 - ◊ European Language Social Science Thesaurus (ELSST), 14 language, more forthcoming
- ◊ DDI international documentation standard vocabularies to describe research methods. For example:
 - ◊ Mode of Collection (questionnaire, interview, experiment, observation, content coding etc.)
 - ◊ Time Method (longitudinal, cross-section, time series etc).
- ◊ Can be used for both quanti and quali data
- ◊ Controlled vocabularies used for broad categorisations of methods. More detailed explanations can be added in free-text elements

Access

Publisher

Finnish Social Science Data Archive (FSD)

Publication year

2021

Terms of data access

The dataset is (B) available for research, teaching and study.

Topics

Social sciences

Media

Occupational health

Crime and law enforcement



Keywords

Crimes against persons

Journalism

Journalists

News

Professional ethics

Social media

Working conditions





DDI vocabularies

- Aggregation Method
- Analysis Unit (x)
- Character Set
- Contributor Role
- Data Source Type (y)
- Data Type
- Date Type
- General Data Format (y)
- Language Proficiency
- Lifecycle Event Type
- Mode of Collection (x)
- Numeric Type
- Response Unit
- Sampling Procedure (x)
- Software Package
- Summary Statistic Type
- Time Method (x)
- Time Zone
- Type of Instrument (y)
- Etc.



28 Vocabularies

Results per page 30 ▾

Sort by CV name (A-Z) ▾

▸ Agency ?



[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 NO



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

Describes the entity being analyzed in the study or variable. This vocabulary can also be used to describe the unit of observation, which is the unit being observed, or from which data are collected. The unit of observation can be the same as, or different from the unit of analysis.

Version: 2.1 [Download](#)

EN DA FI FR DE IT LT NO PT SR SL SV

cessda

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

Names of organisations providing metadata for CESSDA Data Catalogue.

Version: 3.0 [Download](#)

EN



[CVs search](#) > ModeOfCollection v.4.0



DDI Alliance Controlled Vocabulary for Indsamlingsmetode

CV name	Mode Of Collection
CV short name	ModeOfCollection
CV definition	The procedure, technique, or mode of inquiry used to attain the data.
CV notes	This vocabulary was first published by the DDI Alliance. Please see: https://ddialliance.org/controlled-vocabularies/all .
CV name (da)	Indsamlingsmetode
CV definition (da)	Fremgangsmåden, teknikken eller undersøgelsesmetoden, der er anvendt til at opnå dataene.
Language	Danish (da)

Version	3.0.1	Date of publication	Danish (da) v.3.0.1								
EN	FI	DE	IT	LT	NO	PT	SL	SV	DA	FR	SR

[Details](#)
[Versions](#)
[Identity and general](#)
[Usage](#)
[License and Citation](#)
[Export/Download](#)

Code value

Code descriptive term (da)

Code definition(da)

▼ *Interview*

Interview

En forudplanlagt kommunikation mellem to (eller flere) personer – interviewere(n) og de(n) interviewede - hvor intervieweren(ne) får information fra de(n) interviewede. Hvis gruppeinteraktion er en del af metoden, anvend "Fokusgruppe".

▼ *Interview.FaceToFace*

Personligt interview

Dataindsamlingsmetode, hvor en tilstedeværende interviewer foretager et personligt interview, stiller spørgsmål og noterer svar. Brug dette bredere begreb, hvis det ikke er CAPI eller PAPI, eller hvis det ikke vides, om det er CAPI eller PAPI.

Interview.FaceToFace.CAPIorCAMI

Personligt interview: Computerassisteret (CAPI/CAMI)

Computer-assisted personal interviewing (CAPI) eller computer-assisted mobile interviewing (CAMI). Dataindsamlingsmetode, hvor intervieweren læser spørgsmål op til respondenterne fra skærmen på en computer, bærbar eller en mobil enhed såsom tablet eller smartphone og indtaster svarene på den samme enhed. Administrationen af interviewet styres af et specifikt designet program/applikation.

**DDI Alliance Controlled Vocabulary for Indsamlingsmetode**

CV name	Mode Of Collection
CV short name	ModeOfCollection
CV definition	The procedure, technique, or mode of inquiry used to attain the data.
CV notes	This vocabulary was first published by the DDI Alliance. Please see: https://ddialliance.org/controlled-vocabularies/all .
CV name (da)	Indsamlingsmetode
CV definition (da)	Fremgangsmåden, teknikken eller undersøgelsesmetoden, der er anvendt til at opnå dataene.
Language	Danish (da)

Version 3.0.1 Date of publication 2019-07-23

Language	Version	<input type="checkbox"/> SKOS	<input type="checkbox"/> PDF	<input type="checkbox"/> HTML
English (en)	4.0	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Finnish (fi)	4.0.2 ▾	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
German (de)	4.0.1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Italian (it)	4.0.1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Lithuanian (lt)	4.0.1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Norwegian (no)	4.0.2 ▾	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Portuguese (pt)	4.0.1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Slovenian (sl)	4.0.1	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Swedish (sv)	4.0.1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Alphabetical

Hierarchy

- ▶ DAMAGE
- ▶ DATA
- ▶ DEFENCE
- ▶ DEMOGRAPHY
- ▶ DEVELOPMENT
- ▶ DISABILITIES
- ▶ DISARMAMENT
- ▶ DISCRIMINATION
- ▶ DISEASES
- ▶ DOMESTIC RESPONSIBILITIES
- ▶ DRIVING LESSONS
- ▶ EARTH SCIENCES
- ▶ ECONOMICS
- ▶ EDUCATION
- ▶ EDUCATIONAL ENVIRONMENT
- ▶ EDUCATIONAL PLANNING AND ADMINISTRATION
- ▶ ELECTORS
- ▶ EMOTIONAL STATES
- ▶ ENERGY
- ▶ ENVIRONMENT
- ▶ BIODIVERSITY
- ▶ COUNTRYSIDE
- ▶ **ENVIRONMENTAL AWARENESS**
- ▶ ENVIRONMENTAL CHANGES
- ▶ ENVIRONMENTAL CONSERVATION
- ▶ ENVIRONMENTAL MANAGEMENT
- ▶ ENVIRONMENTAL QUALITY
- ▶ HABITATS

ENVIRONMENT > ENVIRONMENTAL AWARENESS

PREFERRED TERM

 ⓘ ENVIRONMENTAL AWARENESS 

DEFINITION

AWARENESS AND UNDERSTANDING OF THE ENVIRONMENT AND ENVIRONMENTAL ISSUES, AND OF THE IMPACT HUMAN ACTIONS HAVE ON THESE

BROADER CONCEPT

ENVIRONMENT

RELATED CONCEPTS

ENVIRONMENTALISTS

HISTORY NOTE

TERM CREATED APRIL 2021

IN OTHER LANGUAGES

ⓘ MILIEUBEWUSTZIJN	Dutch
ⓘ YMPÄRISTÖTIETOISUUS	Finnish
ⓘ SENSIBILISATION À L'ENVIRONNEMENT	French
ⓘ UMWELTBEWUSSTSEIN	German
ⓘ ΠΕΡΙΒΑΛΜΟΝΤΙΚΗ ΕΥΑΙΣΘΗΤΟΠΟΙΗΣΗ	Greek
ⓘ SAJONINGUMAS APLINKOS APSAUGOS KLAUSIMAS	Lithuanian
ⓘ MILJØBEVISSTHET	Norwegian
ⓘ SENSIBILIZARE FAȚĂ DE MEDIU	Romanian
ⓘ OKOLJSKA OZAVEŠČENOST	Slovenian
ⓘ MILJÖMEDVETENHET	Swedish

Links

- ◊ CESSDA Data Catalogue: <https://datacatalogue.CESSDA.eu/>
 - ◊ API: forthcoming in 2022
- ◊ CESSDA Vocabulary Service: <https://vocabularies.CESSDA.eu/>
 - ◊ API: <https://vocabularies.CESSDA.eu/api-docs>
- ◊ ELSST: <https://elsst.CESSDA.eu/>
 - ◊ API: <https://thesauri.CESSDA.eu/swagger-ui/index.html>
- ◊ DDI Alliance: <https://ddialliance.org/>

Thank you

taina.jaaskelainen@tuni.fi

 cessda.eu

 @CESSDA_Data



Licence: CC-BY 4.0

To Metadata and Beyond: Describing Data at FSD

Data Description and Metadata - What it takes to produce a good one?

8.12.2021

Emilia Hakkola

Event organised by



Finnish Social Science Data Archive

National resource centre since 1999.

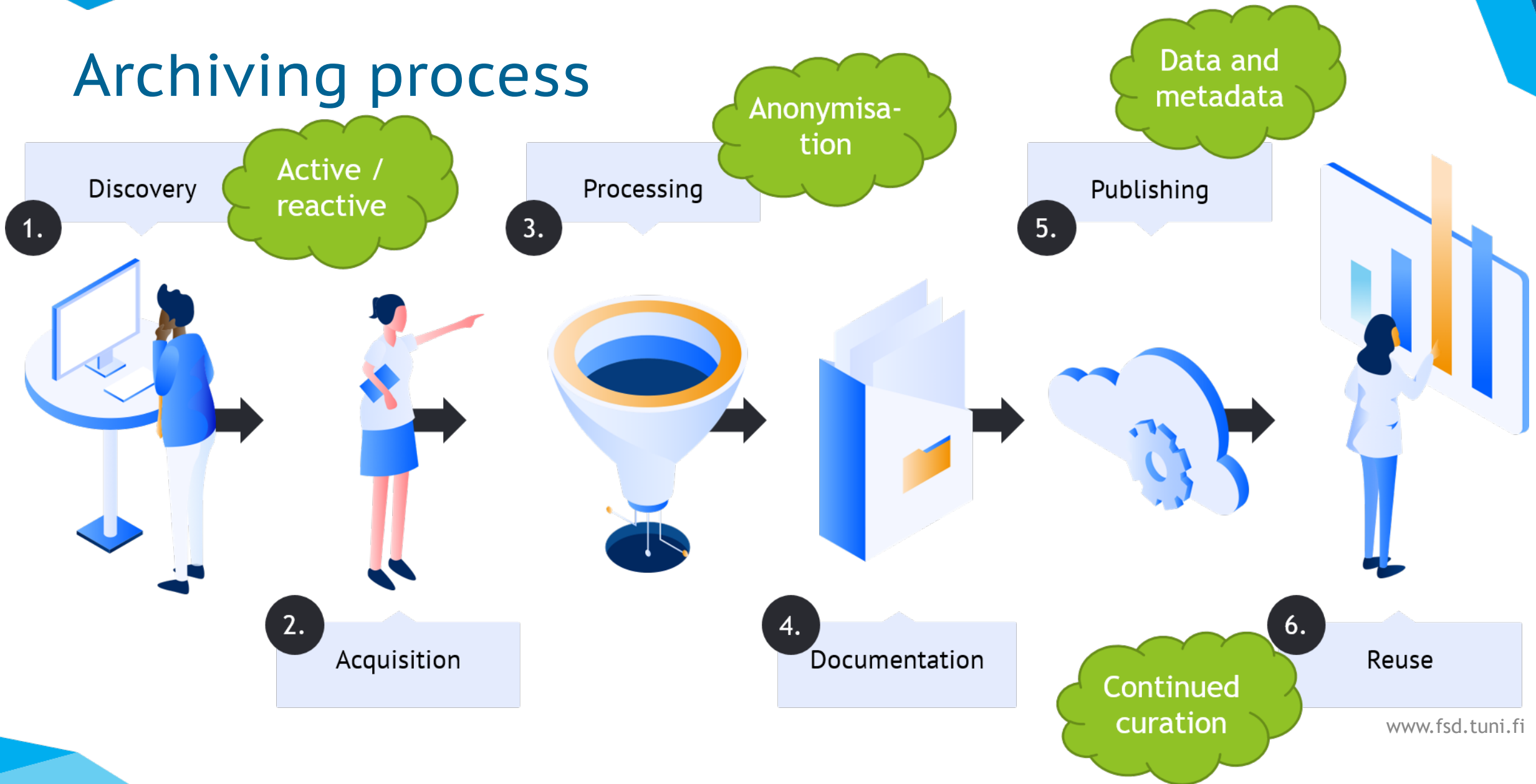
Archives, promotes and disseminates digital research data mainly for research, teaching and learning purposes.

- ▶ Main user groups: researchers, higher education students and teachers
- ▶ All services are free of charge
- ▶ Finland's national service provider for [CESSDA](#)
- ▶ CTS-certified Trustworthy Digital Repository
- ▶ Main functions
 - ▶ ingest, curate and preserve data collected to study (Finnish) society, people and cultural phenomena
 - ▶ information service
 - ▶ promote comparative research
 - ▶ impact (datasets, open science, best practices, standards...)





Archiving process



Systems

Vocabularies, e.g.

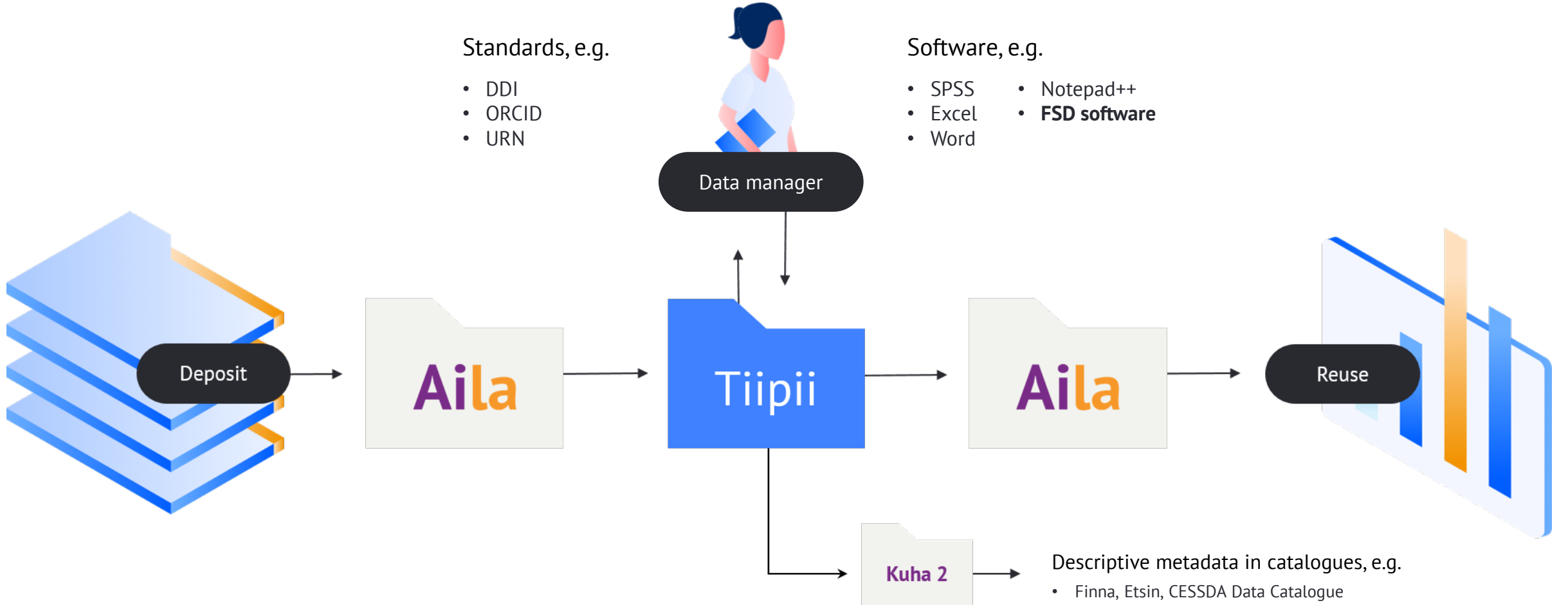
- YSO
- DDI CV
- CESSDA
- ELSST

Standards, e.g.

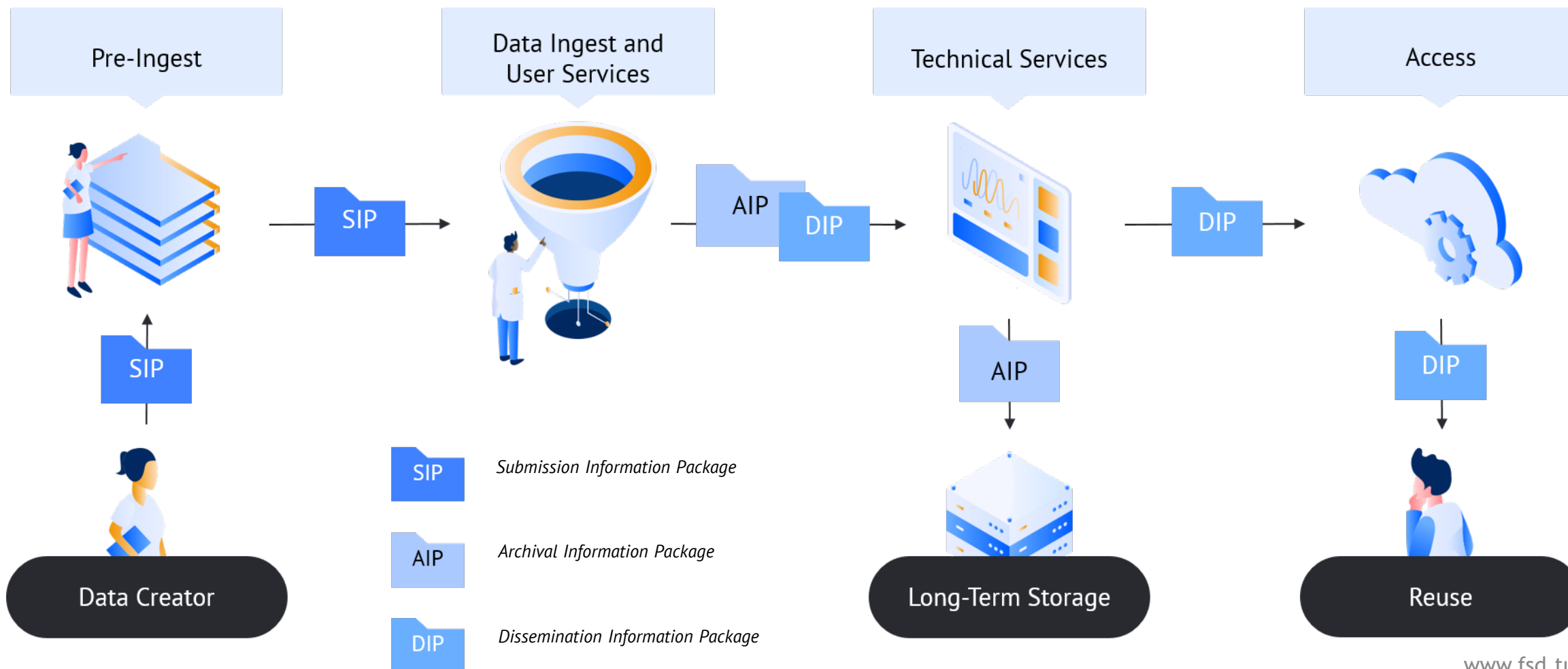
- DDI
- ORCID
- URN

Software, e.g.

- SPSS
- Excel
- Word
- Notepad++
- **FSD software**



OAIS (Open Archival Information System)



We have an extensive digital storage structure of our own. Server environment is maintained by Tampere University. On bit-level preservation we utilise the National Digital Preservation solution.

Documentation and metadata

- ▶ When creating metadata, it is important to focus on **describing the dataset itself** instead of the results based on the data.
- ▶ For each individual research dataset, it is **advisable to create a separate directory where both the data and the metadata are stored.**
- ▶ **Data description can be preserved**, for example, as a **text file** by including the basic information on the data:
 - ▶ description of how the study was conducted
 - ▶ information on data collection instrument
 - ▶ description of data files
 - ▶ description of variables
 - ▶ information on data availability
 - ▶ contextual information and paradata
- ▶ **Another alternative** is to select a **suitable metadata standard** and **store the metadata in a database in structured form.**

Describing data at FSD 1

- ▶ **Data archived at FSD are described in XML** in the international **DDI** format, which is specifically designed for social, behavioural, economic, and health sciences, both **on study and variable level**
- ▶ Study level descriptions in **Finnish and English**
- ▶ **Detailed description** is a prerequisite **for long-term preservation and reuse of data and findability.**
- ▶ DDI in XML enable **effective searches** and **generating the metadata in various formats.**
- ▶ Using an international standard also allows easily sharing the metadata to many **national and international catalogues.**



Describing data at FSD: DDI-Codebook

~300 elements are divided into 5 parts:

▶ **Document Description**

- ▶ includes e.g. bibliographical information of the metadata + license



▶ **Study Description**

- ▶ Description of the content of the dataset itself (not the research done on it!), e.g. dataset authors, keywords, abstract, sampling procedures, data collection, units of observation, target population, terms of access

▶ **Data Files Description**

- ▶ includes e.g. data structure and format, number of variables, size of files, software

▶ **Variable Description**

- ▶ includes e.g. variable and value labels and question texts

▶ **Other Study-Related Material**

The data archived by FSD are also transferred to a national digital preservation service. This package is complemented **with technical metadata and provenance data.**

Standardized vocabularies and classifications in use

- ▶ Analysis Unit
- ▶ Time Method
- ▶ Sampling Procedure
- ▶ Mode Of Collection
- ▶ Type of instrument
- ▶ Fields of Science Classification, MEC (OKM)
- ▶ Keywords, Finnish study descriptions, YSO
- ▶ Keywords, English study descriptions, ELSST
- ▶ CESSDA Topic Classification



cessda

<https://vocabularies.cessda.eu/>

<https://www.fsd.tuni.fi/en/services/data-management-guidelines/examples-and-vocabularies/>

Data description at FSD point by point

Examples mainly from the FSD3467 Finnish National Election Study 2019,
<http://urn.fi/urn:nbn:fi:fsd:T-FSD3467>

- ▶ We publish our study descriptions and related materials on Aila.
- ▶ You can also find these datasets e.g. via [Etsin](#), [Research.fi](#), [Finna](#) and [CESSDA Data Catalogue](#)
- ▶ Study description in machine readable DDI 2.0 format,
<https://services.fsd.tuni.fi/catalogue/FSD3467/DDI/FSD3467e.xml>

BASIC INFORMATION OF THE STUDY

Study title in Finnish and English

Possible alternative title of the study in Finnish and English

Dataset ID Number and Persistent identifier (URN)

FSD3467 Finnish National Election Study 2019

[Overview](#) [Detailed description](#) [Variables](#) [Publications](#) [Download data](#)

Study title

Finnish National Election Study 2019

Dataset ID Number

FSD3467

Persistent identifier

[urn:nbn:fi:fsd:T-FSD3467](https://nbn-resolving.org/urn:nbn:fi:fsd:T-FSD3467)

Data Type

Quantitative

The dataset is (B) available **for research, teaching and study.**

[Download the data](#)

Study description in other languages

- [in Finnish](#)

Related files

- [Codebook \(PDF file, in English\)](#)

AUTHOR INFORMATION

Author(s) Author(s)= person(s) or organization(s) responsible for the substantive and intellectual content of the dataset. Last name, First name (organization/affiliation at the time of data collection) or Organization

Other Identification/Acknowledgements Persons/bodies that have been involved in, for example, collecting or processing the material, but who are not the actual authors. Such e.g. those who have encoded, recorded or transcribed material.

Producers Organization with the financial or administrative responsibility for the physical processes whereby the dataset was brought into existence. For example, an organization has commissioned the study or been the initiator of the collection of the data but is not actually the author of the research.

“When several researchers participate in a research project, the responsibilities and rights of the researchers should be agreed on”
FSD, DMG

Authors

- Grönlund, Kimmo (Åbo Akademi University. Social Science Research Institute)
- Borg, Sami (Tampere University. Faculty of Management and Business)

Other Identification/Acknowledgements

- The following members of the election study consortium participated in questionnaire design (in addition to Sami Borg): Aino Tiihonen (Tampere University), Peter Söderlund (Åbo Akademi University) and Kim Strandberg (Åbo Akademi University).
- Current (1.1.2018-31.12.2021) members of the board of the election study consortium are: Kimmo Grönlund (chair, Åbo Akademi University), Åsa von Schoultz (University of Helsinki), Sami Borg (Tampere University), Hanna Wass (University of Helsinki), Elina Kestilä-Kekkonen (Tampere University) and Kim Strandberg (Åbo Akademi University).

Data Producers

- Election Study Consortium

Series

[Finnish National Election Studies](#)

Distributor

[Finnish Social Science Data Archive](#)

ABOUT ARCHIVING

Distributor, Depositor, Date of Deposit, Version & Series Information

Suggested citation

<bibCit>Grönlund, Kimmo (Åbo Akademi University) & Borg, Sami (Tampere University): Finnish National Election Study 2019 [dataset]. Version 1.0 (2020-09-30). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3467></bibCit>

Data Citation

The data and its creators shall be cited in all publications and presentations for which the data have been used. The bibliographic citation may be in the form suggested by the archive or in the form required by the publication.

Suggested citation:

Grönlund, Kimmo (Åbo Akademi University) & Borg, Sami (Tampere University): Finnish National Election Study 2019 [dataset]. Version 1.0 (2020-09-30). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3467>

More information on [citing data](#).

You cannot download the data yet.

If you are already a registered user, [log in](#) to be able to download. If not, you need to [register](#) yourself as a user first.

“From an archiving point of view, when it comes to research material, it is extremely significant as to who has the right to decide on the handover of research data for reuse and to decide the terms which apply to such reuse.” FSD, DMG

CONTENT, CONTEXT AND RESPONDENTS

Keywords [YSO - General Finnish ontology](#) & [ELSST - European Language Social Science Thesaurus](#)

Fields of Science/Topic Classification [Fields of Science Classification](#) (broadest level terms) & [CESSDA Topic Classification](#)

Abstract Compact summary describing the purpose, nature, and scope of the data collection, special characteristics of its contents, major subject areas covered, and what questions the PIs attempted to answer when they conducted the study. A listing of major variables in the study is important here. Use terminology different ways to make the data as findable as possible. Synonyms and, where possible, lower and upper terms should be used. Mention the scales & tests & measures used and list the background variables. Research project name and research funder.

Analysis/Observation Unit Type Basic unit of analysis or observation that the file describes: individuals, families/households, groups, institutions/organizations, administrative units.... [CVS: DDI Controlled Vocabulary for Analysis Unit](#)

Universe Group of persons or other elements that are the object of research and to which any analytic results refer. *People living in Finland and entitled to vote in the Finnish parliamentary elections in 2019 (excluding the Åland Islands)*

Data Type Rough division: Quantitative / Qualitative

Abstract

This survey focused on the 2019 parliamentary elections in Finland. Main themes included political participation, political attitudes, party allegiance, candidate and party choice, and voting behaviour. Further topics included citizens' initiative, different ways of having a say in matters, and future prospects of Finland. The data were collected after the elections through face-to-face interviews and a self-administered drop-off questionnaire. The interview data contain Finland's contribution to the international CSES study (module 5). Data collection was funded by the Ministry of Justice.

First questions in the interview covered the respondents' interest in politics, attention paid to media coverage of the elections (including social media), Internet use, discussions about politics with others, party identification and self-perceived social class.

Keywords

Internet; constituencies; democracy; election campaigns; elections; electoral candidates; mass media; members of parliament; parliamentary elections; political allegiance; political attitudes; political participation; trust; voting

Topic Classification

- Social sciences ([Fields of Science Classification](#))
- Political behaviour and attitudes (CESSDA Topic Classification)
- Elections (CESSDA Topic Classification)



TIME AND AREA

Time Period Covered Time period to which the data refer

Date of Collection Contains the date(s) when the data were collected.

Nation & Geographical Coverage Indicates the country or countries covered in the file and the geographic coverage of the data

Time Method E.g. Cross-section, longitudinal. [CVS: DDI Controlled Vocabulary for Time Method](#)

```
<timePrd event="single" date="2019-00-00"/>
  <collDate event="start" date="2019-04-17"/>
  <collDate event="end" date="2019-07-15"/>
  <collDate event="start" date="2019-09-25"/>
  <collDate event="end" date="2019-10-05"/>
  <nation abbr="FI">Finland</nation>
  <geogCover>Finland</geogCover>
<timeMeth>Longitudinal: Trend/Repeated cross-section
  <concept>Longitudinal.TrendRepeatedCrossSection</concept>
</timeMeth>
```

Time Period Covered

2019

Collection Dates

- 2019-04-17 – 2019-07-15
- 2019-09-25 – 2019-10-05

Nation

Finland

Geographical Coverage

Finland

Time Method

Longitudinal: Trend/Repeated cross-section

SAMPLE DESIGN AND COLLECTION OF DATA

Data collector(s) Refers to the entity collecting the data. Organization or Last Name, First Name (Organization)

Sampling Procedure Type of sample and sample design used to select the respondents to represent the population. E.g Total universe/Complete enumeration, Probability: Simple random, Non-probability: Purposive. [CVS: DDI Alliance Controlled Vocabulary for Sampling Procedure](#) and a brief description of sampling / selection

Collection Mode [CVS: DDI Alliance Controlled Vocabulary for Mode Of Collection](#)

Research Instrument [CVS: DDI Alliance Controlled Vocabulary for Type of Instrument](#)



Sampling Procedure

Non-probability: Quota

The sample was drawn with the help of quota sampling, in which the quotas were based on NUTS3 region of residence, type of municipality of residence, mother tongue, gender and age of the respondents. The quota sampling was based on statistical data on the distribution of the target population according to the mentioned factors. In the first stage, the number of persons required by the study was regionally divided by NUTS3 major regions. In the second stage, municipality types within each NUTS3 region were taken into account by using the Eurostat DEGURBA classification. In the Uusimaa region, Helsinki was separated as its own area. Interviewees were selected based on the sampling. The interviews were conducted by using the starting point method, where the first interview was conducted in a randomly selected starting point. Additionally, some interviews in city areas were conducted in one specific location (e.g. larger hall, convention centre) instead of the door-to-door method. For these interviews, the interviewer used the respondents' postal codes to ensure that all selected areas were sufficiently represented. Interviews of the Swedish-speaking respondents were conducted in regions where the proportion of Swedish-speaking residents was significant: Helsinki, Uusimaa, Finland Proper, and the Swedish-speaking regions of Ostrobothnia and Central Ostrobothnia. After the interview, the respondents were asked to complete a self-administered paper questionnaire with additional questions (drop-off questionnaire). 753 respondents completed the drop-off questionnaire. Approximately two thirds of the interviews were conducted in April and May. Some phone interviews were conducted between September 25 and October 5 of 2019 to reach the respondents who had mistakenly not been asked all questions in the interview.

Collection Mode

Face-to-face interview

Self-administered questionnaire: Web-based (CAWI)

Self-administered questionnaire: Paper

Telephone interview

Research Instrument

Structured questionnaire

USE OF DATA

* Dataset availability:

- (A) openly available for all users without registration (CC BY 4.0),
- (B) available for research, teaching and study,
- (C) available for research only (including Master's, doctoral and Polytechnic/University of Applied Sciences Master's theses),
- (D) available only by permission from the data depositor/creator.

Restrictions Access conditions set by deposition agreement between the depositor & FSD. *

Special Terms Possible additional conditions set by data depositor.

Citation Requirement

Deposit Requirement

Disclaimers

<useStmt>

<restrctn>The dataset is (B) available for research, teaching and study.</restrctn>

<citReq>The data and its creators shall be cited in all publications and presentations for which the data have been used. The bibliographic citation may be in the form suggested by the archive or in the form required by the publication.</citReq>

<deposReq>The user shall notify the archive of all publications where she or he has used the data.</deposReq>

<disclaimer>The original data creators and the archive bear no responsibility for any results or interpretations arising from the reuse of the data.</disclaimer>

</useStmt>

OTHER INFORMATION

Data Sources

`<weight>`The data contain a weight variable [paino] which weights the sample to match the mother tongue, age, gender and electoral district distributions in the population as well as the actual vote share of parties in the election.`</weight>`

Weighting Information on the weight variables included in the data: names, how they are made and how they are used

Response Rate

Completeness of Data and Restrictions Possible deficiencies, errors and deletions. Description of modifications made to prevent identification of participants. Other important information.

Related Materials

Related Publications Bibliographic and access information about articles and reports based on the data

Completeness of Data and Restrictions

A mistake occurred in programming of the questionnaire, which resulted in the original data including 288 respondents who had not been asked questions Q12LHA Q12LH-a - K18_SO K18 during the interview. These respondents were contacted later and asked to respond to the missing questions. Of the 288 respondents, 173 agreed to respond to the questions. As a result, the data include 115 respondents who, depending on their responses regarding voting behaviour, did not respond to 3-6 questions. These questions include the following: Nowadays many people do not vote in elections for one reason or the other. Did you vote or not in these parliamentary elections? - (If did not vote): How self-evident was it for you not to vote? - (If did not vote): If you had voted, the candidate of which party or group would you have voted for? - (If voted): The candidate of which party or group did you vote for in these parliamentary elections? - (If voted): How easy or difficult was it for you to choose the party or group whose candidate you voted for? - (If voted): How easy or difficult was it for you to find a suitable candidate? - (If voted): Did you consider voting for a candidate of any other party or group? - (If considered): Which party or group?

To prevent identification of participants, variables D23posti denoting the respondent's postal code and D14 denoting the occupation of the respondent's spouse/partner were removed from the data at FSD. Additionally, open-ended responses in the following variables were removed: D37_so denoting mother tongue, D34_so denoting trade union/professional association and D10_so denoting membership of a church or religious community. Responses in variable D17 denoting municipality of residence were categorised into the five largest cities in Finland, and a variable was created to denote the respondent's NUTS3 region of residence. Individual men's names were removed from two open-ended responses. Variable D07 denoting the respondent's occupation was categorised by using ISCO-08.

DATA FILES DESCRIPTION (quantitative data)

Study description part contains data file information for qualitative data.

```
<fileDscr>  
  <fileTxt>  
    <fileName ID="FSD3467e_file_1">daF3467e.por</fileName>  
    <dimsns>  
      <caseQty>1598</caseQty>  
      <varQty>422</varQty>  
    </dimsns>  
    <fileType>SPSS Portable</fileType>  
  </fileTxt>  
</fileDscr>
```

Number of Cases and Variables

422 variables and 1598 cases.

VARIABLES DESCRIPTION

Variable Groups:

Joint text of a group of variables (usually the text of the question battery)

Variables:

Variable name and label

Question text (preQTxt, qstnLit and postQTxt)

Interviewer Instructions

Summary Statistics

Value labels

Variables are not keyworded.

A tool for semi-automation of keywording is under development in FSD. The tool suggests appropriate YSO and ELSST keywords (algorithms).

```
<varGrp var="K32_1 K32_2 K32_3 K32_4  
K32_5">
```

```
  <txt>To what extent do you trust or  
  mistrust the following? </txt>  
</varGrp>
```

```
<labl level="variable">[q1] Your gender</labl>
```

```
<qstnLit>Your gender</qstnLit>
```

```
<catgry>
```

```
  <catValu>1</catValu>
```

```
  <labl level="category">Male</labl>
```

```
<catStat>500</catStat>
```

```
  <catValu>2</catValu>
```

```
  <labl level="category">Female</labl>
```

```
<catStat>500</catStat></catgry>
```




FSD3467 Finnish National Election Study 2019

[Overview](#)

[Detailed description](#)

[Variables](#)

[Publications](#)

[Download data](#)

Select variable

[d2] Your gender

Question text

Your gender

Frequencies

Category labels	Value	n
Male	1	707
Female	2	889
Other (spontaneous)	3	2
Don't want to say (spontaneous)	7	0

Displayed frequencies are not weighted

Describing data at FSD 2

- ▶ When depositing data package (data files, questionnaires etc.) through Aila, **depositor fills out basic information** regarding data content, data collection, and publications based on the data.
- ▶ This information is needed to be able to produce the necessary metadata. **FSD edits and adds information – and value.**
- ▶ Variable description is based on SPSS data file processed in FSD. **FSD edits and adds information – and value.**

Depositing Form

Perustiedot

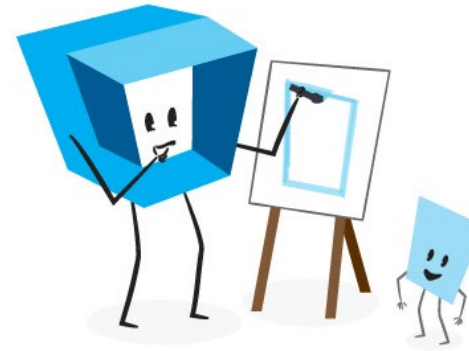
Nimi suomeksi *	FSD Survey 2021
Nimi englanniksi	FSD Survey 2021
Tekijät *	Emilia Hakkola Finnish Social Science Data Archive
Kuvaa, mitä tutkittaville on kerrottu aineiston käsittelystä ja käytöstä. Kerro, sisältääkö aineisto henkilötietoja ja miten aineistoa on anonymisoitu. *	<p>The researcher stores and treats the data as confidential in accordance with the obligation of confidentiality and the Personal Data Act. The name of the subject is not used when processing the interview material or publishing the research results. Subjects are not identifiable in research publications. In addition, it has been reported that the research material will be stored in the Social Science Data Archive for possible further use, with the deletion of data enabling the identification of the respondents.</p> <p>The material contains personal information such as the person's age, gender, country of birth, current home country. The material is mainly in numerical form and it is not possible to identify individuals.</p>
Kuvaa tutkimushankkeen rahoitus ja anna mahdollinen hankenumero.	Academy of Finland, XXXXX
Ilmoita julkaisut ja opinnäytteet, joissa aineistoa on jo hyödynnetty. Kirjaa esimerkiksi tekijä, julkaisuvuosi, julkaisun nimi, pysyvä tunniste, painopaikka, kustantaja.	Hakkola, Emilia (2021) These are our main results. Tampere: Finnish Social Science Data Archive.
Voit antaa lisätietoa aineistosta. Huomaathan, että keruutietoja kysytään seuraavalla välilehdellä.	

Keruutiedot

Kerääjät	Finnish Social Science Data Archive
Keruun aloitusaika	16.09.2020
Keruun lopetusaika	31.10.2020
Keruumenetelmät	Itsetäytettävä lomake: verkkolomake
Lisätietoja keruumenetelmästä	Self-administered questionnaire: Web-based (CAWI)
Keruväline tai -ohje	Strukturoitu lomake
Lisätietoja keruvälineestä	Structured questionnaire
Otantamenetelmät	Todennäköisyysotanta
Lisätietoja otantamenetelmästä	Probability Here is some more information about the sampling procedure
Voit antaa lisätietoja keruusta.	

Tiedostot *

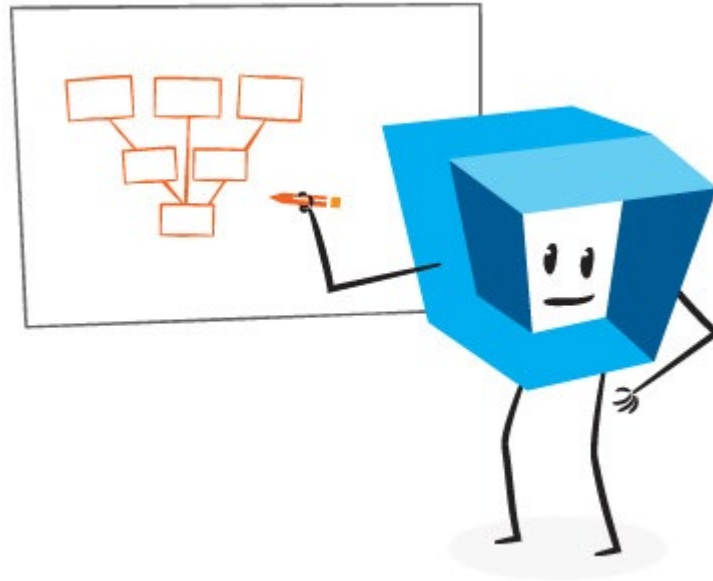
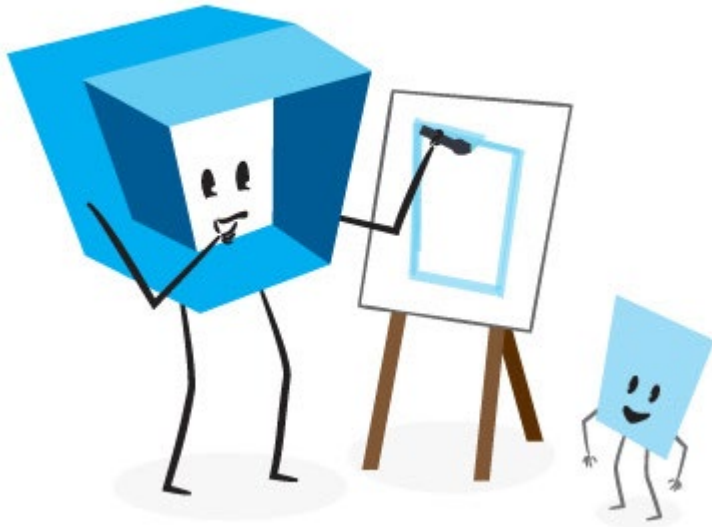
this is our data.xlsx
Tiedoston sisältö
Kieli
Kuvaus
questionnaire_fin.pdf
Tiedoston sisältö
Kieli
Kuvaus



Want to hear more?

- ▶ FSD's Archiving Services
 - ▶ <https://www.fsd.tuni.fi/en/services/depositing-data/>
- ▶ Downloading and Using Data
 - ▶ <https://www.fsd.tuni.fi/en/data/downloading-and-using-data/>
- ▶ FSD's Data Management Guidelines: Data description and metadata
 - ▶ <https://www.fsd.tuni.fi/en/services/data-management-guidelines/data-description-and-metadata/>
- ▶ CESSDA Data Management Expert Guide: Documentation and metadata
 - ▶ <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>
- ▶ FSD's Steps Towards Being More FAIR
 - ▶ <https://www.fsd.tuni.fi/en/news/articles/steps-towards-being-more-fair/>

Questions?



I love
Aila

User services
user-services.fsd@tuni.fi
+358 29 452 0411



Metadata in reuse: harvesting, licensing, repurposing and FAIR

*Data Description and Metadata - What it takes to produce a good one?
December 8, 2021*

Tuomas J. Alaterä, FSD
tuomas.alatera@tuni.fi
<https://orcid.org/0000-0002-3448-3448>

 cessda.eu

 @CESSDA_Data



Licence: CC-BY 4.0

Human readable metadata is fine...

- ...but having machine-readable metadata too is better (even though one cannot even see it!)
 - Allows harvesting
 - Allows automagical enriching
 - Allows wider discoverability
 - Allows citing
 - Allows building on top of interoperable metadata
- Linked open data and APIs are crucial

Persistent Identifiers

- Identifiers for everything
 - Not only for resolving to the resource
 - PIDgraphs rely on persistent identifiers to build enriched and meaningful relations
 - Researcher IDs (predominantly ORCID)
 - Research Organisation / Funder IDs (ROR, ISNI, URN ...)
 - Publications, articles and such (ISBN, ISSN, DOI, Handle...)
 - Research project IDs (RAiD...)
 - In addition, identifiers for any entities that has relations with the dataset
- Needed that the platform supports the use of IDs
- Doable for example in JSON-LD for basically all actors
- Should be included as text, if types or fields for relations not available

Persistent Identifiers, JSON-LD examples

```
"publisher": [
  {
    "@type": "Organization",
    "sameAs": "https://ror.org/040af2s02",
    "name": "University of Helsinki"
  }
],

"name": [
  {
    "@value": "Finnish Voter Barometer 1973",
    "@language": "en"
  },
  {
    "@value": "Puolueiden ajankohtaistutkimus 1973",
    "@language": "fi"
  }
],

"citation": {
  "@type": "CreativeWork",
  "creator": [{
    "@id": "https://orcid.org/0000-0000-0000-0000",
    "@type": "Person",
    "name": "John Smith",
    "familyName": "Smith",
    "givenName": "John",
    "identifier": "https://orcid.org/0000-0000-0000-0000",
    "email": "j.smith@somedomain.org"
  }],
}
```

Metadata served in different formats

- Offer the metadata for harvesting primarily via an API
 - Multiple formats can be produced from the core discipline specific metadata
 - Basic mapping e.g. to Dublin Core increases usability
 - FSD uses Kuha2 for serving DDI Codebook, EAD3 and OAI Dublin Core
 - DataCite is a recommended format to consider
- Metadata in XML or different LOD format can be embedded or linked to the landing page

Licenses for metadata

- Licenses or conditions for use for data are common
- Metadata should be licensed as well
 - Because of clarity of (re)use (and perhaps merit)
 - At times required by aggregators
 - Recommended formats CC0 and CC BY
 - Metadata license declared in machine-actionable format
 - Like data, metadata should be persistent, and versions monitored

Citation

- Merit depends on citations
- Therefore, data citations must stand the test of time
 - Use of PIDs
 - Repository driven service: offer both a citation example and a citation in machine-actionable format
 - Importance of a landing page for data where relevant information is available, like ID, title, creator, publisher, release date, version.
 - Make sure these are available in machine-readable format using e.g. schema.org or DC.

Machine-actionable metadata and FAIR

- FAIR depends on machine-actionable metadata and the use of various (FAIR) controlled vocabularies
- In SSH domain, tools offered by CESSDA (vocabularies.cessda.eu) are of use (DDI, CESSDA vocabs)
- Other vocabularies e.g. for place names, coverage, specimen etc. as needed by the community
- Expressed in a format suited to your needs (using schema.org, DC Terms, Open Graph...)

In Conclusion

- Interoperability is both a technical and content issue
- PIDs and other published identifiers need to be collected early
- Further PIDs need to be minted as needed
- Different forms of metadata may be needed for harvesting
- For machine-actionability a standard for open linked data is needed
- Machine actionable interoperability relies on various sources and should not depend on a manual processing only

Collaborative Notes

Data Description and Metadata - What it takes to produce a good one?

Date: Wednesday 8 December 2021 10:00 - 12:30 CET

<https://www.fsd.tuni.fi/en/news/events/webinar2021-12-08/>

Programme

- Introduction
 -
- CESSDA and its services related to data description and metadata production and use
 - If your organisation dont use ISO 8601, make the move right now :)
 - <https://xkcd.com/1179/>
- Significance of data description quality for cross-national data catalogues
 -
- Describing data in practice: FSD as a case example
 -
- Metadata in (re)use: harvesting, licensing, repurposing and FAIR
 -
- Discussion and Q & A

Questions covered during the discussion

Q: Are there any good templates to use to assist in describing metadata?

A: Not that we know of. Various submission forms provide a structure, but they are not usually open. Best suggestion is to look at any catalogue entry, for example in Aila, where one can see all the elements used and the content entered. E.g. [FSD2316 Electoral Alliances in Parliamentary Elections 1945-2007](#)

Q: What is the authoritative source for DDI vocabularies?

A: When the pipeline is finalised it will be DDI Alliance Vocabulary Service. At this point DDI website has older versions, latest versions are on cessa's vocab service.

Q: Is the standard for service the same as for the dataset ?

A: An API service is capable of serving harvesters with many types of metadata. It does not care if the metadata is about data or books etc. It may be useful that you serve metadata in various formats via your API service because different aggregators may use different metadata models. Therefore some mapping is usually needed. With OAI-PMH providing DublinCore is required.

[Protocol for Metadata Harvesting - v.2.0](#) is used in many archives regardless of content. It can express records in multiple metadata schemas.

Q: How long does it take to produce a metadata description at FSD?

A: Depends on the dataset. Metadata production takes a day or two. This is not including any other things that need to be done, like anonymisation, data cleaning.

Q: Who assigns the Topic Classification - authors themselves or database administrators?

A: Data producers at FSD. Depositors approve the final metadata description and therefore the topics too.

Q: How do you define the owner/author of a dataset?

A: Advised to decide and agree on in advance. Definition for the owner: "The person, corporate body, or agency responsible for the work's substantive and intellectual content." FSD requires contact details to someone who can act on the dataset, and decide that it can be archived (and in the case of those dataset that are disseminated only on a request, contact details of a person granting access.

About authorship: <https://tenk.fi/en/advice-and-materials>

Q: Who curates the data sets on the variable level? FSD? The researchers? FSD together with the researchers?

A: Together, but FSD takes the main responsibility. FSD also has the responsibility to produce metadata needed for long term preservation and relevant technical metadata. Remember to describe the data, not the study or research results.

Q: Are there best practices regarding data ingestion and curation within the CESSDA consortium? Regarding workflows specially.

A: See CESSDA's DMEG ([Data management expert guide](#)) for valuable information.

Most archives have similar ingest and curation principles, but there is variation in depth.

There has been training and learning from others. This could be elaborated. Repositories need to be open about their policies, practices and use of DMPs. Different resources in different archives play a central role in how detailed or extensive the processes can be.

Certified trusted digital repositories need to offer, preferably public, documentation about these processes. See also eg.

<https://www.fsd.tuni.fi/en/data-archive/documents/records-management-and-archives-format-on-plan/>

Q: Preservation metadata, how is it done? How do you update these datasets?

A: FSD uses a national digital preservation solution and relies on digital preservation specifications set by the service. Descriptive DDI metadata is included in the METS document descriptive metadata section as it is. In addition we produce technical metadata, like size of the tile, checksums, image or video file information (MIX, MODS). We use a very

minimal provenience section (PREMIS) and rights metadata (because data are only disseminated via FSD and extensive rights metadata doesn't have to travel with the archival package. Process is semi automated at the moment. Updates are incremental and it would be possible to pull previous versions from digital storage too.

For specifications and details: <https://digitalpreservation.fi/en>

Q: CESSDA Data Catalogue metadata model?

A: [CESSDA Data Catalogue - DDI Codebook Profile](#) This is the CESSDA Data Catalogue metadata model in a table format for DDI 2.5 version but please note that this is only at study level (excluding variable information). Archives often have and need more information, so the catalogue metadata is only a sub-set. It lacks some very relevant elements.

<https://cmv.CESSDA.eu/profiles/cdc/ddi-2.5/1.0.4/profile.xml>

The full CESSDA metadata model: [CMM CESSDA Metadata Model](#)