TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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

cessda

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

Tampereen yliopisto
Tampere University

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...)

Aila

CORE TRUST SEAL

Findable
Accessible
Interoperable
Reusable

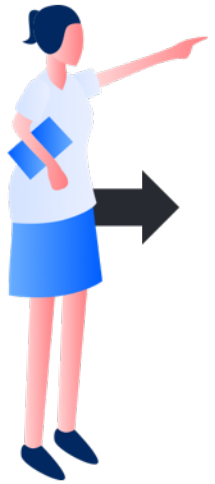
TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

Archiving process

1. Discovery

Active / reactive

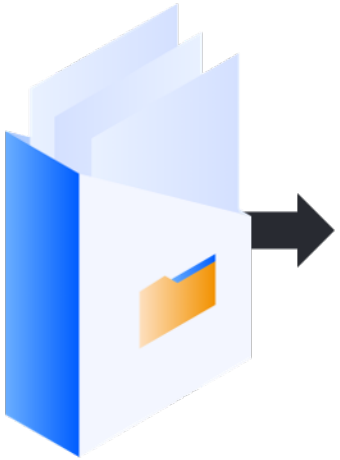
2. Acquisition



3. Processing

Anonymisa-tion

4. Documentation



5. Publishing

Data and metadata

Continued curation

6. Reuse

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

Systems

Vocabularies, e.g.

- YSO
- DDI CV
- CESSDA
- ELSST

Standards, e.g.

- DDI
- ORCID
- URN

Software, e.g.

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

Data manager

Deposit

Aila

Tiipii

Aila

Reuse

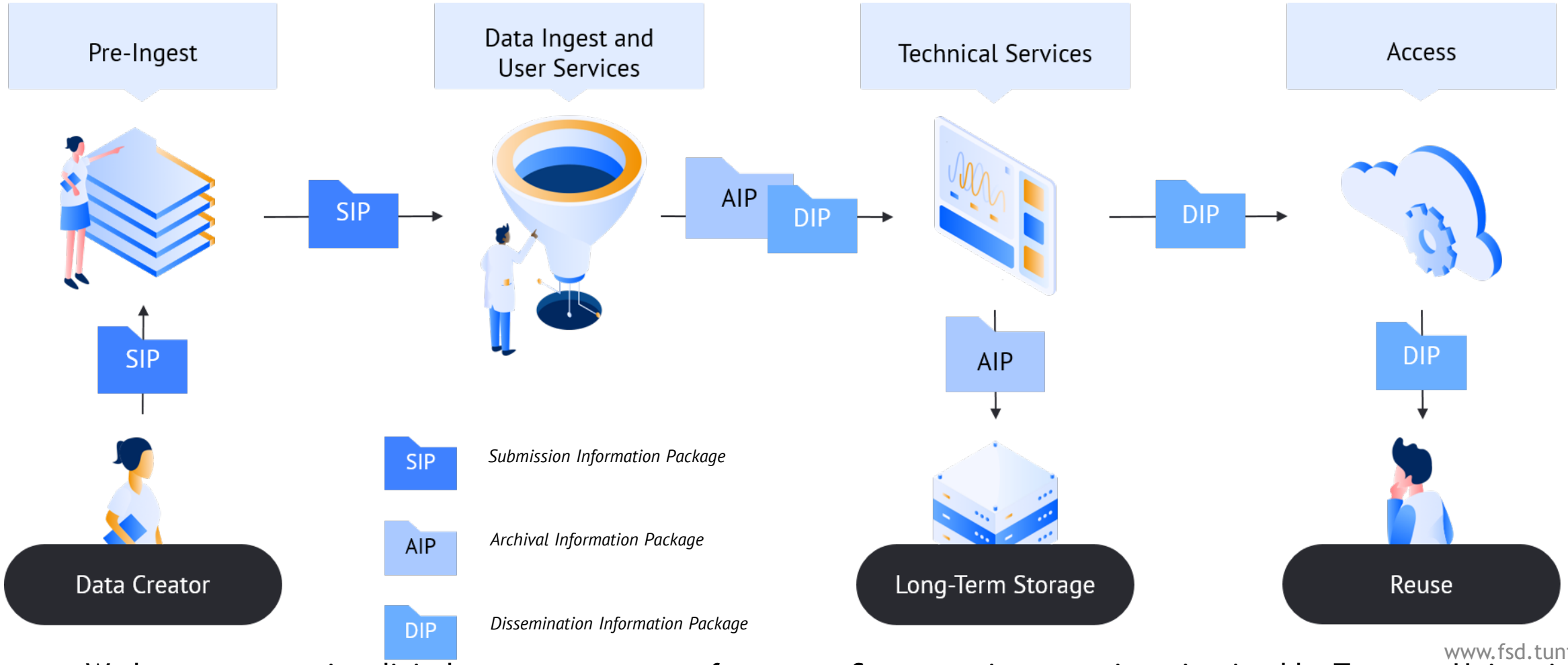
Kuha 2

Descriptive metadata in catalogues, e.g.

- Finna, Etsin, CESSDA Data Catalogue

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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.

TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

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.**

TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

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**.

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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.**

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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**

https://vocabularies.cessda.eu/

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

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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

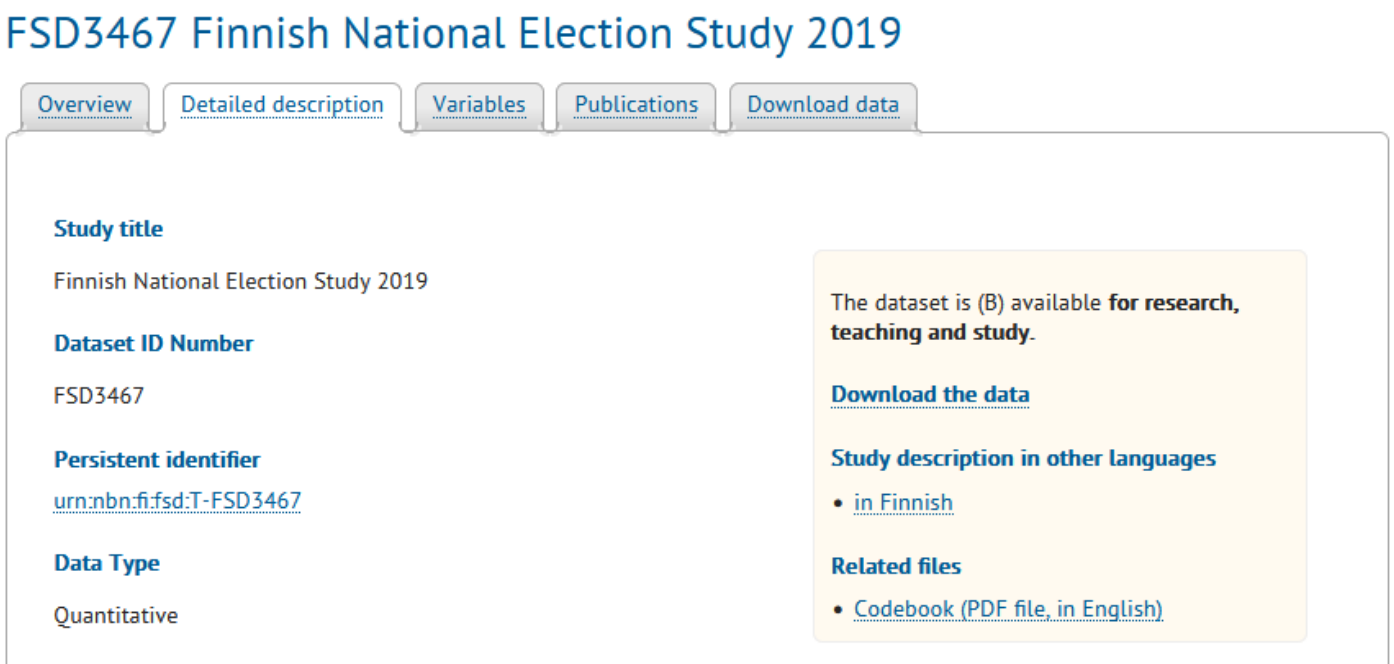
TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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)



TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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

https://www.fsd.tuni.fi/en/services/data-management-guidelines/copyright-and-agreements/

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

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

ABOUT ARCHIVING

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

Suggested citation

<biblCit>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</biblCit>

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

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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)

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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

TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

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

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

* 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.

USE OF DATA

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>

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

<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>

OTHER INFORMATION
Data Sources

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>
   <dimensns>
    <caseQnty>1598</caseQnty>
    <varQnty>422</varQnty>
   </dimensns>
   <fileType>SPSS Portable</fileType>
  </fileTxt>
 </fileDscr>
```

Number of Cases and Variables

422 variables and 1598 cases.

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

VARIABLES DESCRIPTION

Variable Groups:

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

```
<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>
```

Variables:

- Variable name and label
- Question text (preQTxt, qstnLit and postQTxt)
- Interviewer Instructions
- Summary Statistics
- Value labels

```
<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>
```

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).

TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

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

TIETOARKISTO
FINNISH SOCIAL SCIENCE
DATA ARCHIVE

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.**

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

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. *	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. 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.
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

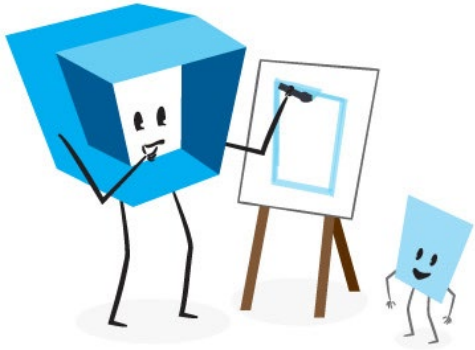
Kerääjät	Finnish Social Science Data Archive
Keruun aloitusaika	16.09.2020
Keruun lopetusaika	31.10.2020
Keruumenetelmät	Itsetäytettävä lomake: verkkolomake
Lisätietoja keruumenetelmästä	Self-administered questionnaire: Web-based (CAWI)
Keruuväline tai -ohje	Strukturoitu lomake
Lisätietoja keruuvälineestä	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	

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

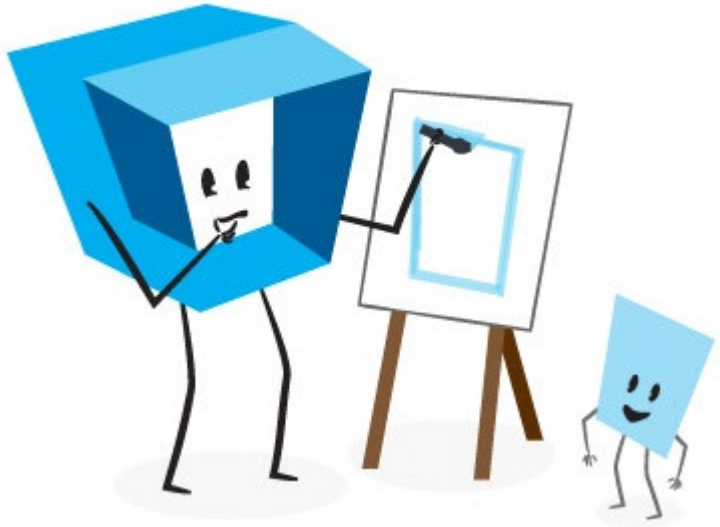
Want to hear more?

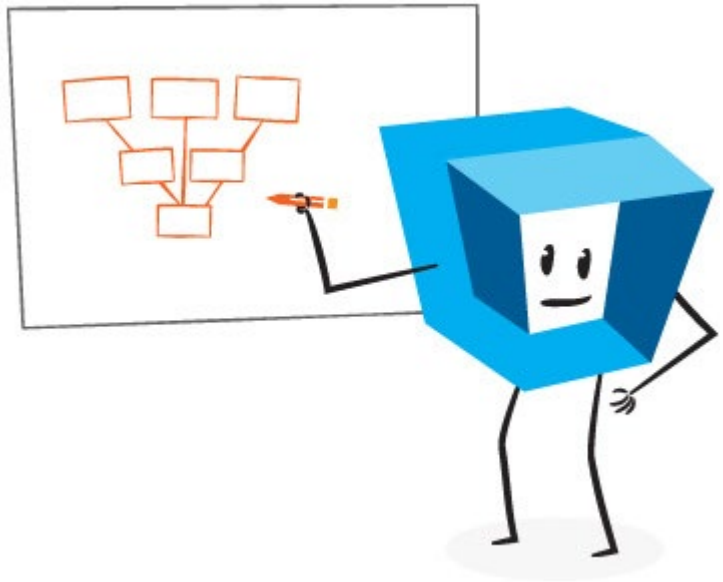


- FSD's Archiving Services
  - https://www.fsd.tuni.fi/en/services/depositing-data/
- Downloading and Using Data
  - https://www.fsd.tuni.fi/en/data/downloading-and-using-data/
- FSD's Data Management Guidelines: Data description and metadata
  - https://www.fsd.tuni.fi/en/services/data-management-guidelines/data-description-and-metadata/
- 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/

TIETOARKISTO
FINNISH SOCIAL SCIENCE DATA ARCHIVE

Questions?





I love Aila

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