

CESSDA Data Catalogue

Preparing Metadata for the CDC

*Mari Kleemola (FSD), CESSDA Tools & Services WG leader
with input from John Shepherdson and Taina Jääskeläinen*

*CESSDA Training Day
27-28 November 2019, Cologne*

 cessda.eu

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



In this presentation...

- ◇ CDC in general, harvesting metadata
- ◇ “Must have” and “should have” metadata
- ◇ Forthcoming
- ◇ Further information

21836 results found

[Advanced search](#)[Reset filters](#)[Clear search](#)[Topic](#)[Collection years](#)[Country](#)[Publisher](#)[Language of data files](#)

Results per page

30

Sort by

Relevance



1

2

3

4

...



[VALU 2003 - SVT exit poll survey EMU-election 2003](#)

*Holmberg, Sören (University of Gothenburg, Department of Political Science);
Thedéen, Torbjörn (Royal Institute of Technology, Division of Safety Research);
Hernborn, Hans (Sveriges Television); (1 more)*

September 14, 2003 was the date for the referendum on Sweden's possible participation in the third stage of the Economic and Monetary Union (EMU). The question that all those entitled to vote had to answer was: 'Do you think Sweden should introduce the euro as its currency?' In connection with the referendum an exit poll survey was carried out. In VALU - Swedish exit poll surveys - voters leaving polling stations are asked by public service broadcaster Sveriges Television (SVT) to fill in a...

[The Regional Western Sweden SOM Survey 2017](#)

University of Gothenburg, SOM Institute

Western SOM has been conducted yearly since 1992, but was initially limited to residents of Gothenburg and its surrounding municipalities. In 1998, the survey was extended to include the entire Västra Götaland County plus the municipality of Kungsbacka. The purpose of the regional surveys is to enable SOM researchers to study attitudes and behaviour linked to local and regional issues. They emphasise public services and media, although many of the questions are identical with those used in N

CESSDA Data Catalogue (CDC)

- ◊ Version 2.1 in production on CESSDA Technical Infrastructure: <https://datacatalogue.cessda.eu/>
- ◊ Now listed in EOSC Marketplace: <https://marketplace.eosc-portal.eu/services/cessda-data-catalogue>
- ◊ User guide: <https://datacatalogue.cessda.eu/documentation/>
- ◊ Compatible with Google Data Search, special markup embedded in pages
- ◊ Issue tracker for logging and tracking progress against any metadata issues (bitbucket).

Harvesting metadata from SPs

- ◆ Fully automated harvesting of metadata from Service Providers ensures that CDC content is up to date.
- ◆ Automatically updates its content every 24 hours.
- ◆ Any SP that has publicly accessible harvestable OAI-PMH endpoint that serves CMM compliant metadata in DDI 1.2/2.5 XML format can be part of CDC.
- ◆ Currently CDC holds circa 21000 records in English and 10000 records in other languages.

SP harvesting status

Service Provider	Acronym	NESSTAR	OAI-PMH	Other
Social Science Data Archives	ADP	Yes	No	No
Portuguese Social Information Archive	APIS	Yes	No	No
Austrian Social Science Data Archive	AUSSDA	No	No	Dataverse
Czech Social Science Data Archive	CSDA	Yes	No	No
Data Archiving and Networked Services	DANS	No	Yes	No
Data Centre Serbia for Social Sciences	DCS	No	No	Dataverse
Danish National Archive	DNA	Yes	No	No
<i>Estonian Social Science Data Archive</i>	<i>ESTA</i>	Yes	No	No
Swiss Centre of Expertise in the Social Science	FORS	Yes	Yes	No
Finnish Social Science Data Archive	FSD	No	Yes	No
Leibniz Institute for the Social Sciences	GESIS	Yes	Yes	No
<i>Irish Social Science Data Archive</i>	<i>ISSDA</i>	Yes	No	No
Norwegian Centre for Research Data	NSD	Yes	No	No
PROGEDO Research Infrastructure	Progedo	Yes	No	No
Slovak Archive of Social Data	SASD	No	Yes	No
Swedish National Data Service	SND	Yes	Yes	No
Social Sciences and Humanities Data Archive	SODA	No	No	No
Greek research infrastructure for the social scie	SoDaNet	Yes	No	No
Tárki Data Archive	TARKI	No	No	No
UK Data Service	UK Data Servi	Yes	Yes	No
<i>Blcocca Data Archive</i>	<i>UniData</i>	Yes	No	No

But it takes more than having a harvestable endpoint!

- ◇ CDC can be only as good as the metadata powering it.
- ◇ Better metadata => better search results, better facets
- ◇ As a rule, metadata should be corrected at the source.

An example of a CDC record

▼ Similar results

[European Values Study 2017: Finnish Data](#)

[EVS - European Values Study 1999 - Integrated Dataset](#)

[EVS - European Values Study 1981 - Integrated Dataset](#)

[EVS - European Values Study 1990 - Integrated Dataset](#)

[European Values Study: Wave 4, Great Britain, 2009-2010](#)

[← Back](#)

[</> View JSON](#)

[🔗 Go to study](#)

Study title

European Values Study 2009: Finnish Data

Creator

Pehkonen, Juhani (TNS Gallup Finland)

Kääriäinen, Kimmo (Church Research Institute)

Study Persistent Identifier

FSD2459 (FSD)

urn:nbn:fi:fsd:T-FSD2459 (The National Library of Finland)

Abstract

The survey charted a variety of values, attitudes and life situations of Finnish people. The respondents rated the importance of various domains in life (e.g. family, work, leisure time). The survey also investigated whether the respondents belonged to any voluntary organisations and participated in their activities. In addition, the respondents indicated whether they discussed politics with their friends and whether they were willing to participate financially in reducing the pollution of the environment. The respondents were asked to tell how often they see their friends and colleagues. In relation to prejudices against different groups with various characteristics, the respondents were asked which groups they would not like to have as neighbours (e.g. people of a different race, drug addicts, immigrants, homosexuals). Further questions covered general trust or mistrust of people. Views were also probed on ha

[Report a problem](#)

Methodology

Data collection period

26th June 2009 - 12th July 2009

Country

Finland

Time dimension

Longitudinal: Trend/Repeated cross-section

Analysis unit

Individual

Sampling procedure

Not available

Data collection method

Self-administered questionnaire: Web-based (CAWI)

Language of data files

FI

▼ Access

Publisher

Finnish Social Science Data Archive

Year of publication

2014

Terms of data access

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



Terms of data access

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

Study number

oai:fsd.uta.fi:FSD2459

▼ Topics

Social sciences

Tieteet humanities

Political behaviour and attitudes

Social behaviour and attitudes

Religion and values

▼ Keywords

Abortion

Education

Environmental degradation

Ethics

Families

Happiness

Immigration

Labour and employment

Living conditions

Marriage

Moral values

Political attitudes

Poverty

Prejudice

Religious communities

Satisfaction

Spouses

Trust

Unemployment

Values

Voluntary work

“Must have” metadata

- ◆ Study number <IDNo>
- ◆ Abstract <abstract>
- ◆ Study title <titl>
- ◆ Publisher <distrbtr>
- ◆ Study Url <URI>
- ◆ The language of the metadata file should be stated using <codeBook>@xml:lang, otherwise it defaults to English.

Language of metadata is important



flash 155

750 results found

Advanced search

Reset filters

Clear search

Topic ?

Collection years ?

Country ?

Publisher ?

Language of data files ?

Results per page 30

Sort by Relevance



1 2 3 4 ...



[Flash Ευρωβαρόμετρο 155 \(Δικαιοσύνη και εσωτερικές υποθέσεις\)](#)

Soufflot de Magny, Renaud (Ευρωπαϊκή Επιτροπή, Βρυξέλλες); Debyser, Ariane (Ευρωπαϊκή Επιτροπή, Βρυξέλλες)

Το συγκεκριμένο Flash Eurobarometer διερεύνησε τις απόψεις των ερωτώμενων που αφορούν τη δικαιοσύνη και τις εσωτερικές υποθέσεις.

[Flash Eurobarometer 155 \(Justice and Home Affairs\)](#)

Soufflot de Magny, Renaud (Europäische Kommission, Brüssel); Debyser, Ariane (Europäische Kommission, Brüssel)

Attitudes towards EU legislation. Topics: attitude towards the following legal issues: recognition of judicial decisions in civil and family matters throughout the EU, equal rights of defence for accused persons throughout the EU, equal rules for asylum seekers throughout the EU, validity of decisions on asylum applications in one EU member state in all other member states, equal rights of legal immigrants and national citizens, need of some sectors of the national economy for immigrant work...

[Participation and agency when ageing in place, 2012](#)

Eide, Arne Henning (SINTEF); Vik, Kjersti (Sør-Trøndelag University College)

Unwanted variations in <distrbtr>

Topic ?

Results per page 30

Sort by Relevance

Collection years ?



1 2 3 4 ...



Country ?

Publisher ?

Search publishers

Centre de données socio-politiques (45)

Centre for Socio-Political Data (1)

Danish Data Archive (1221)

Finnish Social Science Data Archive (1474)

GESIS Data Archive for the Social Sciences (5767)

GESIS Αρχείο Δεδομένων για τις Κοινωνικές Επιστήμες (121)

GESIS Αρχείο Δεδομένων για τις Κοινωνικές Επιστήμες (21)

VALU 2003 - SVT exit poll survey EMU-election 2003

Holmberg, Sören (University of Gothenburg, Department of Political Science);

Thedéén, Torbjörn (Royal Institute of Technology, Division of Safety Research); Hernborn, Hans (Sveriges Television); (1 more)

September 14, 2003 was the date for the referendum on Sweden's possible participation in the third stage of the Economic and Monetary Union (EMU). The question that all those entitled to vote had to answer was: 'Do you

...the referendum an exit poll survey ...tions are asked by public service

...residents of Gothenburg and its ...tire Västra Götaland County plus

...Enable SOM researchers to study attitudes and behaviour linked to local and regional issues. They emphasise public services and media, although many of the questions are identical with those used in N...

"Should have" metadata

- ◊ Topic <topcClas>
- ◊ Keyword <keyword>
- ◊ Time dimension <timeMeth>
- ◊ Country <nation>
- ◊ Analysis unit <anlyUnit>
- ◊ Year of publication <distDate>
- ◊ Study Persistent Identifier <IDNo>
- ◊ Creator <AuthEnty>
- ◊ Language of data file(s) <fileDscr><fileTxt>@xml:lang>
- ◊ Data collection method <collMode>
- ◊ Data collection period: start date
<stdyDscr><stdyInfo><sumDscr><collDate>@date @event="start">
- ◊ Data collection period: end date
<stdyDscr><stdyInfo><sumDscr><collDate>@date @event="end">
- ◊ Terms of data access <restrctn>

Strongly recommended: use topic and use it properly

▼ Topic



Results per page

30

Sort by

Relevance

Search topics

ADMINISTRATIVE BURDEN (1)

ADMINISTRATIVNE OVIRE (1)

AFRICA AND THE EUROPEAN UNION (1)

AFRIKA IN EVROPSKA UNIJA (1)

AGRICULTURE (2)

AIR TRANSPORT SERVICES (1)

AKTERI I ODLOČANJE, ORVEŠČANJE (1)

► Collection years

► Country

► Publisher

► Language of data files



connection with the referendum an exit poll survey was carried out. In VALU - Swedish exit poll surveys - voters leaving polling stations are asked by public service broadcaster Sveriges Television (SVT) to fill in a... sible participation in the that all those entitled to euro as its currency? In

Report a problem

Use both the ISO country code in the attribute and country name

Collection years

Country ?

Search countries

(1)

48 držav je članic ISSP skupine. Skozi leta in posamezne izvedbe so sodelovale različne države. Podrobno informacijo najdete v opisu posamezne raziskave. (1)

ANDALUZIJA (1)

ARGENTINA (1)

ARMENIJA (1)

AVSTRALIJA (1)

[VALU 2003 - SVT exit poll survey EMU-election 2003](#)
*Holmberg, Sören (University of Gothenburg, Department of Political Science);
Thedéen, Torbjörn (Royal Institute of Technology, Division of Safety Research);*

[The Regional Western Sweden SOM Survey 2017](#)
University of Gothenburg, SOM Institute

Western SOM has been conducted yearly since 1992, but was initially limited to Gothenburg and its surrounding municipalities. In 1998, the survey was extended to the entire Västra Götaland County plus the municipality of Kungsbacka. The purpose of the surveys is to enable SOM researchers to study attitudes and behaviour linked to regional issues. They emphasise public services and media, although many of

sible p
that all
euro a
VALU -
ce bro

Mostly unused: lang of data files

▶ Collection years



▶ Country



▶ Publisher



▼ Language of data files



× en (7320)

× fi (1326)

× sk (3)



1

2

3

4

...



[Estudo Eleitoral Português, 2005](#)

Lobo, Marina Costa (ICS-UL); Magalhães, Pedro (ICS-UL); Barreto, António (ICS-UL); (1 more)

O Estudo Eleitoral Português 2005 é um inquérito pós-eleitoral conduzido junto da população adulta portuguesa com o objectivo de medir o comportamento eleitoral (participação e opção de voto) em eleições legislativas, assim como uma série de potenciais correlatos, incluindo exposição aos media, discussão política e uma série de atitudes e valores políticos. Inclui o módulo 2 do Comparative Study of Electoral Systems e várias perguntas do Comparative National Elections Project.

[Informação sobre saúde da população portuguesa, 2012](#)

Cabral, Manuel Villaverde (Institute of Social Sciences, University of Lisbon); Barros, José Henrique Dias Pinto de (Institute of Public Health, University of Porto); Carmo, Maria Isabel Augusta Cortes do (Faculty of Medicine, University of Lisbon); (5 more)

O conhecimento sobre os fatores que promovem e defendem a saúde, enquanto fatores de prevenção das doenças, e sobre fatores de risco para a saúde, é um dos principais determinantes de comportamentos e estilos de vida salutogénicos. Assim sendo, a efetividade das intervenções de saúde pública passa por compreender o nível de conhecimento da população visado de forma a planear de forma estratégica ações de sensibilização e de transmissão de con

Report a problem

About dates

- ◇ Strongly recommended:
 - ◇ Year of publication
 - ◇ Data collection period (start and end)
- ◇ Ideally: 'YYYY-MM-DDThh:mm:ssZ'
 - ◇ can accept 'YYYY-MM-DD', 'YYYY-MM' or 'YYYY'

Check your metadata in CDC

- ◊ Choose your own organisation from Publisher list
- ◊ See the lists in other filters to chase incorrect country names, different spellings of organisation names, topic classification problems etc.

Feedback welcome



Find Social and Economic Research Data

21844 results found

Advanced search

Reset filters

Clear search

Topic



Collection years



Country



Publisher



Language of data files



Search languages

Results per page

30

Sort by

Relevance



1

2

3

4

...



[VALU 2003 - SVT exit poll survey EMU-election 2003](#)

*Holmberg, Sören (University of Gothenburg, Department of Political Science);
Thedéen, Torbjörn (Royal Institute of Technology, Division of Safety Research);
Hernborn, Hans (Sveriges Television); (1 more)*

September 14, 2003 was the date for the referendum on Sweden's possible participation in the third stage of the Economic and Monetary Union (EMU). The question that all those entitled to vote had to answer was: 'Do you think Sweden should introduce the euro as its currency?' In connection with the referendum an exit poll survey was carried out. In VALU - Swedish exit poll surveys - voters leaving polling stations are asked by public service broadcaster Sveriges Television (SVT) to fill in a...

Report a problem

See also last slide for contact email addresses!

Metadata issues in bitbucket

CESSDA / CESSDA Managed Content / cessda.metadata.office

































Issues

Create issue

Filter by: All **Open** My issues Watching

Advanced search

Issues (1–22 of 22)

Title ▾	T	P	Status	Votes	Assignee	Version	Created	Updated	
#19: Extract Study Url from NESSTAR records (Progedo)	CDC	✓	⬆️	OPEN	 Genevieve MICH	CMM 1....	2019-06-05	2019-10-31	
#26: Differing Topic term variant (GESIS)	CDC	■	⬆️	OPEN	 Wolfgang Zenk-N	CMM 1....	2019-08-07	2019-10-31	
#11: Workflow for MDO issue tracker	CDC	✓	⬆️	OPEN	 John Shepherdso	CMM 1....	2019-05-16	2019-10-31	
#18: Extract Study Url from NESSTAR records (NSD)	CDC	✓	⬆️	OPEN	 John Shepherdso	CMM 1....	2019-06-05	2019-10-31	
#4: Inconsistent values for Publisher (NSD)	CDC	✓	⬆️	OPEN	 John Shepherdso	CMM 1....	2019-04-18	2019-10-31	
#21: Extract Study Url from NESSTAR records (CSDA)	CDC	✓	⬆️	OPEN	 Tomáš Čížek	CMM 1....	2019-06-05	2019-10-23	
#30: Country filter values: many different values in use	CDC	✓	⬆️	NEW	 Andre Foerster	CMM 1....	2019-08-15	2019-09-11	
#2: Harvesting of HTML href and other embedded elements broken (UKDA)	CDC	✓	⬆️	OPEN	 John Shepherdso	CMM 1....	2019-04-18	2019-09-06	
#15: Inconsistent values for Publisher (CDSP)	CDC	✓	⬆️	NEW	 Andre Foerster	CMM 1....	2019-06-05	2019-09-02	
#29: Topic filter values: many different formats in use	CDC	✓	⬆️	NEW	 Andre Foerster	CMM 1....	2019-08-15	2019-08-21	
#13: Some of the SoDaNet metadata records don't have language tags	CDC	✓	⬆️	OPEN	 alinardis	CMM 1....	2019-05-29	2019-07-10	
#25: Harvest of data from DNA	CDC	✓	⬆️	OPEN	 Andre Foerster	CMM 1....	2019-07-10	2019-07-10	
#22: Extract Study Url from NESSTAR records (Unidata)	CDC	✓	⬆️	NEW	 Carlo Maria Pisan	CMM 1....	2019-06-05	2019-06-17	
#3: Inconsistent values for Country (UKDS)	CDC	✓	⬆️	OPEN	 Sharon Bolton	CMM 1....	2019-04-18	2019-06-12	
#24: Inconsistent values for Country (SoDaNet)	CDC	✓	⬆️	NEW	 alinardis	CMM 1....	2019-06-12	2019-06-12	
#23: Inconsistent values for Country (ADP)	CDC	✓	⬆️	NEW	 Irena Vipavc Brva	CMM 1....	2019-06-12	2019-06-12	

Summary: remember to...

- ◇ record the metadata language
- ◇ use controlled vocabularies whenever possible
- ◇ use attribute vocab name when using CESSDA topic classification
- ◇ be consistent
 - ◇ for example, don't use many variants of the name of your organization unless the name has really changed
- ◇ share experiences, provide feedback

Forthcoming

- ◆ Bug fixes
- ◆ Machine-actionable CDC metadata profile and a tool that can check a metadata file against the profile (see [announcement in SPF Basecamp](#)).
- ◆ Training videos for researchers and for SPs
- ◆ In the future: variable information

Further information

- ◇ CDC User Guide <https://datacatalogue.CESSDA.eu/documentation/index.html>
- ◇ EDDI2018 presentation: The New CESSDA Catalogue <https://zenodo.org/record/2530106#.XJliRSj7SUK>
- ◇ CESSDA Vocabulary Service <https://vocabularies.CESSDA.eu/>
- ◇ CESSDA Metadata Model <https://zenodo.org/record/3543756#.Xda7ZkGxVhE>
- ◇ CESSDA Metadata Model User Guide <https://zenodo.org/record/3236194#.Xda7S0GxVhE>
- ◇ CV Usage <https://docs.google.com/spreadsheets/d/10yeLWJnXU6-mIQTs-dXMhYb4LvVbvtNb3L9bhuzA-g0/edit#gid=1056082761>
- ◇ Metadata Office Basecamp <https://3.basecamp.com/3584575/projects/1716357> (not public)
- ◇ *metadata-office@CESSDA.eu* / metadata issues
- ◇ *support@CESSDA.eu* / getting your collection into CDC

Thank you!

 cessda.eu

 @CESSDA_Data

