CESSDA

How to Find and Access Data in Europe

Irena Vipavc Brvar (ADP - Slovenian Social Science Data Archives)
Jennifer Buckley (UKDS – United Kingdom Data Service)

A practical introduction

cessda.eu @CESSDA_Data

CC BY SA

Cite as: Vipavc Brvar I., J. Buckley (2020). CESSDA Data Management Expert Guide. Chapter 7. Discover: How to Find and Access Data in Europe [presentation]. Bergen: CESSDA ERIC.

Overview

- Data types and sources
- Identify what you need
- Searching data archives
- Evaluating data: quality and usefulness
- Accessing data

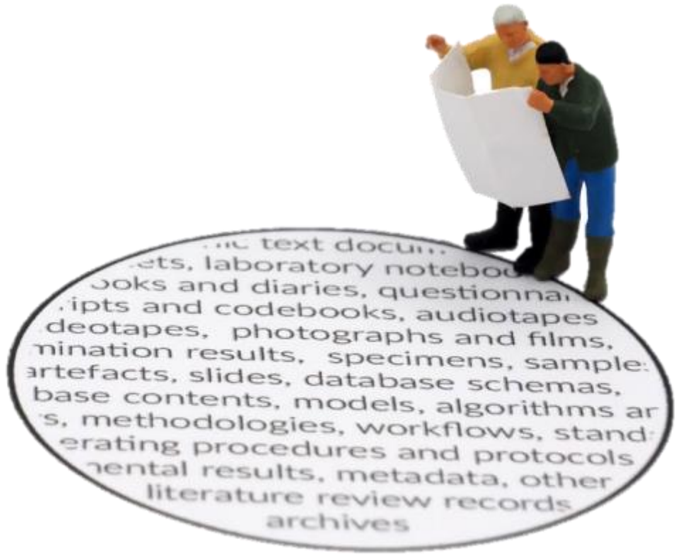
Data types and sources

Activity 1
Your knowledge and experience of the data landscape

- Introduce yourself
- Tell us about your research work (current, future, past)
- Did you use or you intend to use available data for your work? Tell us about it.

Types of data

Thinking about the types of data available can help you work out what you need and how to find it.



Quantitative and Qualitative

Types of data: level of analysis

Macro data

- Aggregate
  - about populations, groups, regions and countries constructed by combining information on lower level units (e.g. unemployment rate, fertility)
- System level
  - characteristics of higher-level units such as the state or the political system e.g. electoral system (PR or single-member districts) and member of EU

Meso data

- data on collective and cooperative actors such as commercial companies, organizations or political parties

Microdata

- data from individual units (often people or households) often from surveys, a census and administrative records

Types of data: time

Cross-sectional	• one-point of time (a snap shot) • usually information on multiple cases and variables
Repeated cross sectional	• cross-sectional surveys repeated with new samples • data from the different samples allows analysis of trends
Time series	• series of data points in time order (often equally spaced in time) • aggregate macro data are often time-series data. • time points may come from sample surveys e.g. unemployment from labour force surveys
Longitudinal	• follow the same units over time e.g. household panel studies collect information from a sample of households in regular 'waves'

New data types

- Big Data, which are often described by their attiributes (3 Vs):
  - *Volume* means that Big Data are very large and that processing them demands great computational power.
  - *Velocity* stands for the fact that Big Data are produced successively and new data emerge every moment.
  - *Variety* reminds us that Big Data are unstructured and messy and thus not ready for immediate analysis.

- Depending on their source, the OECD defines six categories of Big Data:
  - A: Data stemming from the transactions of government, for example, tax and social security systems.
  - B: Data describing official registration or licensing requirements.
  - C: Commercial transactions made by individuals and organisations.
  - D: Internet data, deriving from search and social networking activities.
  - E: Tracking data, monitoring the movement of individuals or physical objects subject to movement by humans.
  - F: Image data, particularly aerial and satellite images but including land-based video images.

Social media data

- Social media data (category D in the OECD taxonomy) are the data from platforms like Facebook, Twitter, Instagram or YouTube,
- These data are created by the users of such platforms.
- Researchers can access these data in three main ways:
  - 1) Direct cooperation with the companies/platforms,
  - 2) Buying from data resellers,
  - 3) Via APIs (one might add web scraping to the list but most platforms/companies discourage its use).

Sources of microdata

CC BY SA

CESSDA Training Working Group (2017)

There are many sources of data.

European social science data archives

Data collections include:

- variation between archives
- quantitative data - major source of individual level data
- qualitative
- outputs of
  - major academic projects
  - government/policy
  - small research teams
  - individual researchers
- recent and less recent data
- different languages

SND Swedish National Data Service

Leibniz Institute
for the Social Sciences

cessda

Consortium of European Social Science Data Archives

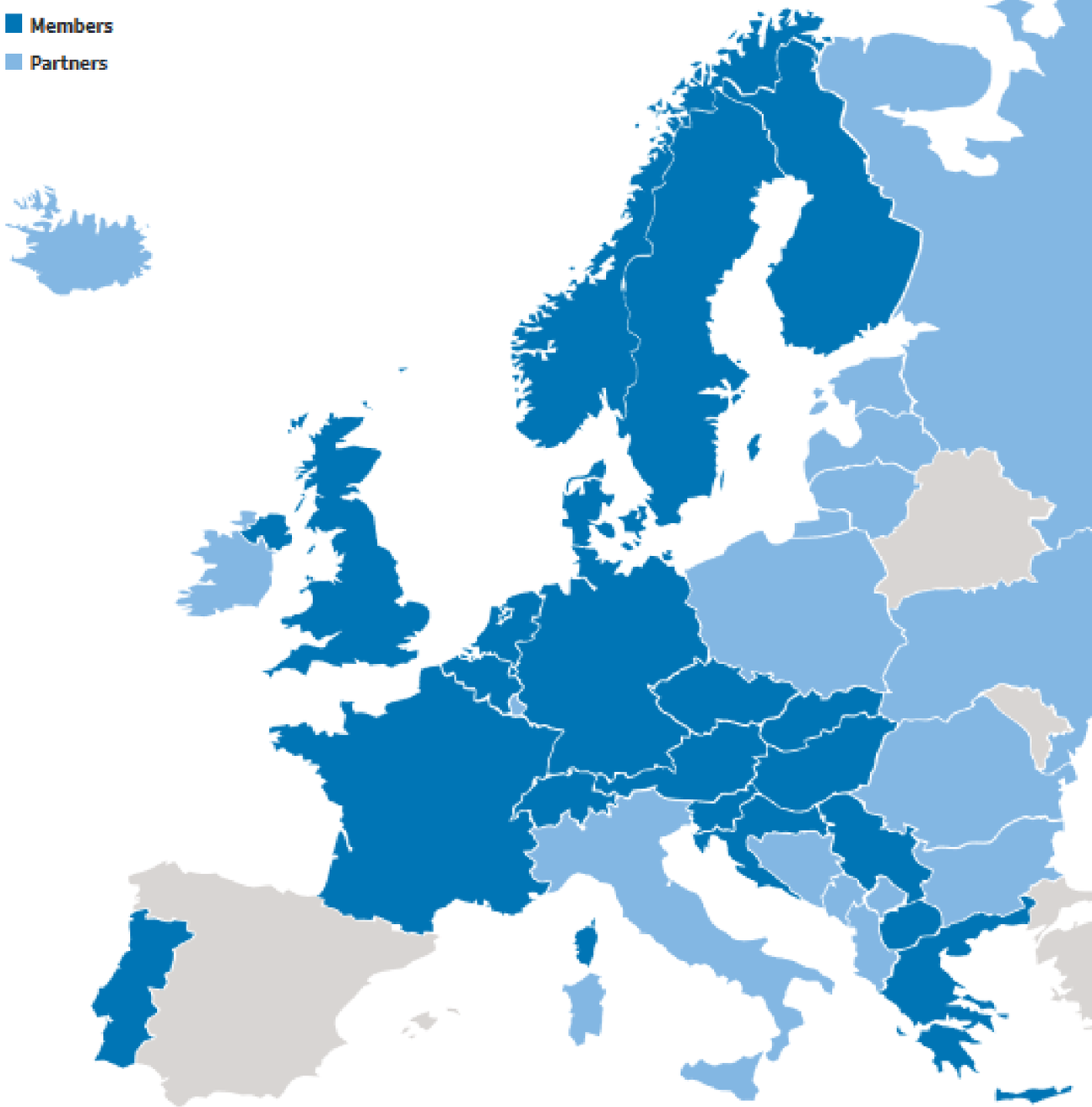
"Enabling the research community to conduct high-quality research in the social science"

Key tasks:

- Developing standards and best practices around the management and archiving of social science data.
- Facilitating access to important data resources
- Work done by developing tools, training and co-ordinating network.
- CESSDA data catalogue.

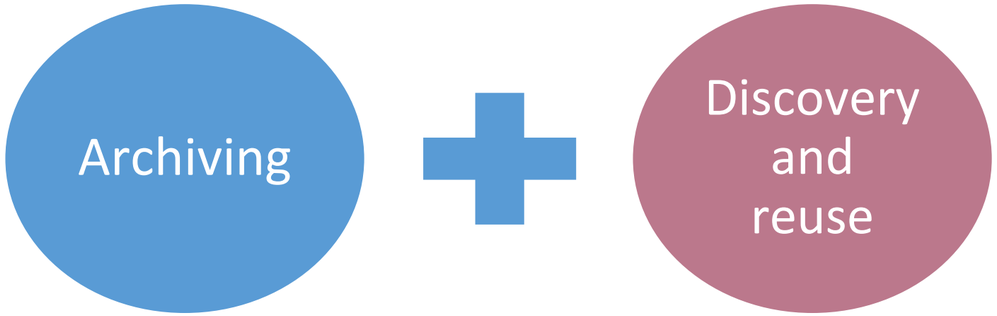
Members

- Austria
- Belgium
- **Croatia**
- Czech Republic
- Denmark
- France
- Finland
- Germany
- Greece
- Hungary
- Netherlands
- **North Macedonia**
- Norway
- Portugal
- **Serbia**
- Slovakia
- Slovenia
- Sweden
- Switzerland
- UK



cessda

National data services



Activities include:

- checking the quality of data and metadata,
- maintaining catalogues,
- managing access to data through appropriate licensing,
- obtaining data and
- training for both those creating and using data.

Open Access to research data (European Commission)

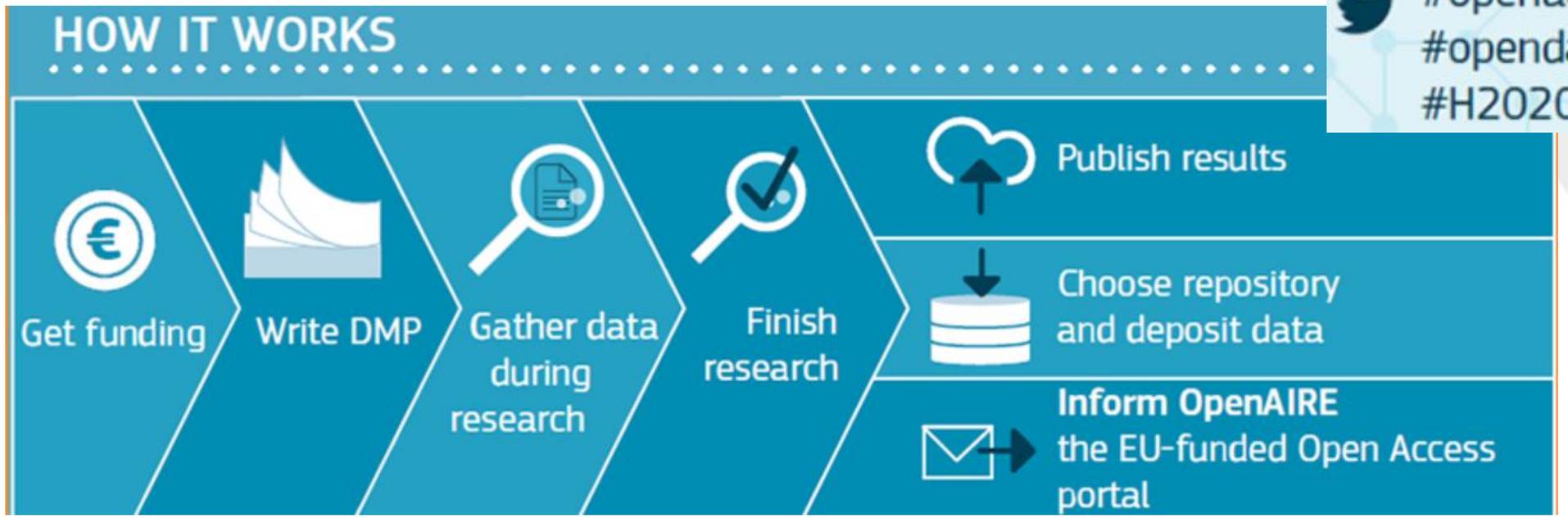
Open access (OA) can be defined as **the practice of providing on-line access to scientific information that is free of charge to the user and that is re-usable**.

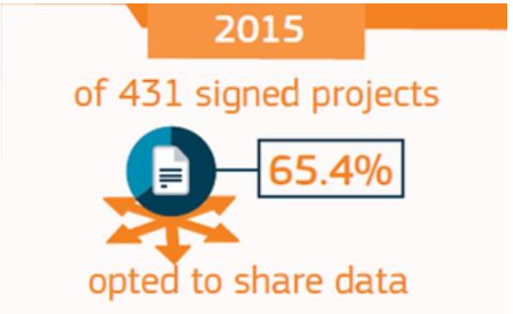
Open access to 'scientific information' refers to two main categories:

- **Peer-reviewed scientific publications** (primarily research articles published in academic journals)
- **Scientific research data: data underlying publications and/or other data** (such as curated but unpublished datasets or raw data)

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

#openaccess
#opendata
#H2020

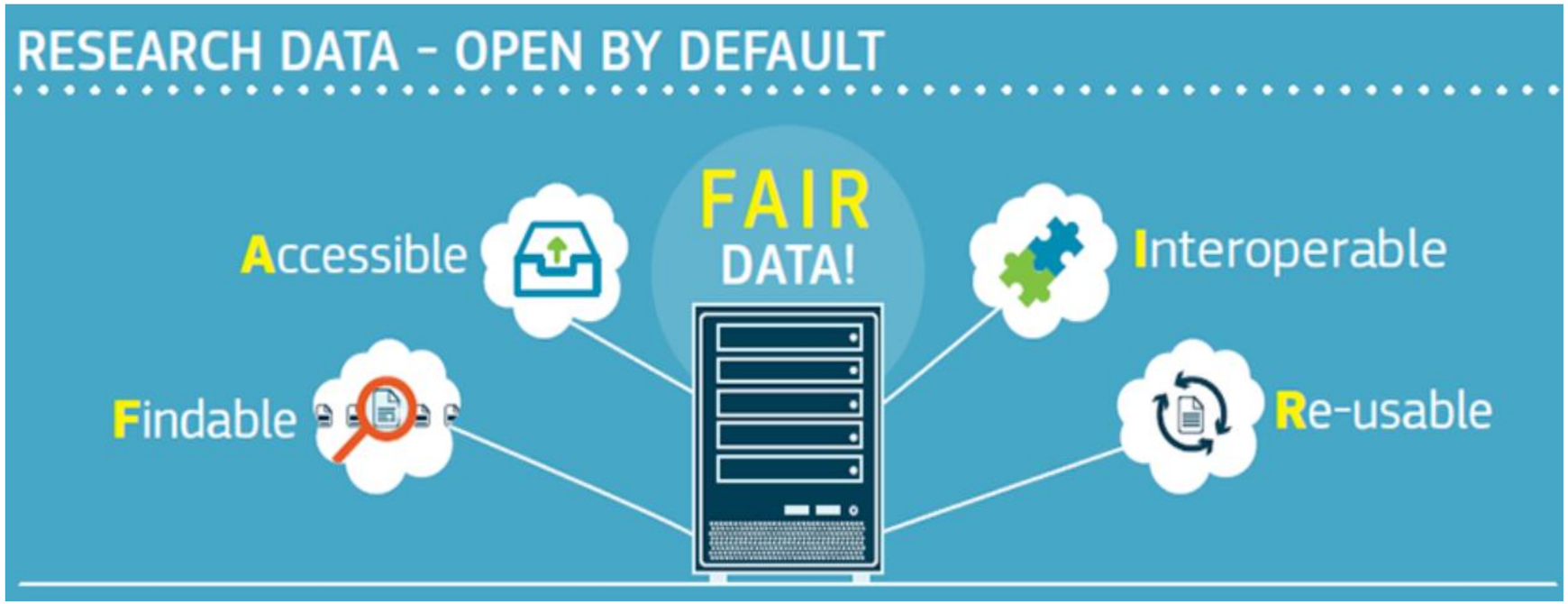




Source: EC

Open Access to research data

Importance of research infrastructures / data repositories



***Source*: EC**

Horizon Europe

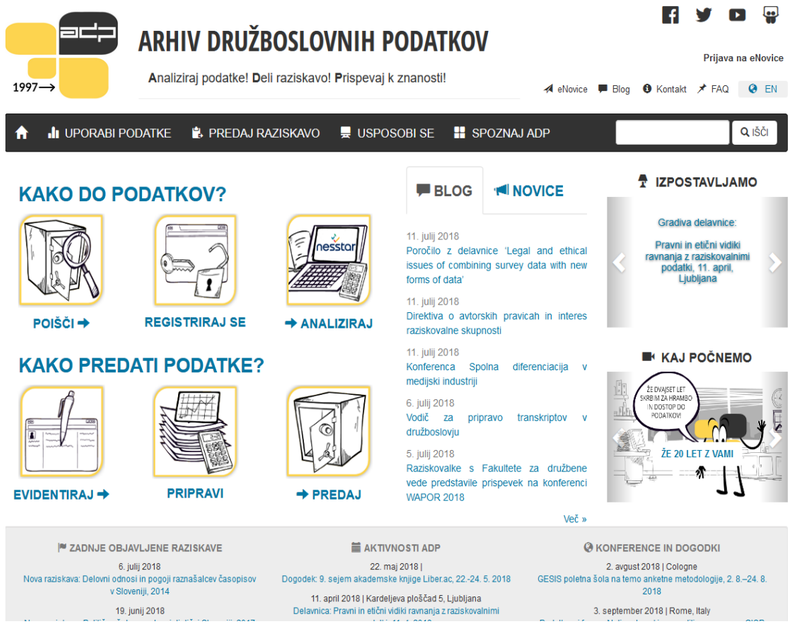
2021-2027

Reinforce openness

Open Science will become the modus operandi of Horizon Europe. It will go beyond the open access policy of Horizon 2020 and require open access to publications, data, and to research data management plans.

Slovenian Social Science Data Archives (ADP)

- Founded in 1997
- Slovenian national data repository for social sciences
- 600 social science surveys with data in a data catalogue + 150 with metadata
- Cca. 800 users registered in 2017 (90 % education, 10 % scientific/research purpose)
- 168 survey data used for detailed secondary-analysis in 2017



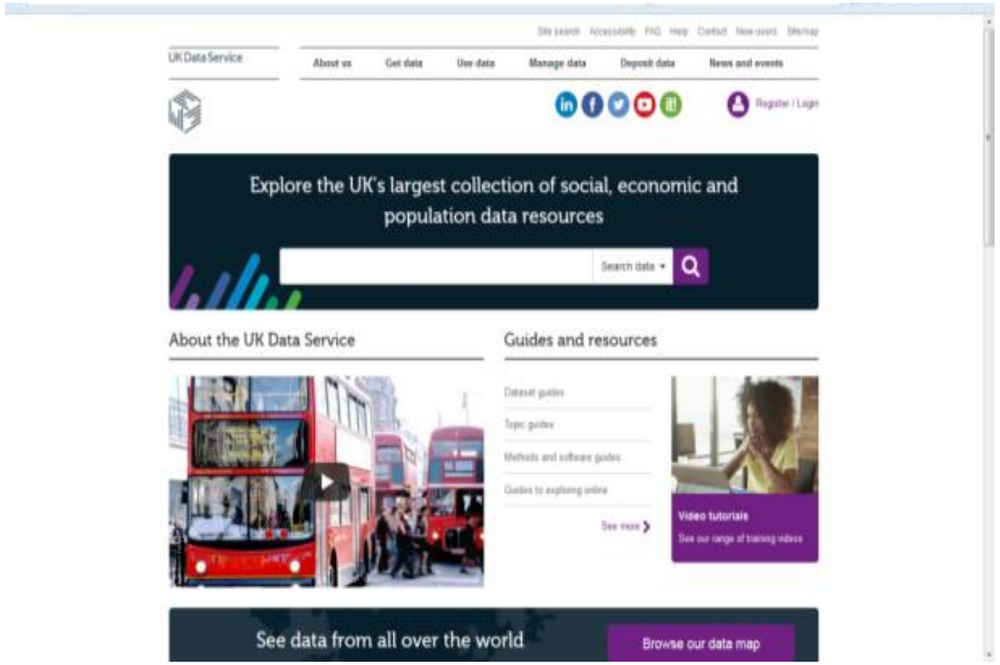
- **Oldest data sets in the archive (public opinion polls) are from 1966**
- **Wide range of topics covered**
- **In most cases data relates only to Slovenia / few international**
- **Metadata in SI and EN, datafiles mostly in SI**

cessda

UK Data Service

Access to the UK's largest collection of social, economic and population data

Support for users with training and guidance.

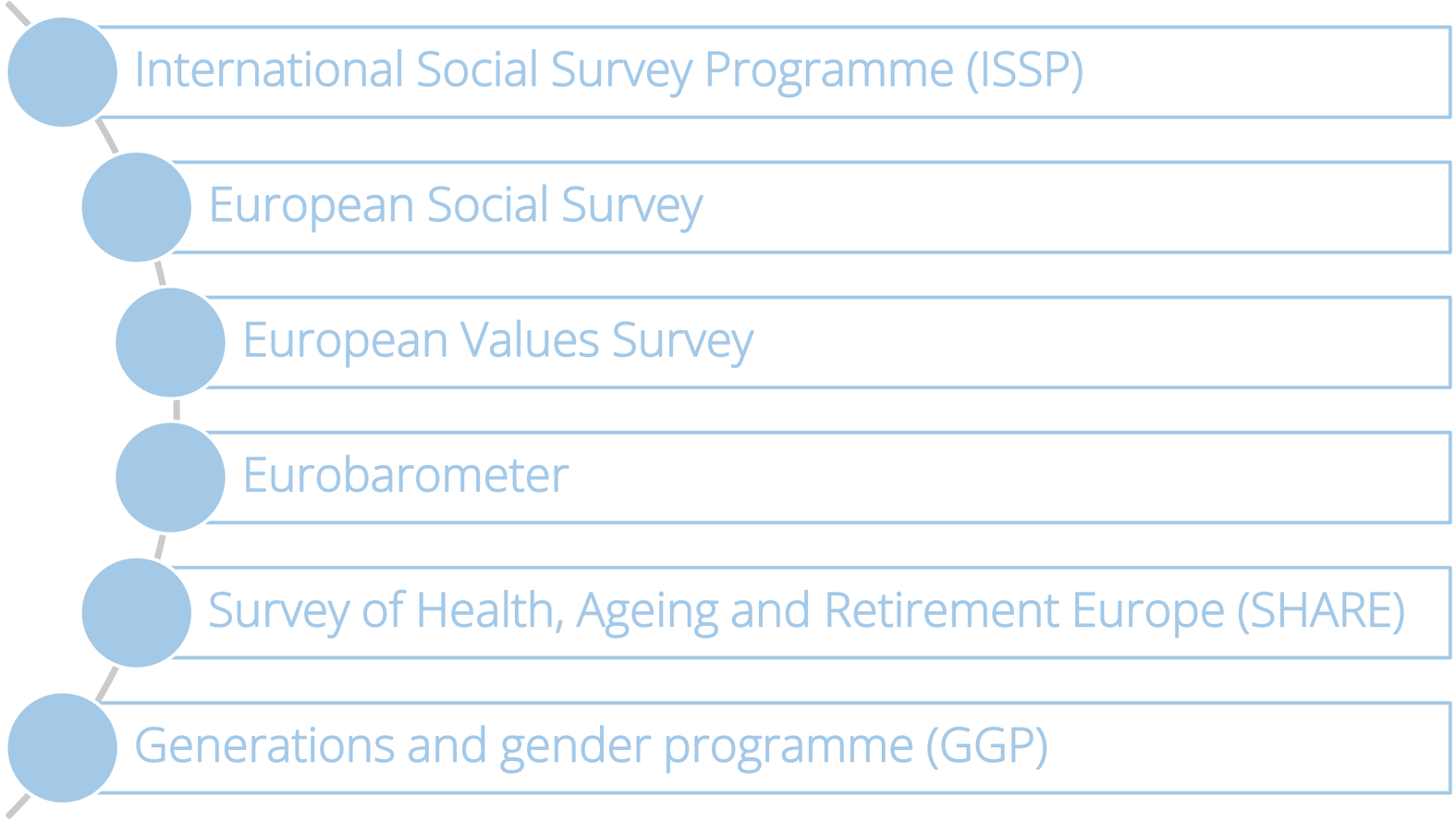


- **major UK and cross-national surveys**
- **longitudinal studies (household panel and cohort studies)**
- **UK Census 1971-2011**
- **qualitative data collections**
- **research data in a researcher repository (Reshare)**

UK data service

Cross-national studies

International survey research programmes include many European countries



International Social Survey Programme (ISSP)

- annual programme (started in 1984)
- cross-national collaboration
- rotating thematic modules e.g.
  - Citizenship: 2004 and 2014
  - Work Orientations: 1989, 1997, 2005, 2015
  - Role of Government: 1985, 1990, 1996, 2006, 2016

Data Download

Data Catalogue

gesis Leibniz-Institut für Sozialwissenschaften

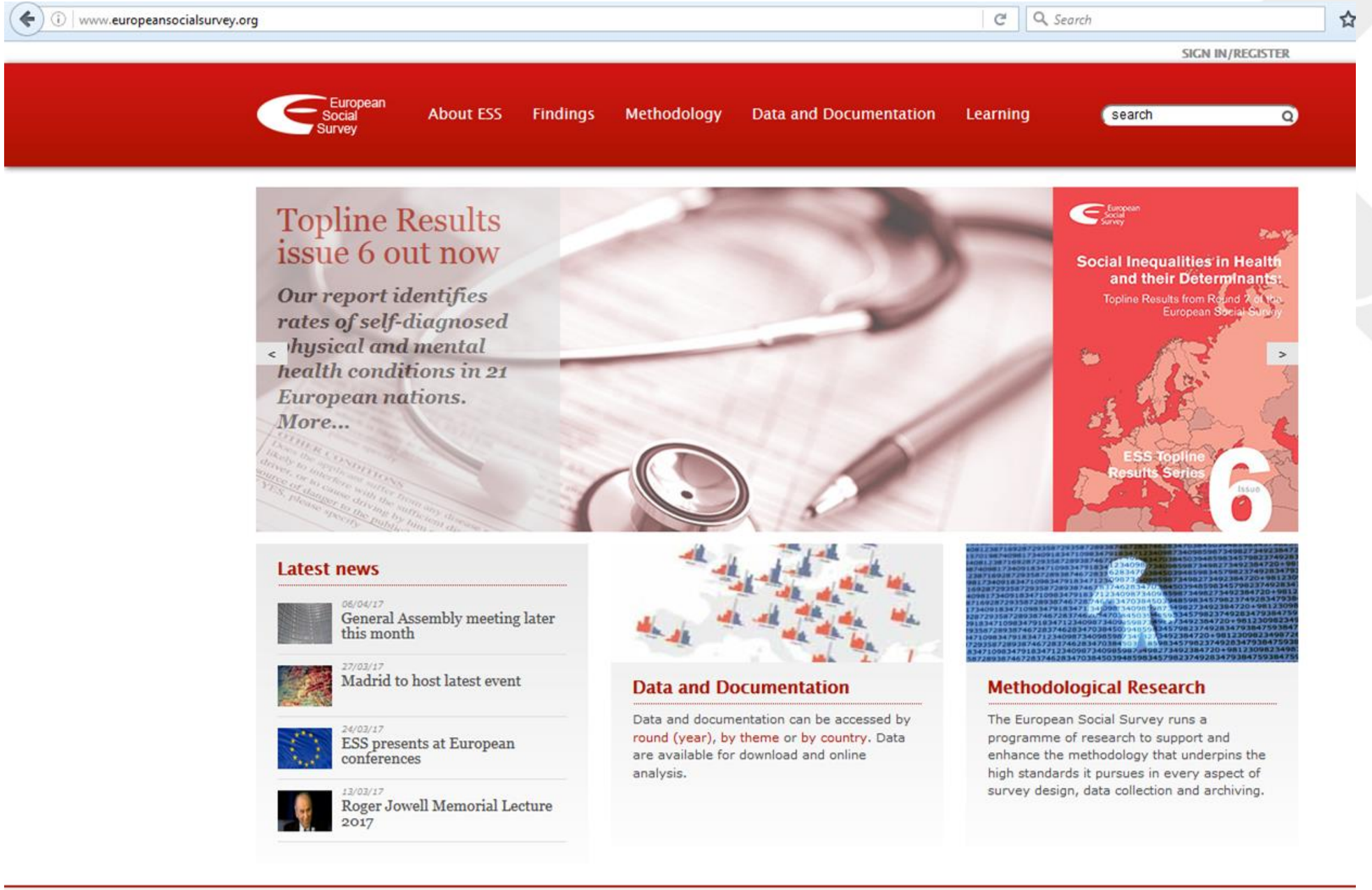
European Social Survey (ESS)

- A biennial cross-national survey (started in 2002)
- Highest methodological standard
- Freely available data for 36 countries (23 countries in 2016)

	R1 2002	R2 2004	R3 2006	R4 2008	R5 2010	R6 2012	R7 2014	R8 2016
Media and social trust	•	•	•	•	•	•	•	•
Politics	•	•	•	•	•	•	•	•
Subjective well-being...	•	•	•	•	•	•	•	•
Gender, Household	•	•	•	•	•	•	•	•
Socio demographics	•	•	•	•	•	•	•	•
Human values	•	•	•	•	•	•	•	•
Immigration	•						•	
Citizen involvement	•							
Health and care		•						
Economic morality		•						
Family work and well-being		•			•			

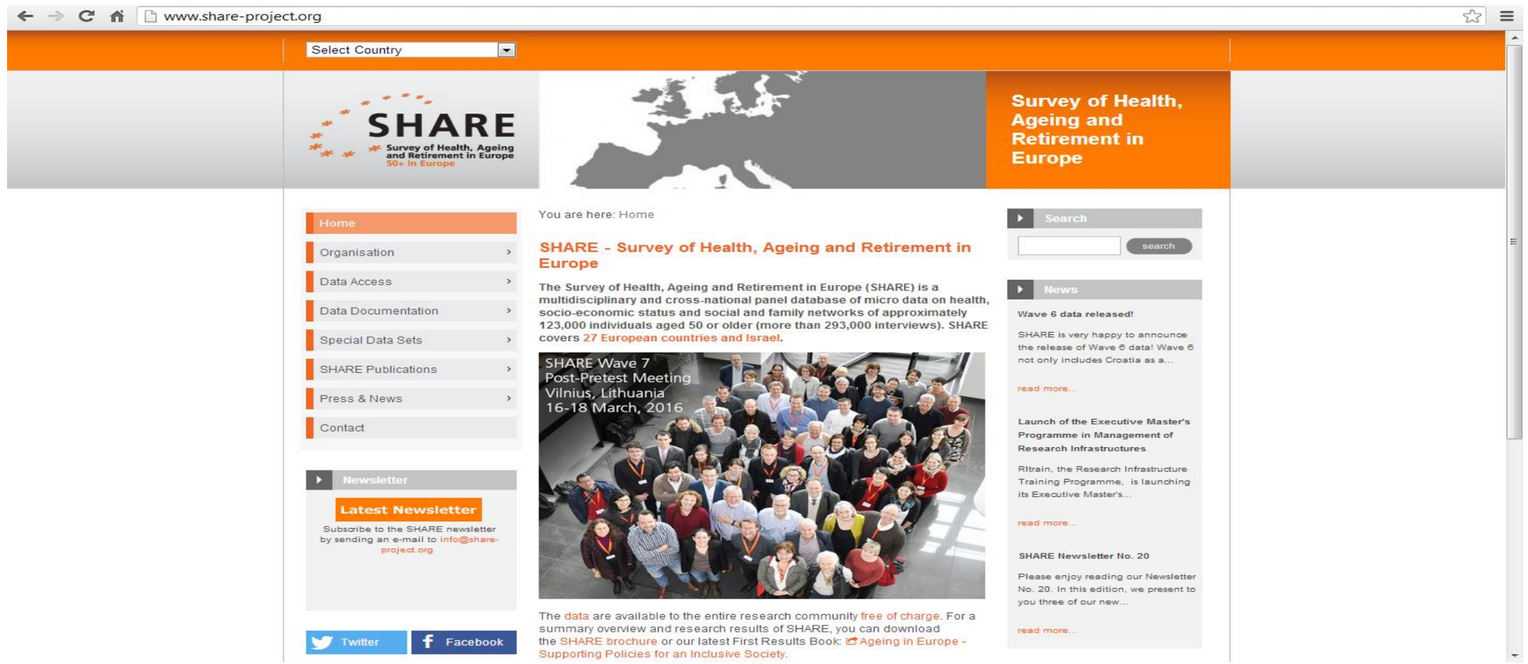
Probably most used / cited data. 125 T registered users, 89 T data downloads

Source: ESS


www.europeansocialsurvey.org
Search
SIGN IN/REGISTER
European Social Survey
About ESS
Findings
Methodology
Data and Documentation
Learning
search
Topline Results issue 6 out now
Our report identifies rates of self-diagnosed physical and mental health conditions in 21 European nations. More...
Social Inequalities in Health and their Determinants:
Topline Results from Round 7 of the European Social Survey
ESS Topline Results Series
6
Issue
Latest news
06/04/17
General Assembly meeting later this month
27/03/17
Madrid to host latest event
24/03/17
ESS presents at European conferences
13/03/17
Roger Jowell Memorial Lecture 2017
Data and Documentation
Data and documentation can be accessed by round (year), by theme or by country. Data are available for download and online analysis.
Methodological Research
The European Social Survey runs a programme of research to support and enhance the methodology that underpins the high standards it pursues in every aspect of survey design, data collection and archiving.

Survey of Health, Ageing and Retirement in Europe (SHARE)

- longitudinal study
- more than 140,000 individuals aged 50
- 27 European countries and Israel
- micro data on health, socio-economic status and social and family networks



cessda

Examples: Longitudinal studies

Household panel studies

following households over time and asking questions on a broad range of topics such as household composition, employment, earnings, health, social and political participation and life-satisfaction

- **German Socio-Economic Panel (SOEP)**
- **Understanding society (and the British Household Panel Study)**
- **Swiss Household** Panel

Five key data providing organizations

- Eurostat – Statistics office of European Union
- LIS - harmonised socio-economic micro datasets
- OECD – key source of comparable statistical, economic and social data
- World Bank - Free and open access to global development data
- IMF - time series data on economic and financial indicators

Eurostat

Statistical office of the European Union

Provides national and sub-national data

- economy and finance, population and social conditions, industry, trade, agriculture and fisheries, transport, environment and energy and science, technology and innovation

Microdata

- e.g European Community Household Panel, European Union Labour Force Survey, European Union Statistics on Income and Living Conditions

Metadata for Official Statistics

MISSY (**Microdata Information System**) is an online service platform that provides structured metadata for official statistics. MISSY includes metadata at the study and variable level as well as reports and tools for data handling and analysis. All documentation in MISSY refers to microdata available for scientific purposes. MISSY currently documents the following official statistics microdata:

EU-Data

- AES (Adult Education Survey)
- CIS (Community Innovation Survey)
- EU-LFS (European Union Labour Force Survey)
- EU-SILC (European Union Statistics on Income and Living Conditions)
- SES (Structure of Earnings Survey)

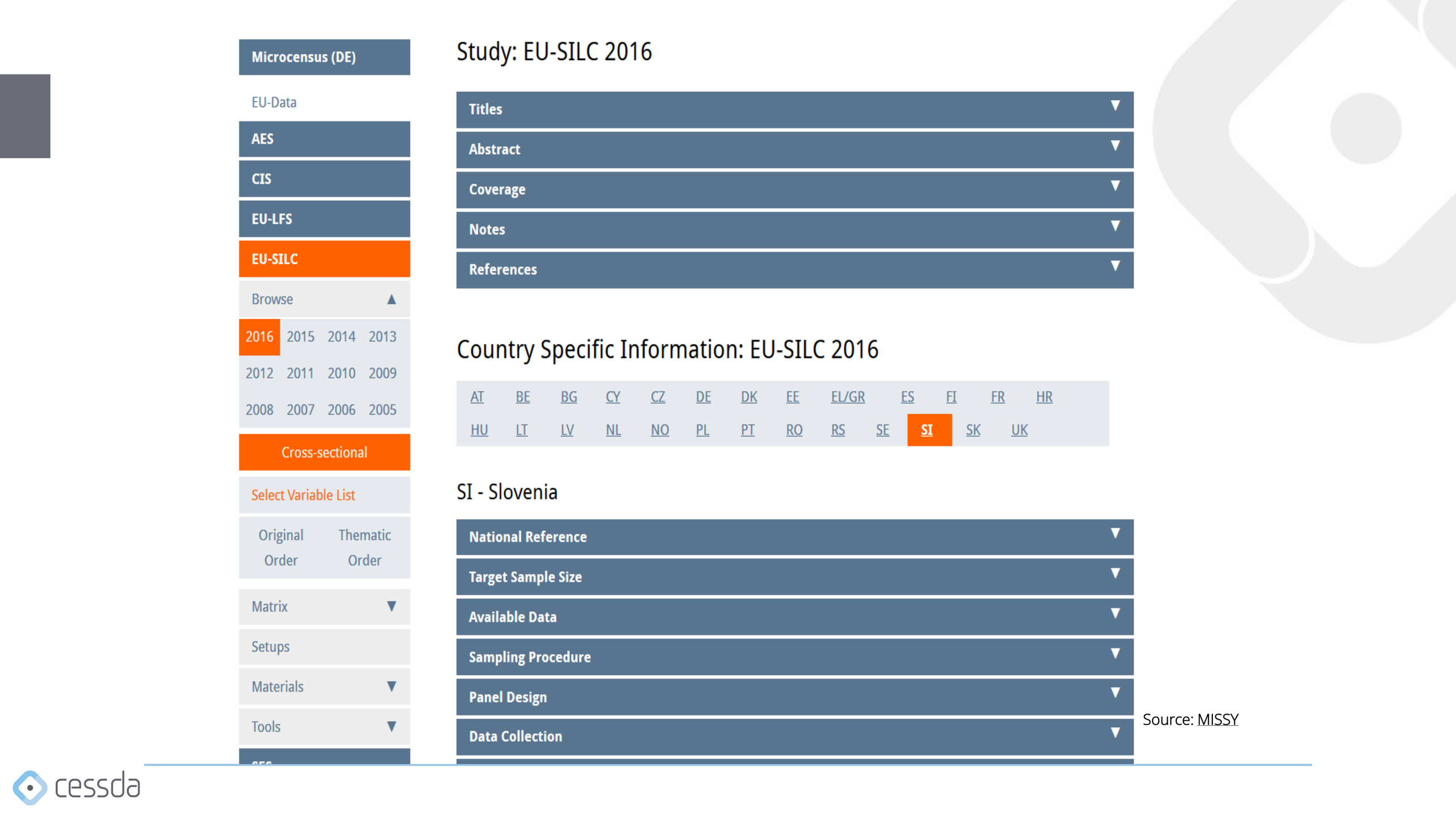
National Data

- MZ (German microcensus)

fdz German Microdata Lab

gesis Leibniz-Institut für Sozialwissenschaften

cessda

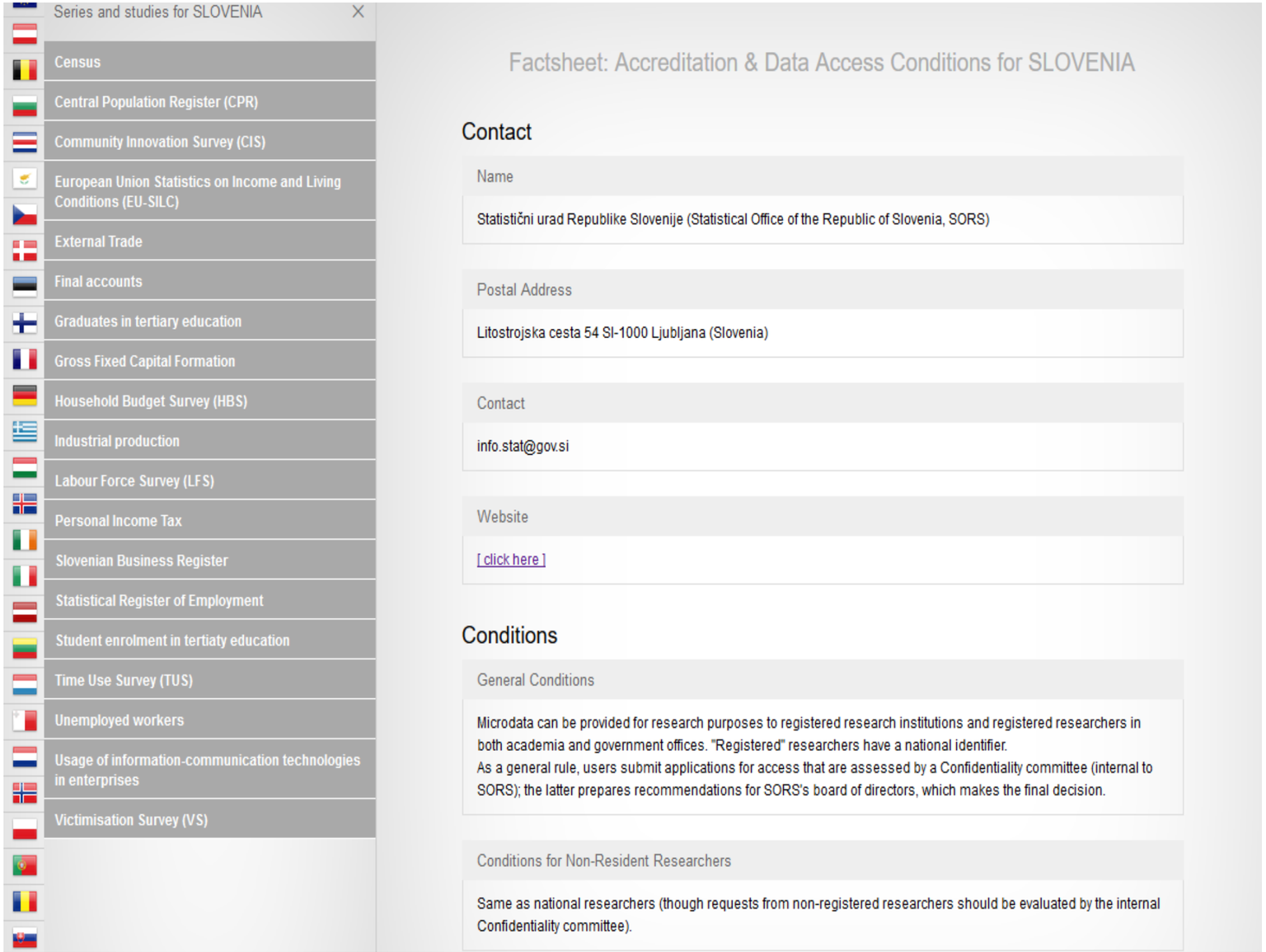


Source: MISSY

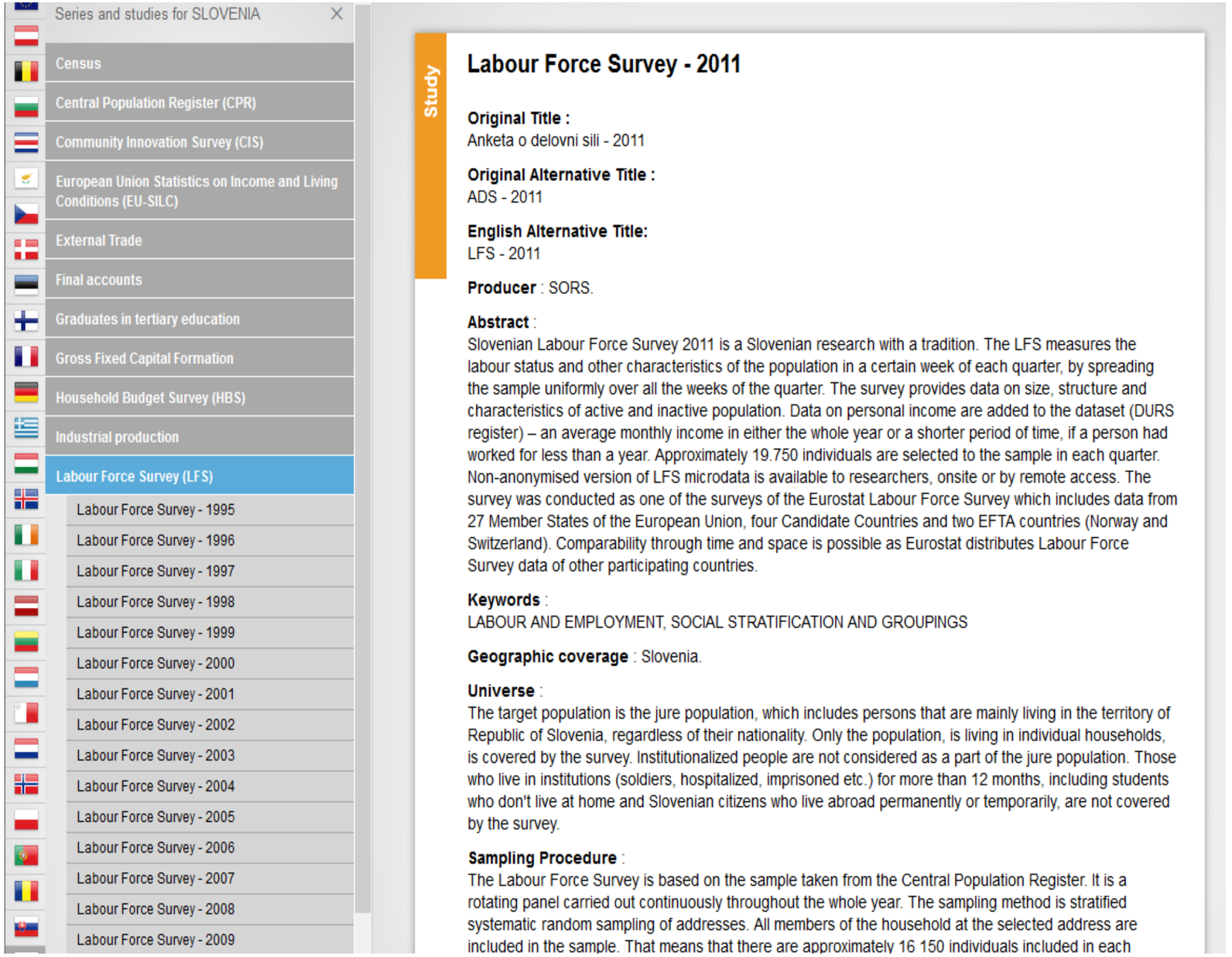
cessda

CIMES

Centralising and Integrating Metadata from European Statistics



Source: CIMES



CIMES
Centralising and Integrating Metadata from European Statistics

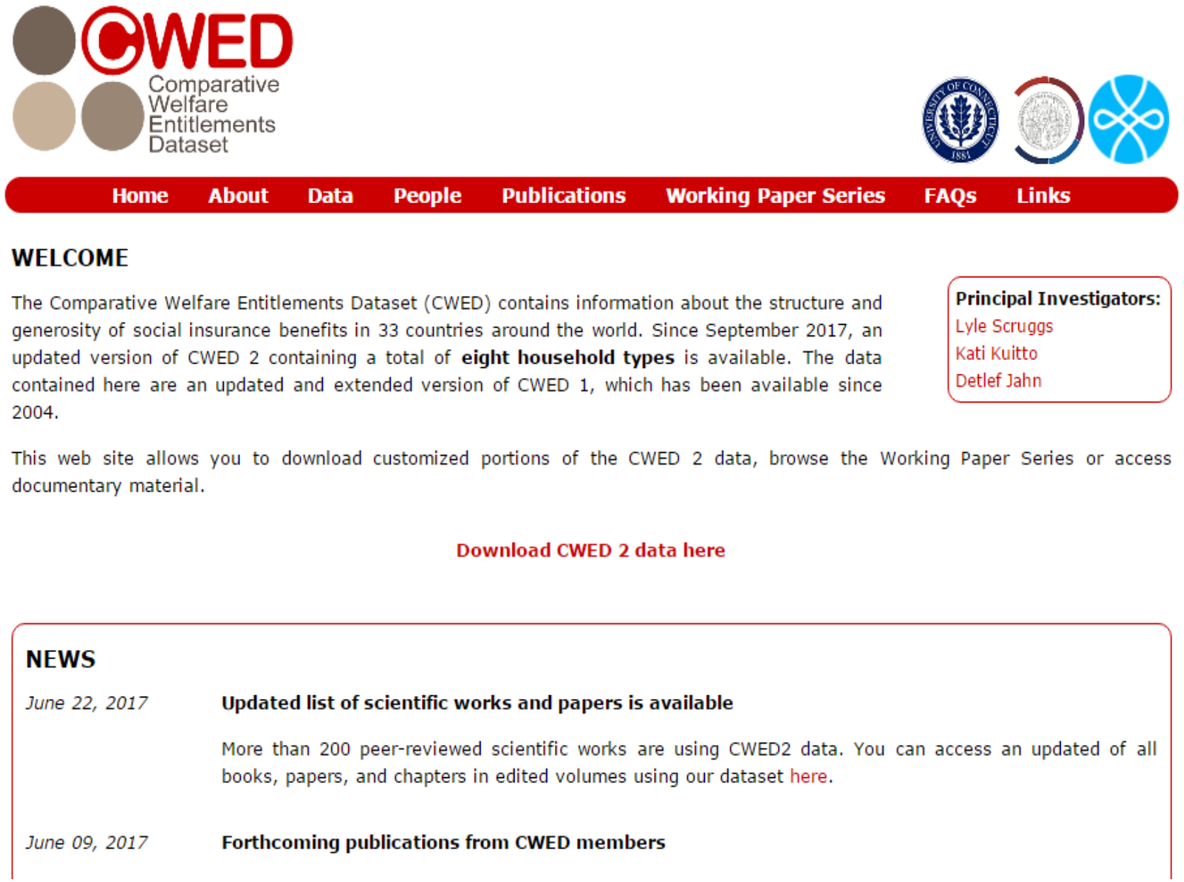
Source: CIMES

cessda

Direct from project websites

Some research projects share research data through project websites

http://cwed2.org/



Data repositories

Digital archives collecting, preserving and displaying datasets, related documentation and metadata.

Types of repository

domain-specific trusted repositories (e.g. CESSDA archives) - focus on high-quality data with a potential for reuse

institutional research data repositories e.g. universities

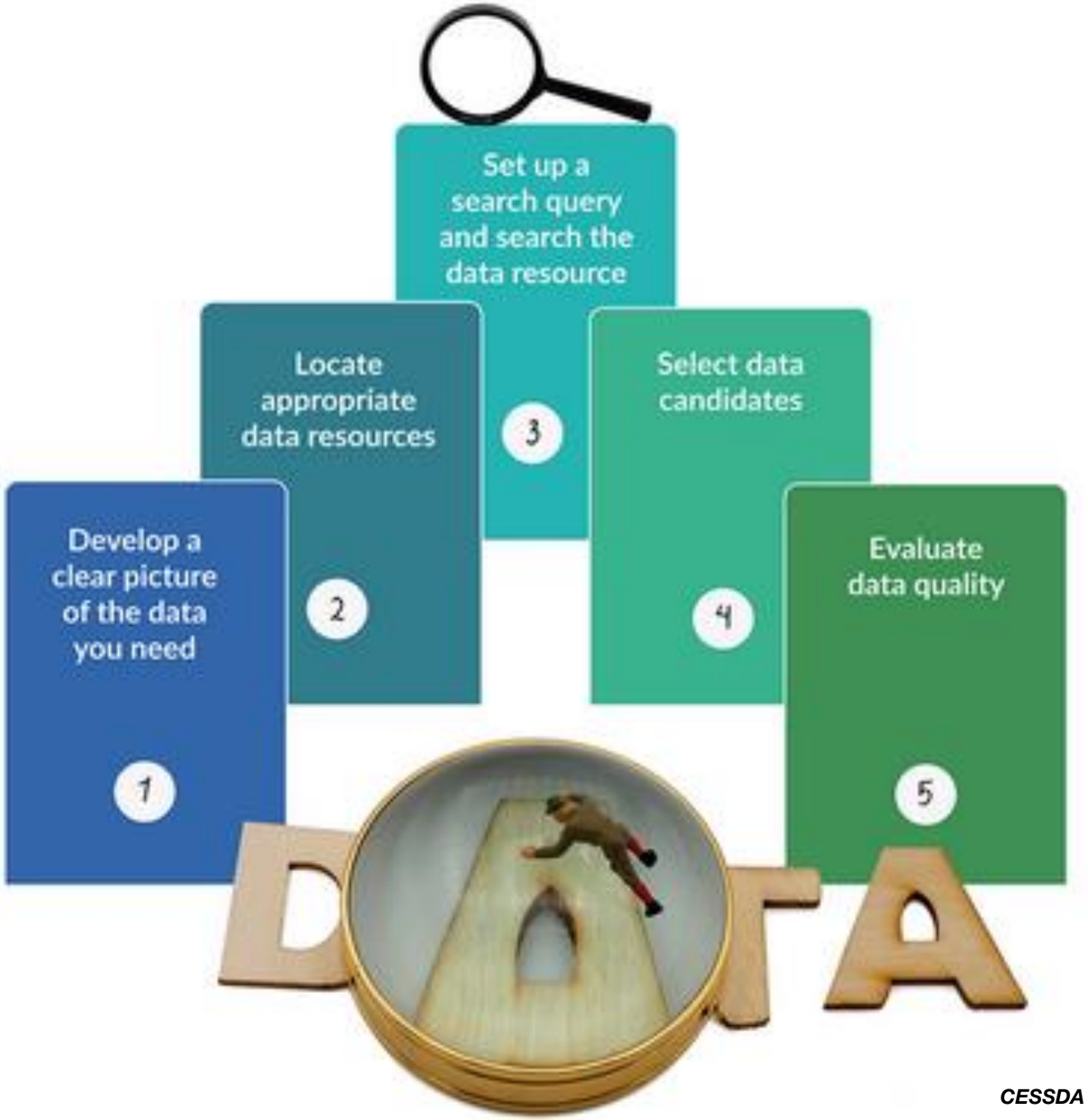
general purpose repositories e.g. Zenodo, Figshare, Harvard Dataverse

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

A registry of research data repositories

Search

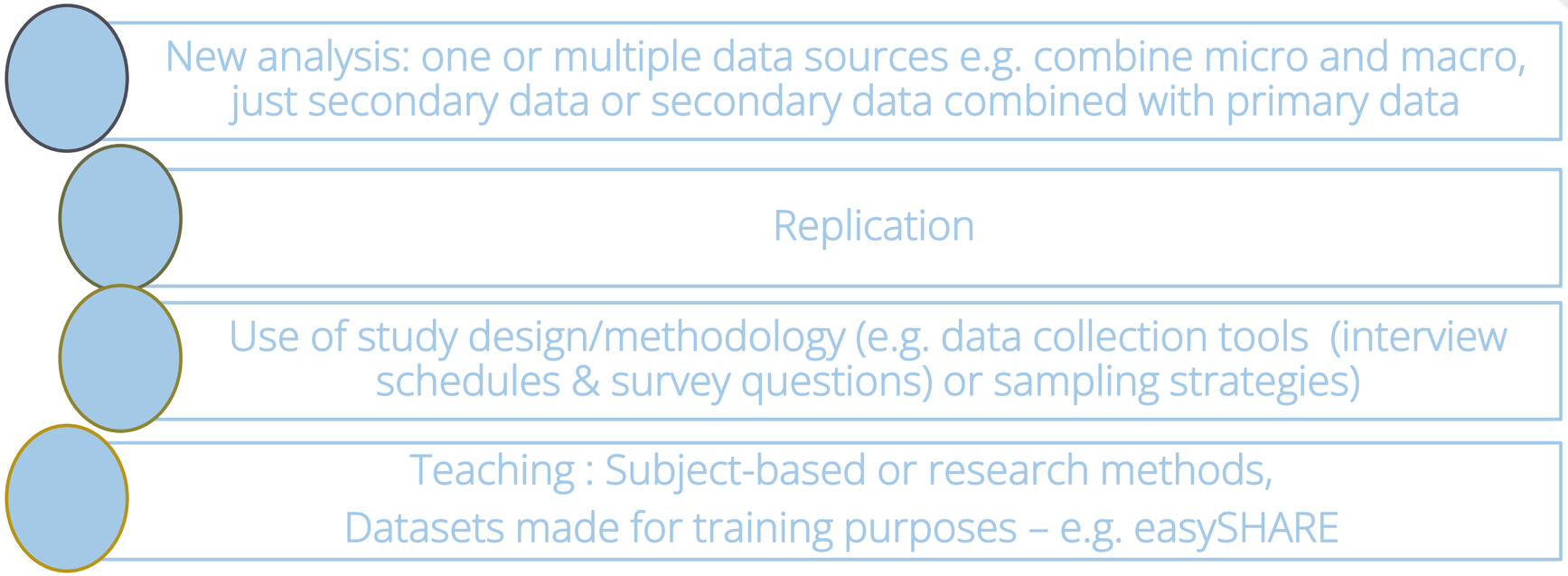
- by subject, content type and country
- for data archives with a certificate (a trusted repository), open access or for data sets that have a persistent identifier



CESSDA Training Working Group (2017)

Identify what you need

Four ways we can use archived data



Identifying data needs

Research Question

What is the ideal dataset for addressing this question?
(Compromises needed in reality)

Key concepts

- Key features
- Multidimensional
- Groups of people
- Dependent/ independent variables

How to operationalise?

(concepts can be complex and difficult to measure)

- What variables/multiple variables?
- Comparable/established measures (e.g. Schwarz Human Values)

Identifying data needs

- Who are you concerned with?
- e.g. people/adults/EU citizens, migrants, local authorities

- As most recent as possible
- a specific period (e.g. 2008-2018)
- a long a period as possible
- data from people at multiple time points?

- Do you need representative (random) sample?
- Size (large sample for inferences about small groups)

- e.g. specific countries or regions,
- all EU countries or A10 countries (2004)

- individuals, households, regions or countries?

Activity: Identify data needs

- Task: identify data needs - Evaluating data worksheet

Searching data archives

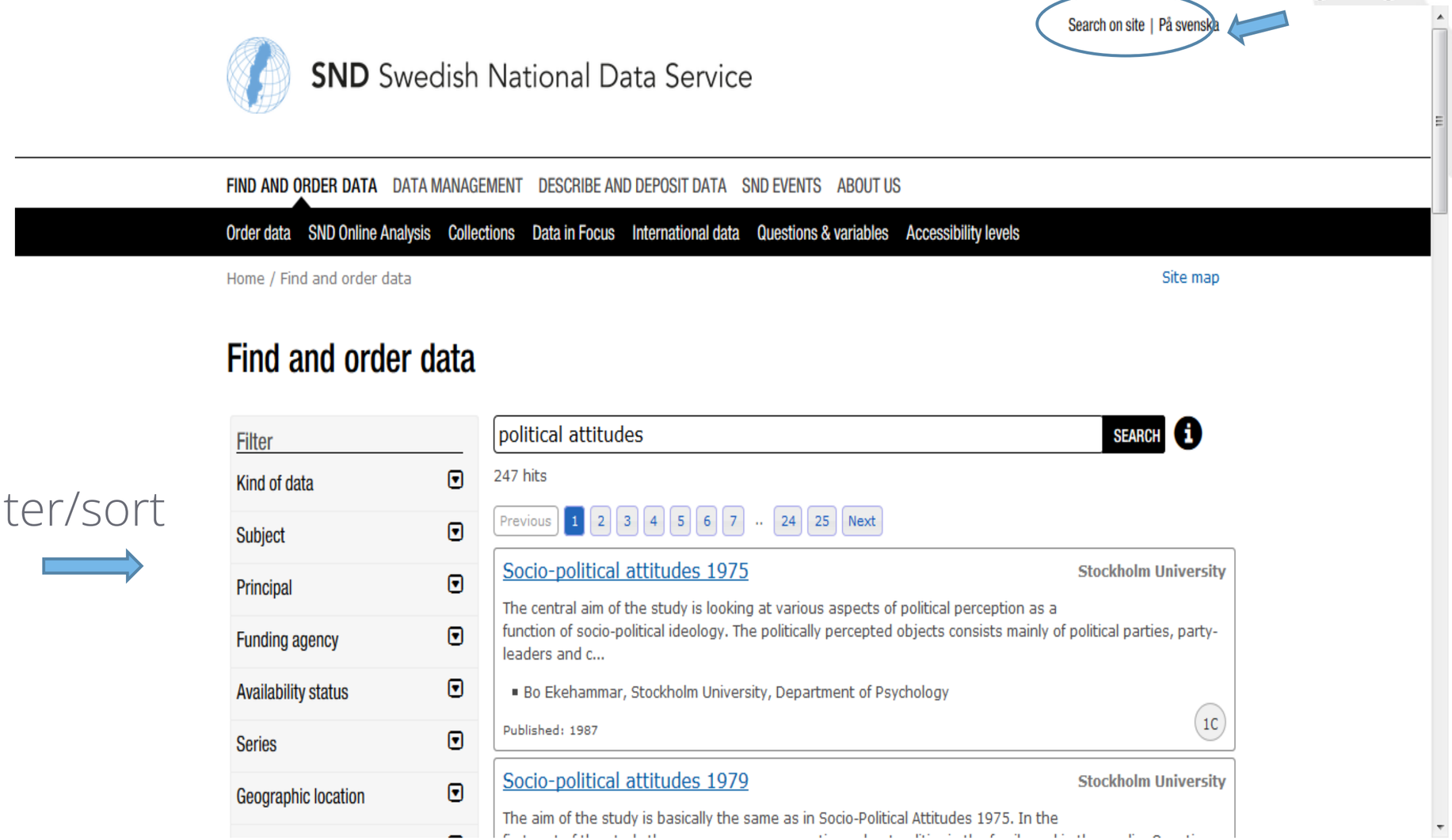
CESSDA Training Working Group (2017)

Three types of search

- Search for data on a topic
- Search for a specific dataset
- Browse data collections by type or theme

Online catalogues – searching (browsing)

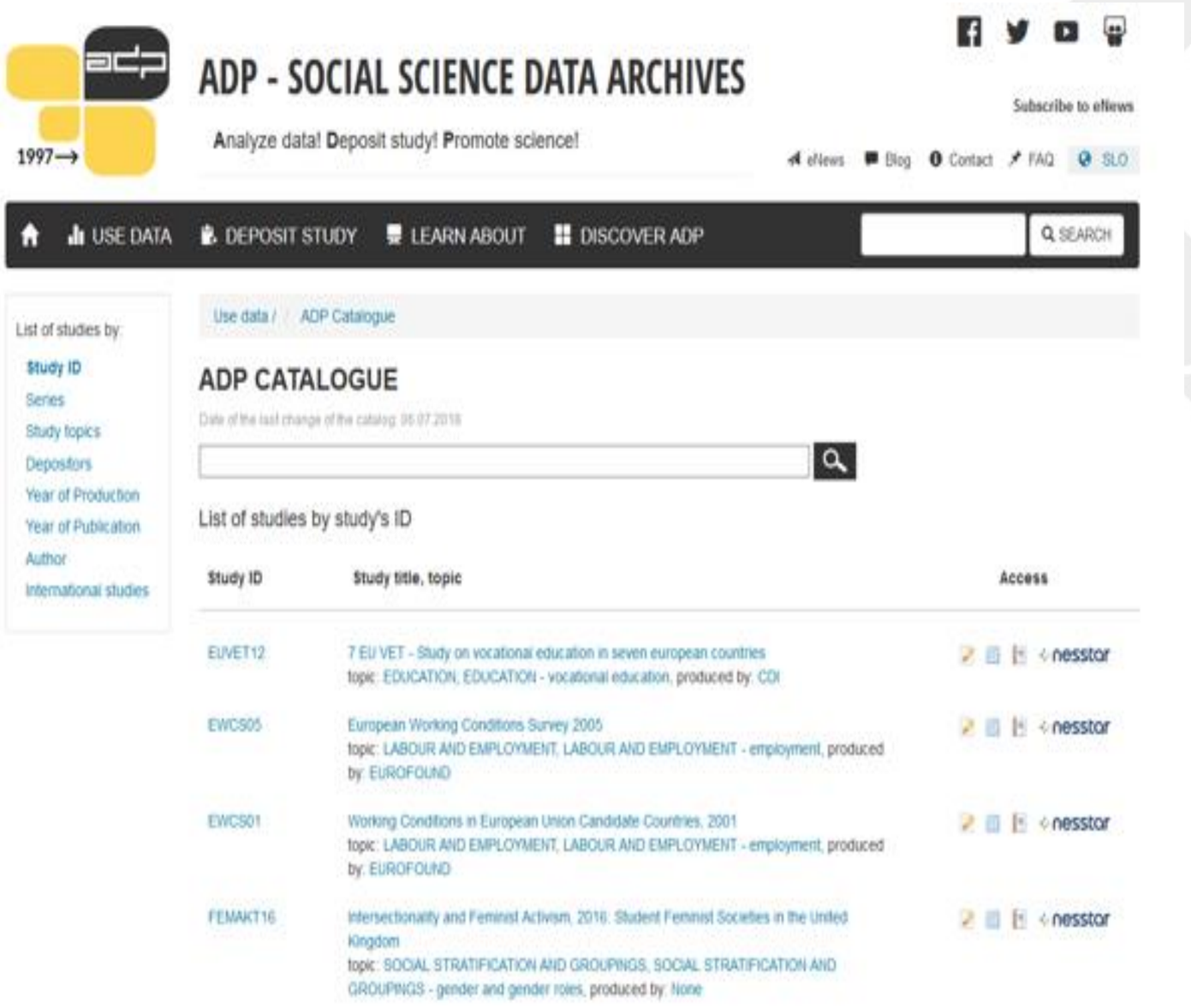
language



Filter/sort

Study Topics

- DEMOGRAPHY, POPULATION, VITAL STATISTICS AND CENSUSES
- ECONOMICS
- EDUCATION
- HEALTH
- HOUSING AND LAND USE PLANNING
- INFORMATION AND COMMUNICATION
- LABOUR AND EMPLOYMENT
- LAW, CRIME AND LEGAL SYSTEMS
- NATURAL ENVIRONMENT
- Not categorized
- OTHER
- POLITICS
- PSYCHOLOGY
- SCIENCE AND TECHNOLOGY
- SOCIAL STRATIFICATION AND GROUPINGS
- SOCIAL WELFARE POLICY AND SYSTEMS
- SOCIETY AND CULTURE
- TRADE, INDUSTRY AND MARKETS
- TRANSPORT, TRAVEL AND MOBILITY



Economic Morality (ESS2 2004)

Data/Variables - Round 2 (2004)

- Citizens should spend some free time helping others
- Society better off if everyone looked after themselves
- Citizens should not cheat on taxes
- Trust plumber/builder/mechanic/other repairer deal honestly with you
- Trust financial companies/bank/insurers deal honestly with you
- Trust public officials deal honestly with you
- Plumber/builder/mechanic/repairer overcharged you, how often last 5 years
- You were sold food packed to conceal worse bits, how often last 5 years
- Bank/insurance company failed to offer best deal, how often last 5 years
- You were sold things second-hand that proved faulty, how often last 5 years
- Public official asked favour/bribe for service, how often last 5 years
- How worried are you of being treated dishonestly
- Someone paying cash without receipt to avoid VAT or tax, how wrong

Source: ESS

Documentation often extensive

ESS Rounds

- ESS Round 8 (2016)
- ESS Round 7 (2014)
- ESS Round 6 (2012)
- ESS Round 5 (2010)
- ESS Round 4 (2008)
- ESS Round 3 (2006)
- ESS Round 2 (2004)
- ESS Round 1 (2002)

ESS8 - 2016 Data Download

The ESS8-2016 Edition 1.1 was released on 9th of April 2018. Please see Version notes for complete information.

Users are obliged to read the ESS conditions of use. Please see Deviations in data for an overview of errors or deviations in different countries.

ESS8 - integrated file
Edition 1.1

Integrated files and documents

- ESS8 - integrated file, edition 1.1
- ESS8 - data from Interviewer's questionnaire, edition 1.0
- ESS8 - test data (MTMM), edition 1.0
- ESS8 - data from Contact forms, edition 2.0
- ESS8 - data from Media claims, edition 1.0
- Guide to weighting of ESS data
- FAQ: Combining data files, Renaming variables, Other data formats.
- Fieldwork Summary and Deviations

Survey Documentation

- ESS8 Data Documentation Report ed. 1.0
- ESS8 Appendix A1 Education ed. 1.0
- ESS8 Appendix A2 Income ed. 1.0
- ESS8 Appendix A3 Political Parties ed. 1.0
- ESS8 Appendix A4 Legal Marital and Relationship Status ed. 1.0
- ESS8 Appendix A5 Population Statistics ed. 1.0
- ESS8 Appendix A6 Classifications and Coding Standards ed. 1.0

ESS Findings

Findings from the European Social Survey are available in a number of publications. >>

ESS Data Alerts

- ESS8 Swedish data removed - 09/04/18
- ESS7 Error in INWMME and INWYYE for Netherlands - 12/02/18
- ESS8 Error in INWMMS and INWYYS for Netherlands - 24/01/18
- ESS8 Second edition (2.0) of Contact form data - 15/12/17

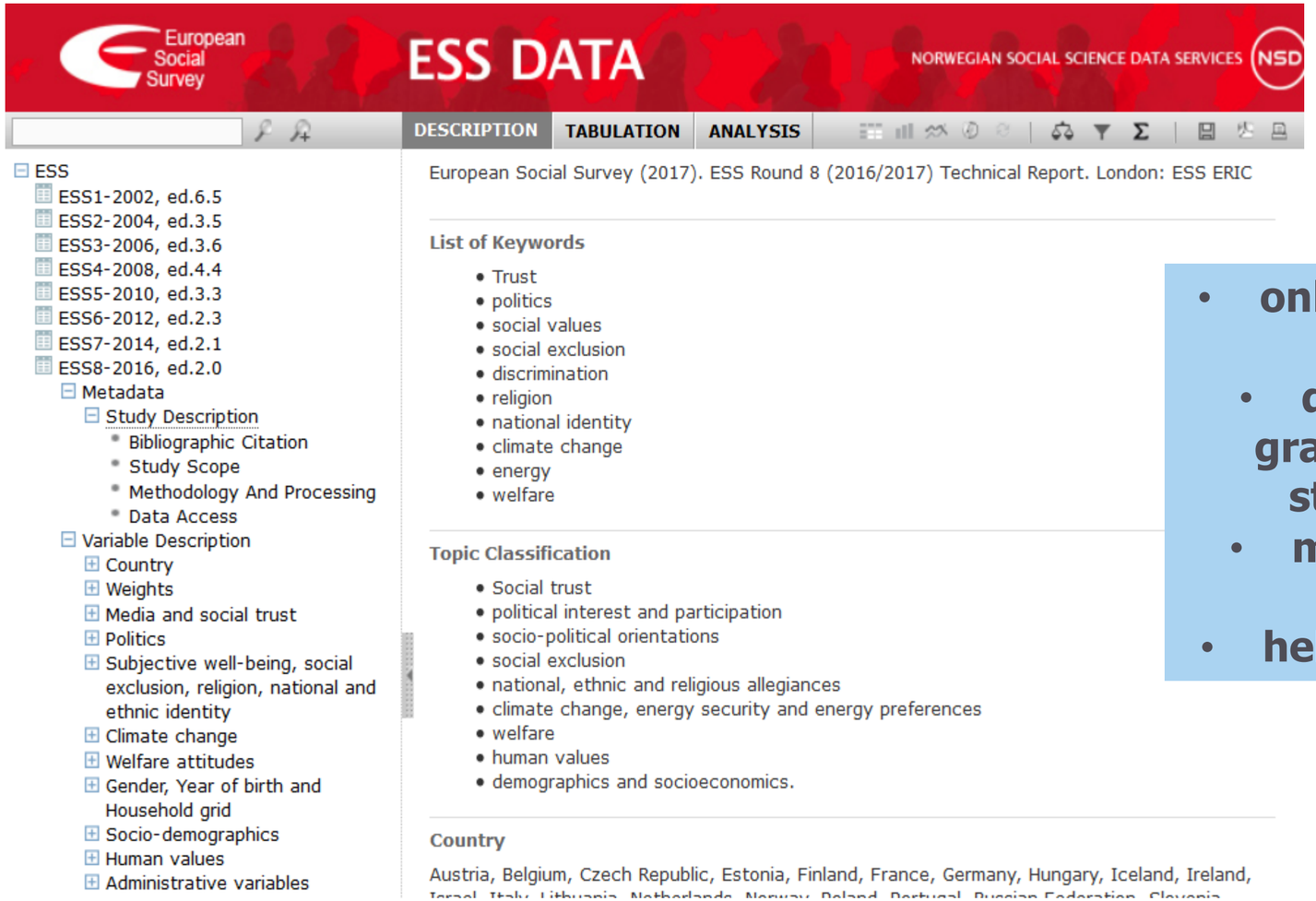
Questions?

Questions regarding data or documentation, please contact essdatasupport@nsd.no

Integrated File – Download

- Download ESS Round 8 (2016)
- Download ESS Round 7 (2014)
- Download ESS Round 6 (2012)
- Download ESS Round 5 (2010)
- Download ESS Round 4 (2008)
- Download ESS Round 3 (2006)

NESSTAR for online browsing and analysis



- **online data browsing and analysis**
- **download tables, graphs, data files and study descriptions**
- **main catalogue or additional tool**
- **help pages [? at top]**

Frequency distribution of one variable

Variable mnactic: Main activity, last 7 days. All respondents. Post coded

LITERAL QUESTION

F17c2. POST CODE: MAIN ACTIVITY

Values	Categories
1	Paid work
2	Education
3	Unemployed, looking for job
4	Unemployed, not looking for job
5	Permanently sick or disabled
6	Retired
7	Community or military service
8	Housework, looking after children, ot
9	Other
66	Not applicable
77	Refusal
88	Don't know
99	No answer

SUMMARY STATISTICS

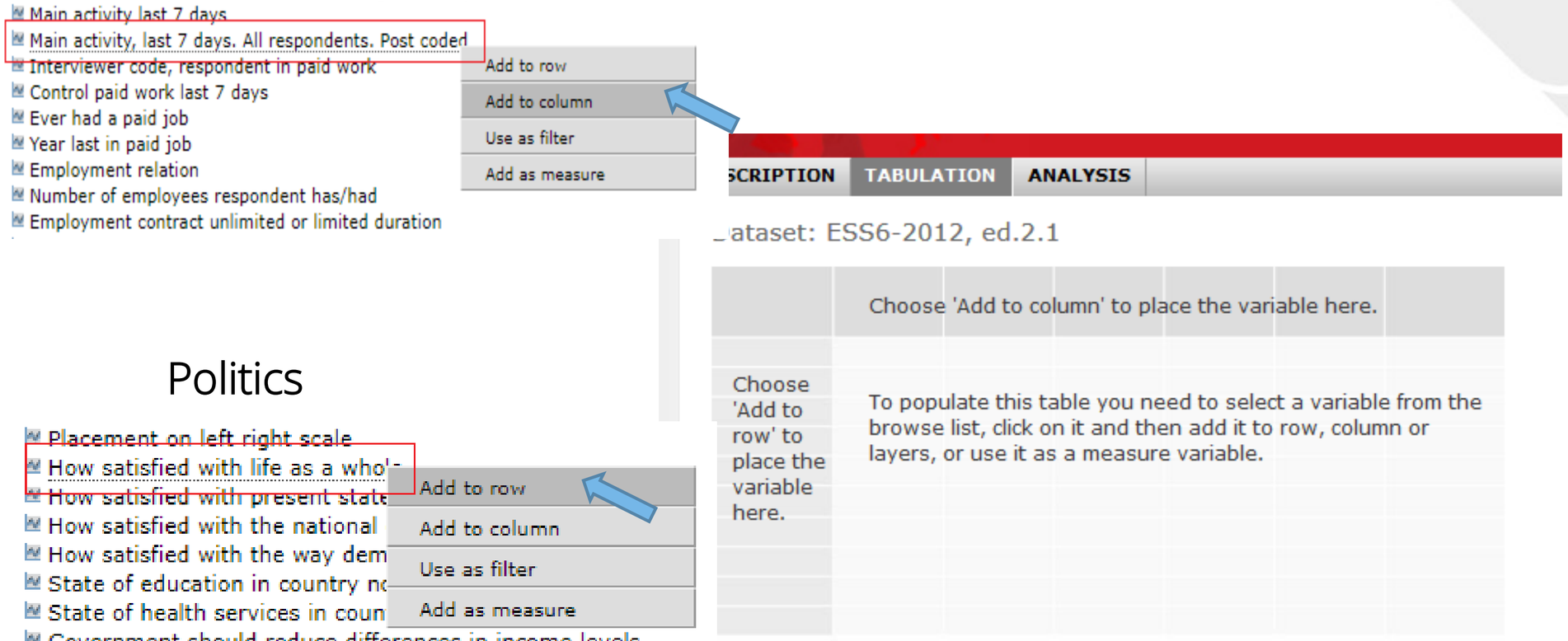
This variable is numeric

Main activity, last 7 days. All respondents. Post coded	Code	Frequency	% of all	% of valid
Paid work	1	25,760	47.1	47.6
Education	2	4,704	8.6	8.7
Unemployed, looking for job	3	2,978	5.4	5.5
Unemployed, not looking for job	4	1,155	2.1	2.1
Permanently sick or disabled	5	1,291	2.4	2.4
Retired	6	12,819	23.4	23.7
Community or military service	7	90	0.2	0.2
Housework, looking after children, others	8	4,796	8.8	8.9
Other	9	520	1.0	1.0
Refusal	77	39	0.1	-
Don't know	88	105	0.2	-
No answer	99	416	0.8	-
Total		54,673	100.0	100.0

Source: ESS; Dataset: ESS6-2012, ed.2.4

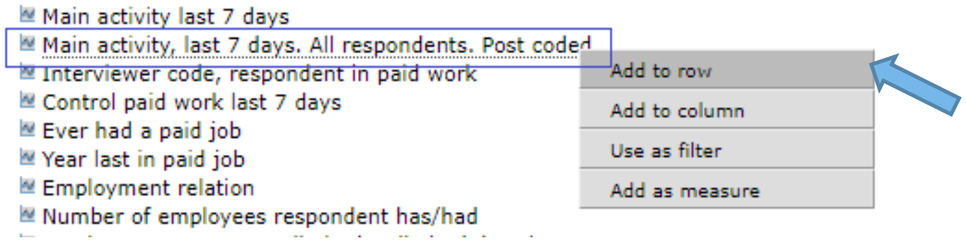
Crosstabs – frequency distribution of two variables

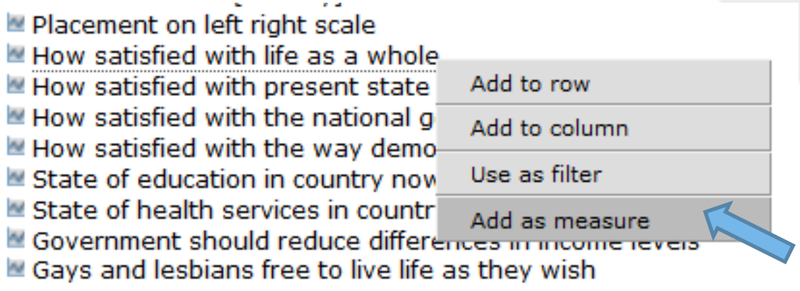
Socio-demographics



Politics

Comparing **life satisfaction measures** of two groups **– the unemployed people looking for work** versus **people in paid work**. Calculate the means for the two groups across Europe.





B20. All things considered, how satisfied are you with your life as a whole nowadays?

Main activity, last 7 days. All respondents. Post coded	Median	Average	Minimum	Maximum	Standard deviation	Sum	Count
Paid work	7.5	7.0	0.0	10.0	2.2	179,423.0	25,653
Education	8.0	7.7	0.0	10.0	1.9	35,885.0	4,676
Unemployed, looking for job	5.6	5.4	0.0	10.0	2.7	16,028.0	2,956
Unemployed, not looking for job	5.7	5.6	0.0	10.0	2.7	6,347.0	1,139
Permanently sick or disabled	5.6	5.5	0.0	10.0	2.8	7,037.0	1,278
Retired	7.0	6.5	0.0	10.0	2.5	83,123.0	12,725
Community or military service	7.6	6.9	0.0	10.0	2.6	618.0	90
Housework, looking after children, others	7.2	6.6	0.0	10.0	2.6	31,469.0	4,767
Other	7.6	7.0	0.0	10.0	2.4	3,599.0	517
Total	7.30	6.76	0.00	10.00	2.40	363,529.00	53,801

Source: ESS; Dataset: ESS6-2012, ed.2.4

Finding data

nesstar_gesis

ADVANCED SEARCH

Please click on ⊞ in the left f

Advanced Search - Google Chrome

https://zacat.gesis.org/webview/velocity?mode=searchview&searchtype=advanced&&

Advanced Search

Search criteria:

Variable ▾ contains health + -

Search for: datasets variables

SEARCH

Study Description ▾

- Study Description
  - Bibliographic Citation
    - Title Statement
      - Full Title
      - Identification Number
    - Responsibility Statement
      - Authoring Entity
    - Production Statement
      - Producer
      - Copyright
      - Funding Agency/Sponsor
    - Distributor Statement
      - Data Distributor
        - Abbreviation
      - Depositor
  - Study Scope
    - Subject Information
      - List of Keywords
      - Topic Classification
    - Abstract

EVS 2008: Greece
EVS 2008: Integrated Dataset
EVS 2008: Estonia
EVS 2008: Portugal
EVS 2008: Lithuania
describe your state of health these days (Q4)
[Open in context]
do you belong to: voluntary health organisations (Q5aN)
[Open in context]
how much confidence in: health care system (Q63M)
[Open in context]
do you work unpaid for: voluntary health organisations (Q5bN)
[Open in context]
kind of job respondent - ISCO88 code (Q112)
[Open in context]
kind of job spouse/partner - ISCO88 code (Q118)
[Open in context]
kind of job father/mother - ISCO88 code (Q129a)
[Open in context]
EVS 2008: Turkey
EVS 2008: Azerbaijan
International Social Survey Programme: Social Inequality

ZA4768: EVS 2008: Lithuania

ZA4768 Datafiles and Documentation download (via data catalogue)

Variable v9: describe your state of health these days (Q4)

LITERAL QUESTION

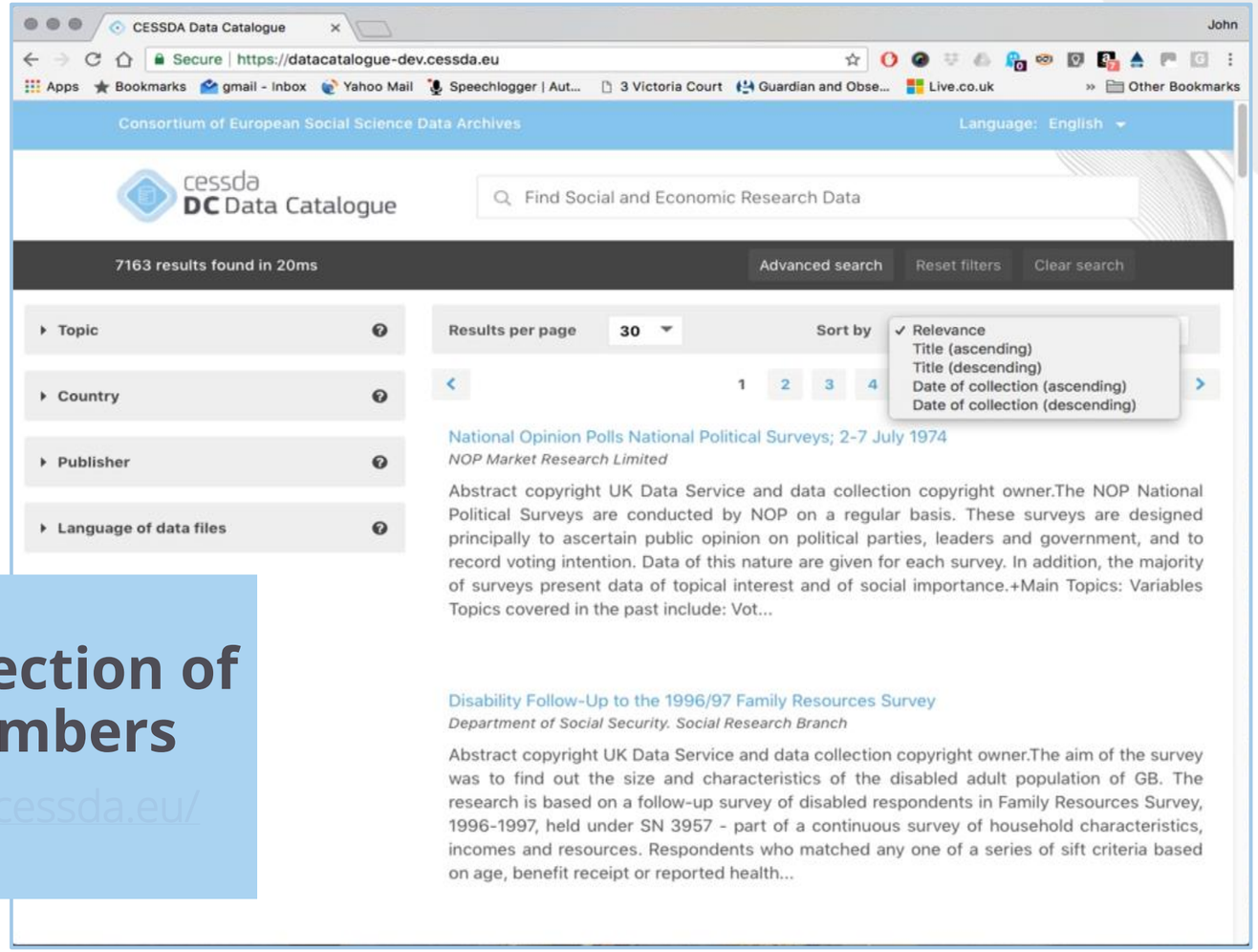
Q4

<SHOW CARD 4>

All in all, how would you describe your state of health these days? Would you say it is

-5 other missing
-4 question not asked
-3 nap
-2 na
-1 dk
1 very good
2 good
3 fair
4 poor
5 very poor

CESSDA data catalogue



Search data collection of all CESSDA members

https://datacatalogue.cessda.eu/

Finding data in practice

Searching can be hard

- Too many results
- No results
- Results not relevant

Evaluate search terms

- How well do they relate to your data needs?
- Spelling/language
- "Exact terms", Boolean Logic (AND OR) – check how search tool works

Sort, filter, advance search

ELSST

Multilingual thesaurus of social science concepts.

Hierarchical and non-hierarchical relationships between concepts.

Use to:

- broaden or narrow a search
- find terms used to index data in other languages

In future, ELSST will be used more widely to index data & embedded within search tool.

elsst.ukdataservice.ac.uk

Activity: Searching for data

Task

Search for data using a data catalogue

- **Any national data service**
- **See CESSDA for links:** www.cessda.eu/Consortium

Evaluating data: quality and usefulness

Metadata and documentation

Metadata ("data about data")

- descriptors that facilitate cataloguing data and data discovery.

Documentation

- user guides, survey questionnaires, interview schedules and fieldwork notes

- Catalogue records (with links to documentation)
- Quality can vary
- Efforts to improve data documentation
- Check for helpdesks/training

What to look for when assessing quality?

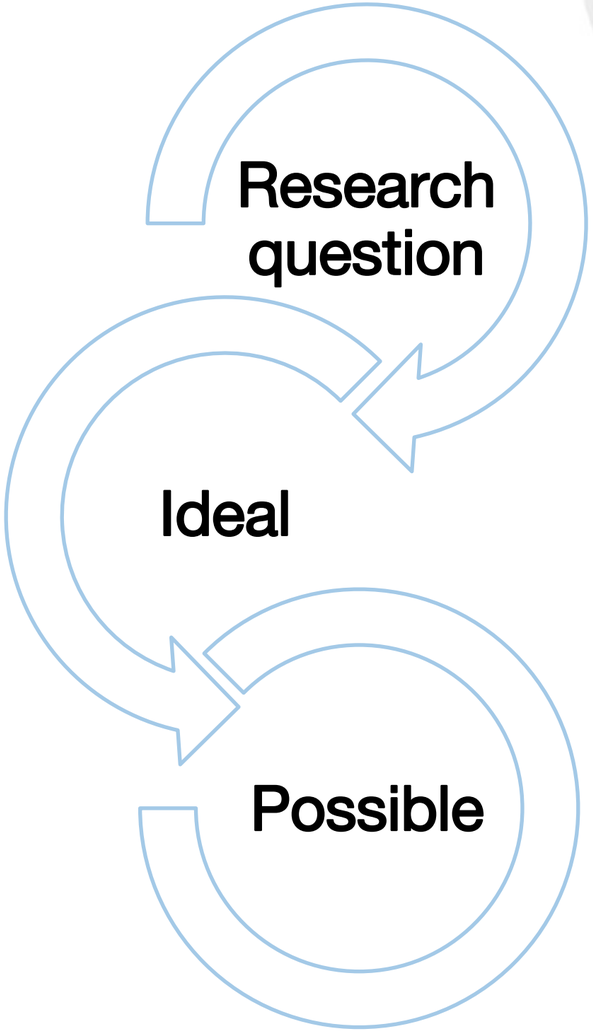
Metadata ("data about data"):

- Why the data was created?
- What the dataset contains?
- How data was collected?
- Who collected the data and when?
- How was the data processed?
- Any manipulations done to the data?
- What quality assurance procedures were used?

But is it useful?

Compare:

- key concepts
- population
- geographical area
- time period
- units of analysis
- study/sample design



Accessing data

Now finally, I've found some great data, how to I get it?

- **Licenses**
- **Access process**
- **Getting started**

CC BY SA

Data access arrangements 1

Open data

any user, no registering (acknowledge source)

Registration

- **often with institutional user name and password**
- **may wait for user name or password**
- **register use of data**

Terms and conditions

- **not trying to identify individuals, households or organisations**
- **not distributing data to others**
- **“data is for non-commercial use only” or for “use in research or teaching” only.**

Download

from catalogue (but sometimes complete a request form)

Images by CESSDA Training Working Group (2017)

Data access arrangements 2

- Sometimes permission from the data owners required (= a additional stage)
- Sensitive or confidential data = more strict (and lengthy) process
- Some services operate a dedicated safe room or safe access service
- Access by users outside the country can be prohibited for confidential data
- Free (except for commercial use and supplementary services)

If you are unsure, ask the relevant data service for help.

And finally…remember to cite data

Why?

- It gives credit the data creators
- It makes data easier to find

How?

- Give enough information to locate the exact version of the data
- Look for recommended citation
- Use persistent identifiers (Digital Object Identifier - DOI)

CESSDA Training Working Group (2017)

ESS Round 2 (2004)
ESS Round 1 (2002)

Citation requirements

The Core Scientific Team of the ESS requests that references to ESS data and the Data Documentation Reports should use the form of words listed below.

To ensure that such source attributions are captured for social science bibliographic utilities, citations must appear in the footnotes or in the reference section of publications.

Citation of data

- ESS Round 8: European Social Survey Round 8 Data (2016). Data file edition 1.0. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 7: European Social Survey Round 7 Data (2014). Data file edition 2.1. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 6: European Social Survey Round 6 Data (2012). Data file edition 2.3. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 5: European Social Survey Round 5 Data (2010). Data file edition 3.3. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 4: European Social Survey Round 4 Data (2008). Data file edition 4.4. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 3: European Social Survey Round 3 Data (2006). Data file edition 3.6. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 2: European Social Survey Round 2 Data (2004). Data file edition 3.5. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.
- ESS Round 1: European Social Survey Round 1 Data (2002). Data file edition 6.5. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

Citation of documentation

- ESS Round 8: European Social Survey (2017): ESS-8 2016 Documentation Report. Edition 1.0. Bergen, European Social Survey Data Archive, NSD - Norwegian Centre for Research Data for ESS ERIC.

are available in a number of publications.
>>

ESS Data Alerts

ESS8 Swedish data removed - 09/04/18

ESS7 Error in INWMME and INWYYE for Netherlands - 12/02/18

ESS8 Error in INWMMS and INWYYS for Netherlands - 24/01/18

ESS8 Second edition (2.0) of Contact form data - 15/12/17

Questions?

Questions regarding data or documentation, please contact essdatasupport@nsd.no

Integrated File – Download

Download ESS Round 8 (2016)

Download ESS Round 7 (2014)

Download ESS Round 6 (2012)

Download ESS Round 5 (2010)

Download ESS Round 4 (2008)

Download ESS Round 3 (2006)

Download ESS Round 2 (2004)

Download ESS Round 1 (2002)

Online Analysis

Open ESS Round 8 (2016)

Open ESS Round 7 (2014)

Open ESS Round 6 (2012)

Open ESS Round 5 (2010)

Open ESS Round 4 (2008)

ELEMENTS OF DATA CITATION

- **Author:** Name(s) of each individual or organizational entity responsible for the creation of the dataset.
- **Date of Publication:** Year the dataset was published or disseminated.
- **Title:** Complete title of the dataset, including the edition or version number, if applicable.
- **Publisher and/or Distributor:** Organizational entity that makes the dataset available by archiving, producing, publishing, and/or distributing the dataset.
- **Electronic Location or Identifier:** Web address or unique, persistent, global identifier used to locate the dataset (such as a DOI). Append the date retrieved if the title and locator are not specific to the exact instance of the data you used.

These are the minimum elements required for dataset identification and retrieval. Fewer or additional elements may be requested by author guidelines or style manuals. Be sure to include as many elements as needed to precisely identify the dataset you have used.

Source: IASSIST – Quick guide to Data Citation

ISSP Research Group (2017): International Social Survey Programme: Work Orientations IV - ISSP 2015. GESIS Data Archive, Cologne. ZA6770 Data file Version 2.1.0, doi:10.4232/1.12848

Hafner-Fink, M. and Malešič, M. (2016). Slovenian Public Opinion 2015: Work Orientation (ISSP 2015), Role of Government (ISSP 2016), Mirror of public opinion and National Security Survey [Data file]. Ljubljana: University of Ljubljana, Social Science Data Archives. ADP – IDNO: SJM15. https://doi.org/10.17898/ADP_SJM15_V1

Resources for social media data

- Social media data come from various resources, such as Facebook, Twitter, Reddit, Instagram or YouTube.
  - The elements of social media data may be:
  - individual tweets, comments on Facebook, Twitter or Reddit etc.,
  - visual content, such as photos or videos,
  - network connections between network users (friend connections, groups),
  - data on ratings and/or interests (preferences or likes)

  - The availability to researchers is limited

Availability of social media data

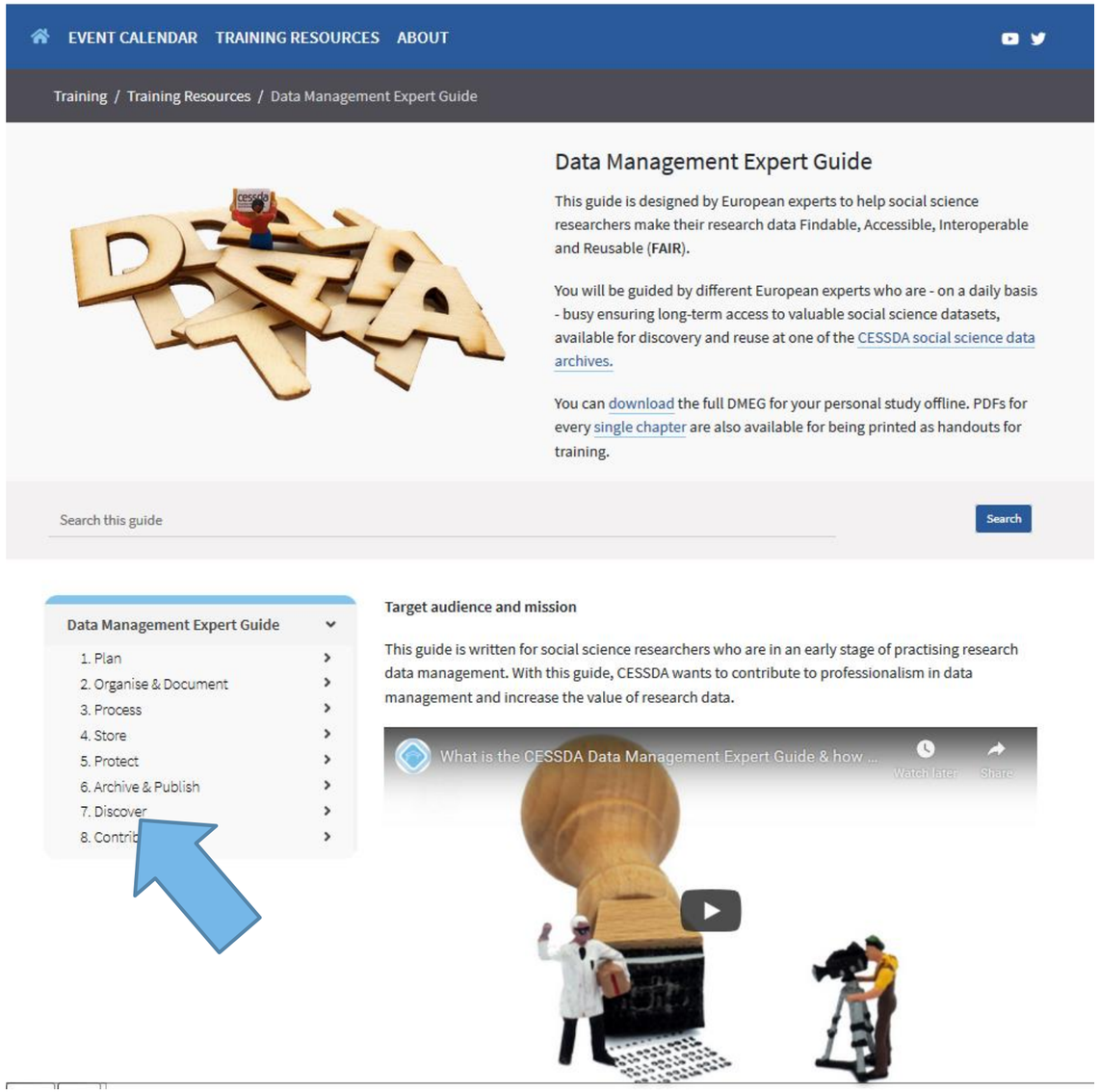
- Social media data are available to researchers, but their availability is restricted by companies that own respective social media platforms (Facebook , Twitter, etc.). Restricted availability of social media data represents serious obstacle for more intensive application of social media data in social research.
- There are several reasons for their limited:
  - *Legal reason*: it deals with the social media content's copyright. The users have copyright for their own content (e. g. Tweets or Facebook posts) and by signing terms of use they give the social media platform a license to use the content for various purposes. The use of the social media data for third parties (private companies, academic researchers etc.) is restricted in the terms of use. This constrains the researchers (and data archives) in using, storing and sharing the data. A good source of guidance on social media data preservation both for researchers and repositories is Thomson, S.D. (2016) "Preserving Social Media".
  - *Ethical reason*: researchers and data archivist must care about the protection of personal information of the social media users.

Platforms as social media data sources

- Social media data can be obtained through the application programming interfaces (APIs) of the social media platforms. However, these APIs usually restrict the type and amount of data you can collect. If researchers request large amounts of data through APIs, they might not get the complete data but samples. Often it is not fully transparent how these data are sampled.
- For those who are not able to handle APIs for downloading the data, there are commercial subjects that sell social media data, such as Gnip (acquired by Twitter Inc. in 2014) or DataSift, but these usually have high costs.

Social media data in European Data Archives

- As on July 2019, only two CESSDA archives store and disseminate social media data so far: GESIS and UK Data Service (UKDS) offer their users limited collection of social media data, Facebook data, geo-coded Twitter data, and specific subsets of Wikipedia. In particular, UKDS holds several Twitter data sets (20 collections of Twitter communication (tweets' IDs, timestamp, hashtags).
- Currently, several CESSDA archives plan strategies to overcome legal and technical issues related to social media data archiving and sharing as they see it as important area.



www.cessda.eu/DMEG

cessda

More literature

– Quick reference guide: Using administrative data for research
– Quick reference guide: Social media and research
– Guidelines on the use of social media data in survey research

- Data Management Expert Guide (CESSDA)
- Offline version of DMEG

Questions

Irena Vipavc Brvar (ADP - Slovenian Social Science Data Archives)
Jennifer Buckley (UKDS – United Kingdom Data Service)

Pictures from
CESSDA Training Working Group (2017). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC.

cessda.eu

@CESSDA_Data

CC BY SA