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

Overview

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

cessda

Provide overview

- This session is about introducing types of data available and how to find data for a research project
- Five sections
- Lectures/talking but also practical activities where the aim is to find data for a specific research question
- Assumes no prior knowledge but we recognize you may come with relevant experience and knowledge, therefore - opportunity to share and learn from each other

2

Data types and sources

cessda

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.

Activity 1: Your knowledge and experience of the data landscape (a warmer activity)

Activity 1 aims are to get participants talking, identify some key data sources/types, and assess participants prior knowledge

1. Participants write some ideas individually (give about 1 min)
2. Then share with neighbour (give 2-3 minutes)
3. Get feedback from each group (possibly making a list on flip chart etc.) and ask participants to introduce each other

Types of data

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

il. text documents, laboratory notebooks and diaries, questionnaires, scripts and codebooks, audiotapes, videotapes, photographs and films, examination results, specimens, samples, artefacts, slides, database schemas, database contents, models, algorithms and methodologies, workflows, standard operating procedures and protocols, experimental results, metadata, other literature review records, archives
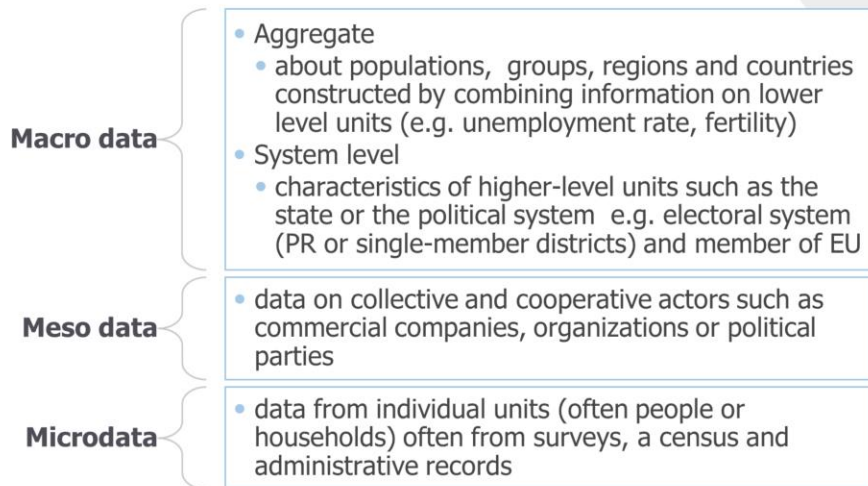
Quantitative and Qualitative

Thinking about the types of data available can help you work out what you need and how to find it. It can help you understand what is possible and help you navigate the volume of data you can find.

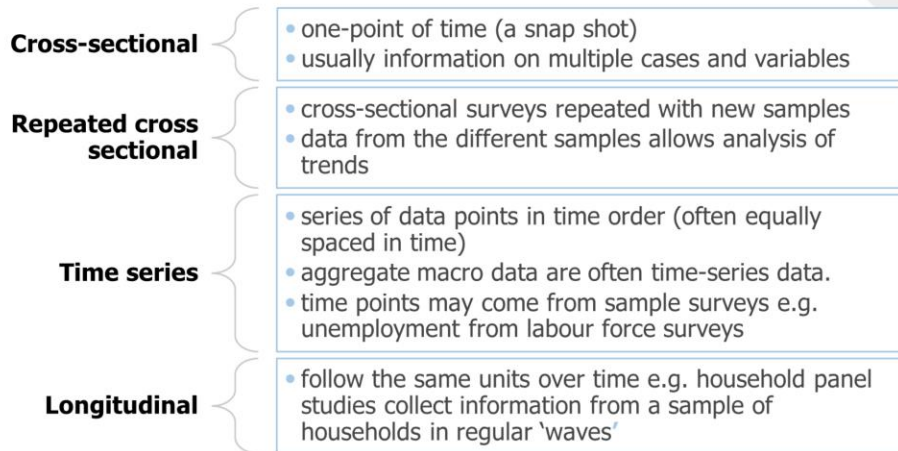We will now discuss some ways to classify data.

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

We can distinguish between types of data at different levels of analysis:

- Microdata - data from individual units (often people or households). Common sources of microdata include surveys, a census and administrative records
- Aggregate macro data refers to data about populations, groups, regions and countries constructed by combining information on lower-level units. Examples of aggregate data include demographics, rate of unemployment for countries (or regions) and GDP. The information on lower-level units may come from surveys, a census and administrative records. System-level macro data refers to information about characteristics of higher-level units such as the state or the political system that is not based on summaries of lower-level units. Specific examples include 'Electoral system (PR or single-member districts)'
- Meso data: a term used to describe data on collective and cooperative actors such as commercial companies, organizations or political parties

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'

cessda

We can differentiate data by how time is represented:

- Cross-sectional – data from one-point of time for multiple cases and variables (a snap-shot)
- Repeated cross-sectional – when cross-sectional surveys, which collect data from a sample in a single point of time, are repeated with new samples we get repeated cross-sectional data. Comparing data from the different samples then allows analysis of trends
- A time series is a series of data points in time order, often at equally spaced points in time. Aggregate macro data for countries and regions are often time-series data. The individual time points may come from sample surveys. For example, the standard measure of unemployment in many European countries comes from countries Labour Force Survey.
- Longitudinal studies – follow the same units over time. Panel studies collect information from the panel members in intervals known as 'waves', which tend to be frequent. [example of household panel studies]
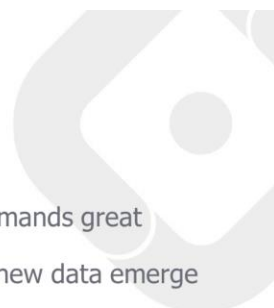
New data types

- 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.

◇ cessda

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).

cessda

Sources of microdata

There are many sources of data.

cessda

Now we will see where you can find 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

YHTEISKUNTATIETEELLINEN
TIETOARKISTO
FINLANDS
SAMHÄLLSVETENSKAPLIGA
DATAARKIV
FINNISH SOCIAL
SCIENCE DATA ARCHIVE

UK Data Service

DANS

SND Swedish National Data Service

NSD NORWEGIAN CENTRE FOR RESEARCH DATA

gesis Leibniz Institute for the Social Sciences

ČSDA Czech Social Science Data Archive Institute of Sociology

cessda

European social science data archives provide access to extensive collections of data.

- Each archive's collection is unique.
- Archives vary in size. Large archives may provide both national and international data, and small archives can have more limited national data collections.
- Quantitative data is most common, particularly data from social surveys. Making social science archives a key source of individual-level data
- Some archives also include qualitative data, for example, interview transcripts and field notes.
- Sources of data include major collaborative studies, official statistics and government studies to small research projects and individual researchers.
- Collections can include current data (though note the processing of data means data might not be archived for many months after data collection). They can also include older data collections including studies that have been collecting data over decades.
- There is a general trend towards greater data archiving and sharing due

to interest in enhancing public investments in research and in making research more transparent.

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.

◉ cessda

Some background
This workshop has been developed by CESSDA.

CESSDA is the Consortium of European Social Science Data Archives.

Its vision is to provide a full-scale sustainable research infrastructure that enables the research community to conduct high-quality research, which in turn leads to effective solutions to the major challenges facing society today.

Key tasks underlying this vision include
- developing standards and best practices around the management and archiving of social science data
- facilitating researcher access to important resources

Ways this work is being developed include e-services such as the CESSDA data catalogue, training and coordinating the network of European data service providers.

Members

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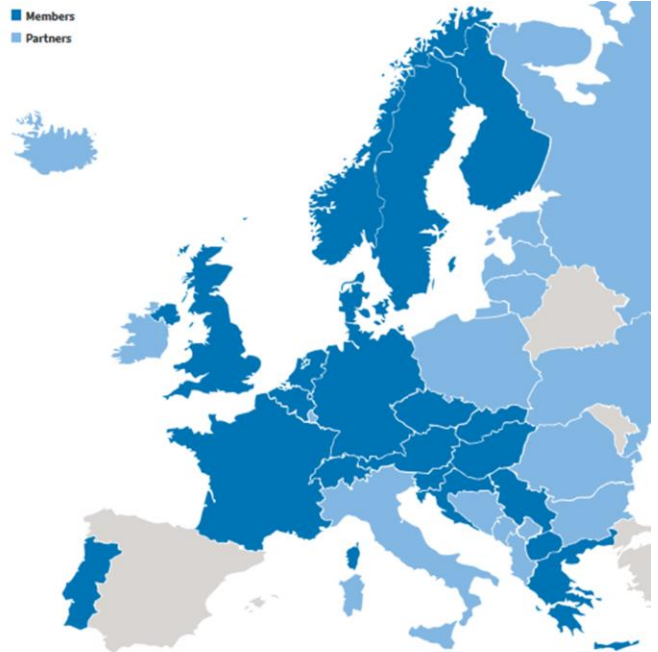
cessda

Many of the data services in Europe are CESSDA members. This map highlights the different national data services in Europe. The darker blue signals member countries. Partner countries are in light blue.

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.

But what do national data services do?
National data services typically combine the archiving of research data with activities to make data available for research, teaching and learning purposes. Core activities include checking data and metadata, maintaining catalogues and managing access to data. Many seek data for their collection and develop training for those creating data and for users of 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

⬦ cessda

In reality, access to results from scientific research is often not completely open. But, we have movement in the rights decoration.

HOW IT WORKS

Get funding / Write DMP / Gather data during research / Finish research

Publish results

Choose repository and deposit data

Inform OpenAIRE the EU-funded Open Access portal

#openaccess
#opendata
#H2020

reach more people, have greater impact

avoid duplication of efforts

preserve data for future researchers

2015
of 431 signed projects
65.4%
opted to share data

Source: EC

cessda
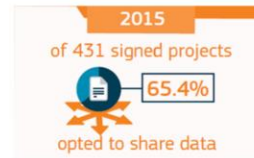
Open Access to research data

Importance of research infrastructures / data repositories

RESEARCH DATA – OPEN BY DEFAULT

FAIR DATA!

Accessible

Findable

Interoperable

Re-usable

Source: EC

cessda
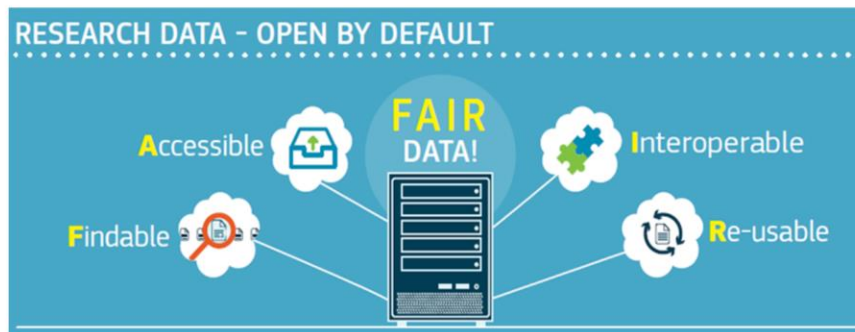
There is an ambitious proposal for a European Open Science Cloud and work to make data FAIR. Principals emphasize the function of research infrastructures especially trusted data depositories such as national data archives.

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.

cessda

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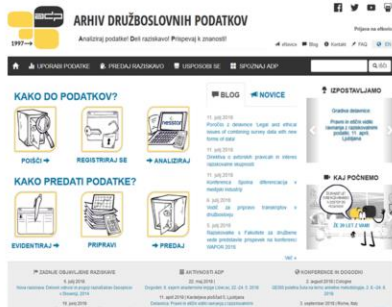
- **Oldest data sets in the archive (public opinion polls) are from 1966**
- **Wide range of topics covered**
- **In most cases data relates only to Slovenia / few international**
- **Metadata in SI and EN, datafiles mostly in SI**

cessda

Now let me tell you something about Slovenian social science data archives:

- It was founded in 1997
- We serve as Slovenian national data repository for social sciences
- Currently we have around 600 surveys with data in data catalogue and 150 with metadata (description of survey)
- We had around 800 users registered in 2017 and 168 survey data used for detailed secondary analysis
- Oldest data sets are from 1966 (public opinion survey)
- Wide range of topics covered
- In most cases data relates only to Slovenia and few international
- Metadata is in SI an EN, data files mostly in SI

UK Data Service

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

Support for users with training and guidance.
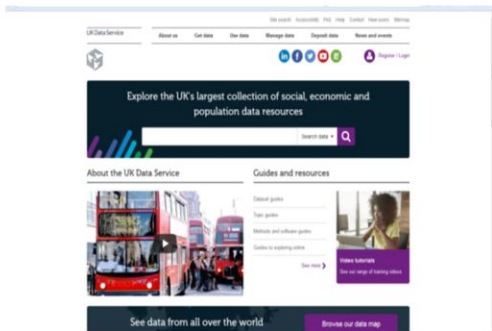
- major UK and cross-national surveys
- longitudinal studies (household panel and cohort studies)
- UK Census 1971-2011
- qualitative data collections
- research data in a researcher repository (Reshare)

UK data service

cessda

The UK Data Service provides access to the UK's largest collection of social, economic, and population data.

We also support users with training and guidance.

The data collection includes major UK and cross-national surveys including many government sponsored surveys.

Longitudinal studies including a large household panel survey called Understanding society and several cohort studies following individuals born in 1958, 1970 and 2000.

There is data from the UK Census from 1971 to 2011. There are also qualitative data collections containing in-depth interview transcripts, diaries, anthropological field notes, answers to open-ended survey questions, audio-visual recordings and images.

The UK Data Service also has an online repository called Reshare for researchers to archive, publish and share research data. It help support the requirement for researchers with grants from the UK's Economic and Social Research Council to archive the data from their research. With data deposited in Reshare, the UK Data Service ensures they conform with ethical and legal requirements.

Cross-national studies

International survey research programmes include many European countries

- International Social Survey Programme (ISSP)
- European Social Survey
- European Values Survey
- Eurobarometer
- Survey of Health, Ageing and Retirement Europe (SHARE)
- Generations and gender programme (GGP)
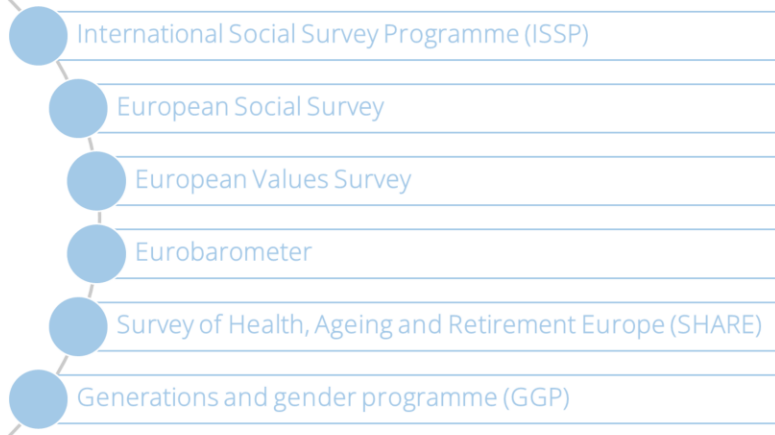
Many European countries participate in international survey research programmes. Current programmes cover a range of themes/topics. Here are some examples…

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

ISSP

Data Download

Data Catalogue

gesis Leibniz-Institut
für Sozialwissenschaften

cessda

And we can discuss some in more depth...

The International Social Survey Programme (ISSP) is an annual programme of cross-national collaboration on surveys. For example, the 2015 data covers 37 countries. The collaborative model means survey questions for each modules are agreed by an expert group and the questions are included in existing national survey programmes.

It consists of thematic modules on a range of topics including the Role of Government Family and Changing Gender Roles, Environment and Health and Health Care. These modules are then rotated (Examples on slide)

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

European Social Survey

Then we have European social Survey which is biennial cross-national survey.

Survey has the highest methodological standards.

Data are freely available for 36 countries, participation of countries vary from year to year.

The ESS data are probably the most used and cited data in Europe.
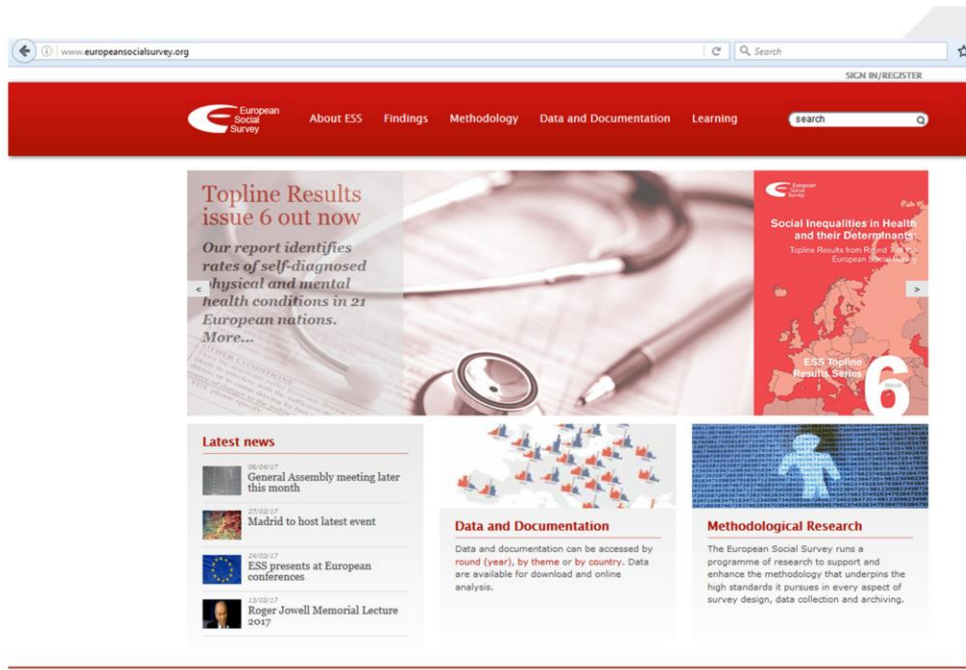
These resources contain valuable information and resources such as survey documentation and publications on substantive findings and

methodological discussions.

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
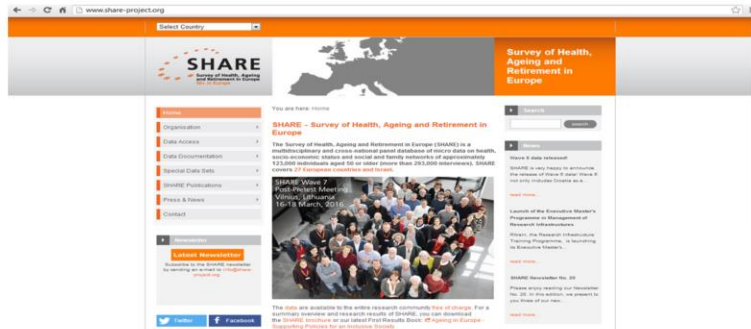
SHARE, the Survey of Health, Ageing and Retirement in Europe
SHARE is a longitudinal study collecting data from more than 123,000 individuals aged 50 or older across 27 European countries and Israel. SHARE provides micro data on health, socio-economic status and social and family networks and is a major resource for research into ageing.

Accessing data from cross-national studies
Many cross-national studies have dedicated infrastructures and websites for distributing data such as SHARE (on slide)
Share data can be accessed at http://www.share-project.org/data-access.html

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
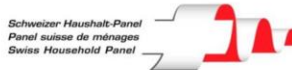
A final source of data to highlight are major longitudinal studies in Europe. Longitudinal studies are a precious resource for social and economic research; as tracking the same individuals over time enables the analysis of change at the individual level. There are a number of longitudinal studies in Europe. These household panel studies offer key examples:

- The German Socio-Economic Panel (SOEP) is a study of nearly 11,000 households and started in 1984.
- Understanding society is a very large study of following the lives of 40,000 UK households. It started in 2009 but also incorporated participants in an older panel study that lasted for 25 years.
- Swiss household panel study started in 1999 with a sample of just over 5,000.

Access to SOEP via the German Institute for Economic Research. Understanding Society and Swiss Household panel study via the national social science data service within each country.
Some harmonisation between these studies

At European level, an important longitudinal study is EU-SILC, coordinated

by EUROSTAT - individual-level changes are measured over a four-year period.

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

A wide range of organisation provide access to key sources of data.
Here are 5 important examples, especially of macro/aggregate time series data.
Eurostat is an important one to highlight for data about European countries …

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

eurostat
Your key to European statistics

cessda

Eurostat is the statistical office of the European Union. Its key task is to provide statistics at European level that enable comparisons between countries and regions (national and sub-national data across a wide range of themes).

Eurostat also provides access to several major sources of microdata including the European Union Statistics on Income and Living Conditions (EU-SILC), collects individual level data on income, poverty, and living conditions and includes a longitudinal element.

The EU-LFS gives harmonised data at on employment and related topics.

These sources of microdata are used to derive many agreement level statistics but researchers can apply to access the microdata itself.

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

Another source of official statistical data beside Eurostat, that include data from various countries is MISSY an online service platform

Includes metadata at the study and variable level as well as reports and tools for data handling and analysis

Missy currently documents the

following official statistics microdata

Source: MISSY

If we look at EU – SILC (European Union Statistics on Income and Living Conditions) study you can see that you can search for metadata according to you needs e.g. year, country

Factsheet: Accreditation & Data Access Conditions for SLOVENIA

Series and studies for SLOVENIA

- Census
- Central Population Register (CPR)
- Community Innovation Survey (CIS)
- European Union Statistics on Income and Living Conditions (EU-SILC)
- External Trade
- Final accounts
- Graduates in tertiary education
- Gross Fixed Capital Formation
- Household Budget Survey (HBS)
- Industrial production
- Labour Force Survey (LFS)
- Personal Income Tax
- Slovenian Business Register
- Statistical Register of Employment
- Student enrolment in tertiaty education
- Time Use Survey (TUS)
- Unemployed workers
- Usage of information-communication technologies in enterprises
- Victimisation Survey (VS)

Contact

Name

Statistični urad Republike Slovenije (Statistical Office of the Republic of Slovenia, SORS)

Postal Address

Litostrojska cesta 54 SI-1000 Ljubljana (Slovenia)

Contact

info.stat@gov.si

Website

[click here]

Conditions

General Conditions

Microdata can be provided for research purposes to registered research institutions and registered researchers in both academia and government offices. "Registered" researchers have a national identifier.
As a general rule, users submit applications for access that are assessed by a Confidentiality committee (internal to SORS); the latter prepares recommendations for SORS's board of directors, which makes the final decision.

Conditions for Non-Resident Researchers

Same as national researchers (though requests from non-registered researchers should be evaluated by the internal Confidentiality committee).

Source: CIMES

cessda

Centralising and integrating metadata from European Statistics

Provides a large and comprehensive **overview** of official micro data disseminated for research purposes by the national statistical institutes (NSI) across Europe.

Series and studies for SLOVENIA ☓

- Census
- Central Population Register (CPR)
- Community Innovation Survey (CIS)
- European Union Statistics on Income and Living Conditions (EU-SILC)
- External Trade
- Final accounts
- Graduates in tertiary education
- Gross Fixed Capital Formation
- Household Budget Survey (HBS)
- Industrial production
- Labour Force Survey (LFS)
  - Labour Force Survey - 1995
  - Labour Force Survey - 1996
  - Labour Force Survey - 1997
  - Labour Force Survey - 1998
  - Labour Force Survey - 1999
  - Labour Force Survey - 2000
  - Labour Force Survey - 2001
  - Labour Force Survey - 2002
  - Labour Force Survey - 2003
  - Labour Force Survey - 2004
  - Labour Force Survey - 2005
  - Labour Force Survey - 2006
  - Labour Force Survey - 2007
  - Labour Force Survey - 2008
  - Labour Force Survey - 2009

Study

Labour Force Survey - 2011

Original Title :
Anketa o delovni sili - 2011

Original Alternative Title :
ADS - 2011

English Alternative Title:
LFS - 2011

Producer : SORS.

Abstract :
Slovenian Labour Force Survey 2011 is a Slovenian research with a tradition. The LFS measures the labour status and other characteristics of the population in a certain week of each quarter, by spreading the sample uniformly over all the weeks of the quarter. The survey provides data on size, structure and characteristics of active and inactive population. Data on personal income are added to the dataset (DURS register) – an average monthly income in either the whole year or a shorter period of time, if a person had worked for less than a year. Approximately 19.750 individuals are selected to the sample in each quarter. Non-anonymised version of LFS microdata is available to researchers, onsite or by remote access. The survey was conducted as one of the surveys of the Eurostat Labour Force Survey which includes data from 27 Member States of the European Union, four Candidate Countries and two EFTA countries (Norway and Switzerland). Comparability through time and space is possible as Eurostat distributes Labour Force Survey data of other participating countries.

Keywords :
LABOUR AND EMPLOYMENT, SOCIAL STRATIFICATION AND GROUPINGS

Geographic coverage : Slovenia.

Universe :
The target population is the jure population, which includes persons that are mainly living in the territory of Republic of Slovenia, regardless of their nationality. Only the population, is living in individual households, is covered by the survey. Institutionalized people are not considered as a part of the jure population. Those who live in institutions (soldiers, hospitalized, imprisoned etc.) for more than 12 months, including students who don't live at home and Slovenian citizens who live abroad permanently or temporarily, are not covered by the survey.

Sampling Procedure :
The Labour Force Survey is based on the sample taken from the Central Population Register. It is a rotating panel carried out continuously throughout the whole year. The sampling method is stratified systematic random sampling of addresses. All members of the household at the selected address are included in the sample. That means that there are approximately 16 150 individuals included in each
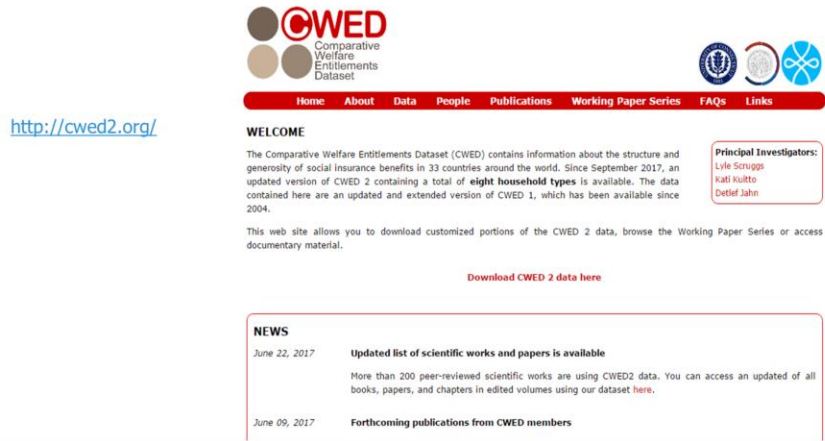
CIMES
Centralising and Integrating Metadata from European Statistics

Source: CIMES

cessda

Direct from project websites

Some research projects share research data through project websites

http://cwed2.org/

Researchers might also share their data via their dedicated websites.

One example is the Comparative Welfare Entitlements Dataset (CWED), which contains information about the structure and generosity of social insurance benefits in 33 countries around the world.

The data contained here are an updated and extended version of CWED 1, which has been available since 2004.

This web site allows you to download customized portions of the CWED 2 data, browse the Working Paper Series or access documentary material. 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

Data can often be accessed through a data repository, which is a digital archive collecting, preserving and displaying datasets, related documentation and metadata (OpenAIRE, 2017). There different types of data repository.

- The CESSDA archives are examples of domain-specific trusted repositories. As a general rule, this kind of repositories includes only high-quality data with a potential for reuse.
- Institutional research data repositories
- General purpose repositories such as Zenodo (n.d.), Figshare (n.d.) or Harvard Dataverse (2017).

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

On this website you can find a registry of research data repositories

Re3data.org is a registry of over 1500 research data repositories,.

You can search by subject, content type and country. In addition, you can select whether you want to search for data archives with a certificate (a trusted repository), with data sets that are available via open access or for data sets that have a persistent identifier.
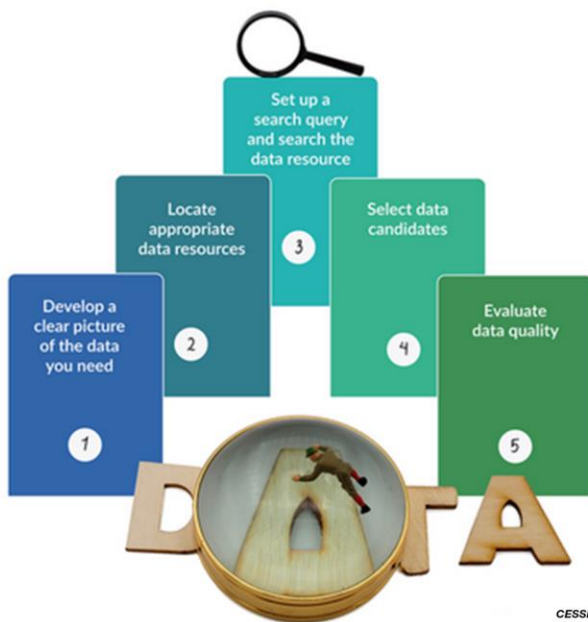
CESSDA Training Working Group (2017)

Identify what you need

Four ways we can use archived data

New analysis: one or multiple data sources e.g. combine micro and macro, just secondary data or secondary data combined with primary data

Replication

Use of study design/methodology (e.g. data collection tools (interview schedules & survey questions) or sampling strategies)

Teaching : Subject-based or research methods,
Datasets made for training purposes – e.g. easySHARE

Archived data can be used in multiple ways.

- One common use is new data analysis - New data analysis can involve one or multiple data sources. For example, you might want to combine data from different levels of analysis such as microdata from a survey (such as the ESS) with regional statistics (from Eurostat). You can also use secondary data only or combine secondary with new primary data.
- Archived data can also be used to replicate the analysis of a previous study.
- Researchers can also make use of archived data's design/methodology including data collection tools and sampling strategies. For example, those designing a survey can find questions from the mass of archived survey projects, which includes surveys designed by experts. Those designing qualitative work can also make use of the detailed notes to replicate a study.
    - EXAMPLE - The [Living and working on Sheppey](#) project offers an interesting example. The project explored the changes in working lives over 40 years on the Isle of Sheppy (Kent, UK) by combining archived and newly collected data. In the new project, researchers replicated a part of the study where school pupils

were asked to complete essay where they imagined themselves in old age reflecting back on their lives.

- Archived data is a great resource for teaching, giving learner experience of real-life research. Data can enhance subject based learning and is key for research methods teaching. Many data providers offer datasets specifically designed for learners. For instance, easyShare is a longitudinal data set created especially for training purposes. It contains only selected variables merged into a single data file. It is more user-friendly than the complete set of SHARE panel data.

In the rest of this session, we are mainly thinking about the first use.

Identifying data needs

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

Key concepts	How to operationalise? (concepts can be complex and difficult to measure)

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

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

cessda

A big part of finding data for a research project is identifying what data you need.

➢ Your data needs largely depend on the research question/objectives. We can think of this process as imagining the ideal dataset to answer your research question. In reality you might need to make compromises but an understanding of the ideal gives you some way to evaluate the potential sources you find.

➢ Key concepts - A first step can be to specify the main concepts of interest.
- Consider the key features of the concepts you want to find within data e.g. if interested in political behaviour, are interested in voting or a broader concept of political behaviour
- For multidimensional concepts, are you interested in any specific theoretical dimensions – e.g. online or offline political behaviour (and possibly how they relate to each other)
- Are you also interested in identifying specific groups/types of people such as men and women?
- Depending on the type of research question/approach it might be useful to specify the dependent and independent variables.

For instance, if looking at gender differences in political participation online, your dependent variable is some measure(s) of online participation and gender is one independent variable – what other independent variables do you need/want to include?

➢ Next, you can consider how to operationalise these concepts (at least ideally)
- Your concepts may be complex and difficult to measure (they often are in the social sciences) but thinking about the key features you want can help find the most suitable data. For example, for operationalising religiosity, we can consider how people self identify as belonging to a religious group or alternatively information about the practice of religious activities. For social class, we can have individuals 'subjective' social class or use occupation based classifications. Think about what is most important for you.
- In quantitative research, you may need information from different variables. To measure complex concepts. For instance, rather than asking about household income directly, surveys might use multiple questions asking about different forms of income and the information can then be combined to create a more accurate measure of overall household income.
- You may want to use measures that are comparable to other studies (or that expand on other studies) In the social sciences, many concepts are measured in standard ways.
  - E.g. The 'Human Values Scale' a well-established 21-item measure of human values, developed by the Israeli psychologist, Professor Shalom Schwartz. It is designed to classify respondents according to their basic value orientations. The Human Values Scale has been included in every ESS round to date.
  - attitudes towards the welfare state are often measured in surveys with questions asking whether it is the government's responsibility to provide certain forms of welfare, often labelled as "role of government".
- There are also important official or harmonised approaches to measuring some concepts, which are designed to provide comparable and validated measures. For instance, all members of the EU must use the International Labour Office (ILO) standardised measure of unemployment in their Labour Force Surveys.
- ILO also developed the International Standard of Occupational Classification

(ISCO) to harmonise the measurement of occupations. Using ISCO codes, the European Socio-Economic Classification (ESeC) was developed in the early 2000s as a comparable measure of socio-economic status (SES).

- Like all measures, standardised approaches have strengths and weaknesses but you will generally find literature examining their validity and reliability.

Identifying data needs

Population	Geography	Time	Unit of analysis	Study design and sample

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

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

- **Time**
  - 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?

- **Unit of analysis**
  - individuals, households, regions or countries?

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

cessda

Ideal Population
What population does your research focus on e.g. European citizens, school children, migrants to the EU, local authorities?

Geographical coverage
Does the population have a specific geographical coverage? e.g., a specific country, countries that are EU member or the "A10" countries joining in 2004?

Time
What period do you want your data to cover? E.g. as recent as possible, from a specific period? For as long a period as possible?
Do you need data from the same sample at different time points?

Units of analysis
Are you interested in individuals, households, regions or countries?

Sample type
A key question in relation to sample type is whether you need data to be representative of a specific population. If so, you most likely need to find data collected from a sample taken using random sampling techniques. Large – especially important for making estimates about small sub-populations – e.g. political activists.

Activity: Identify data needs

- Task: identify data needs - Evaluating data worksheet

Use handout with description of the research question/hypothesis and questions.

ALSO – there is a version of the handout with suggestions/answers.

Searching data archives

CESSDA Training Working Group (2017)

cessda

Now I will tell you how to search data in data archives

Three types of search

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

There are three common main types of searches

1. Search for data on a specific topic
2. Search for a specific dataset (that you know of from other studies)
3. Browse data collections by type or theme (allowed by some archives)

We will focus on the first (which is often the most common)

Online catalogues – searching (browsing)

Most national data services (and other data providers) have online catalogues for searching or browsing
Example – Swedish National Data Service

To get data relevant for you, you can use filters according to what you are looking for, such as: kind of data, subject, who was principal investigator, and geographical coverage, etc.

- Language - Websites and catalogues are often in multiple languages (national languages and English). However, data and documentation may only be available in one language.

Example: ADP

Searching also thru filters such as study topics, author if you know, if the data are part of series, and etc.

Economic Morality (ESS2 2004)
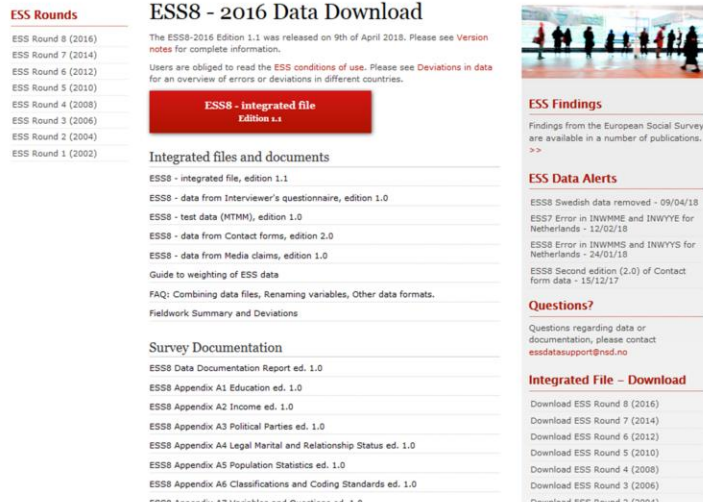
Data/Variables - Round 2 (2004)

- ⓘ 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

European
Social
Survey

Source: ESS

cessda

Documentation often extensive

Documentation of published data is often extensive (obsežna), and before the analysis is important to read everything availabe so you don't misinterpret the results

NESSTAR for online browsing and analysis

Many data service providers use an online platform called NESSTAR

- NESSTAR enables online data browsing and analysis.
- You can also download tables, graphs, data files and study descriptions.
- Some data services use NESSTAR as their main tool for searching and accessing data while others have a main catalogue, providing NESSTAR as a separate tool.
- If using NESSTAR, the help pages include lots of helpful user guidance. NESSTAR 'Help' can be accessed at any time by clicking at the top of the screen.
- A Youtube channel of UK Data Service, there are video tutorials on how to use NESSTAR.
- Future of NESSTAR is now (spring 2018) unclear, as development was stopped. It is possible that NESSTAR will be replaced by Dataverse system.
- [trainer notes: can include a demo to extend]

Frequency distribution of one variable

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

LITERAL QUESTION			Code	Frequency	% of all	% of valid
F17c2. POST CODE: MAIN ACTIVITY		Main activity, last 7 days. All respondents. Post coded				
Values	Categories	Paid work	1	25,760	47.1	47.6
1	Paid work	Education	2	4,704	8.6	8.7
2	Education					
3	Unemployed, looking for job	Unemployed, looking for job	3	2,978	5.4	5.5
4	Unemployed, not looking for job	Unemployed, not looking for job	4	1,155	2.1	2.1
5	Permanently sick or disabled	Permanently sick or disabled	5	1,291	2.4	2.4
6	Retired	Retired	6	12,819	23.4	23.7
7	Community or military service	Community or military service	7	90	0.2	0.2
8	Housework, looking after children, ot	Housework, looking after children, others	8	4,796	8.8	8.9
9	Other	Other	9	520	1.0	1.0
66	Not applicable					
77	Refusal	Refusal	77	39	0.1	-
88	Don't know	Don't know	88	105	0.2	-
99	No answer	No answer	99	416	0.8	-
SUMMARY STATISTICS						
This variable is numeric		Total		54,673	100.0	100.0

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

cessda

Now we will do a simple analysis on ESS data – we will look at the frequency distribution of one variable.

The data in this „educational purposes" is not weighted, but otherwise you have to use weights.
We can see two columns of %: % of all (includes also answers such as refusal, don't know, no answer) and valid % with only valid answers (the others are defined as missing values).

Crosstabs – frequency distribution of two variables

Socio-demographics

- Main activity last 7 days
- Main activity, last 7 days. All respondents. Post coded
- Interviewer code, respondent in paid work
- Control paid work last 7 days
- Ever had a paid job
- Year last in paid job
- Employment relation
- Number of employees respondent has/had
- Employment contract unlimited or limited duration

| Add to row |
| Add to column |
| Use as filter |
| Add as measure |

Politics

- Placement on left right scale
- How satisfied with life as a whol
- How satisfied with present state
- How satisfied with the national
- How satisfied with the way dem
- State of education in country no
- State of health services in coun
- Government should reduce differences in income levels

| Add to row |
| Add to column |
| Use as filter |
| Add as measure |

SCRIPTION · **TABULATION** · **ANALYSIS**

Dataset: ESS6-2012, ed.2.1

	Choose 'Add to column' to place the variable here.
Choose 'Add to row' to place the variable here.	To populate this table you need to select a variable from the browse list, click on it and then add it to row, column or layers, or use it as a measure variable.

cessda

Comparing **life satisfaction measures** of two groups — **the unemployed people looking for work** versus **people in paid work**. Calculate the <u>means</u> 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

cessda

Now we will do a descriptive analysis e.g., comparing life satisfaction of two groups - the unemployed people looking for work versus people in paid work – we will calculate means for two groups:

- In comparison of two selected groups, we can see that people in paid work have higher average of life satisfaction according to unemployed people looking for work
- the highest average have those whose main activity was education (7.7)

Finding data

In Nesstar we can search data with the help of advance search, with click o this magnifying glass with +, where the new window opens where we define our search.

In our case we will search in Gesis – German archive, so we can chose what results do we need, so from this dropbox menu we select variable (but you can also chose any other), so the Nesstar will give us all surveys containing variables that have selected word in it, in our case HEALT, also we have to tell Nesstar what to search for, in this case we chose variable as opposed to dataset. Than we click search. Now we get results   list of all surveys and than we look at one of variables and if we click on it we get frequency distribution where the word we were searching for is coloured yellow.

CESSDA data catalogue

Search data collection of all CESSDA members

https://datacatalogue.cessda.eu/
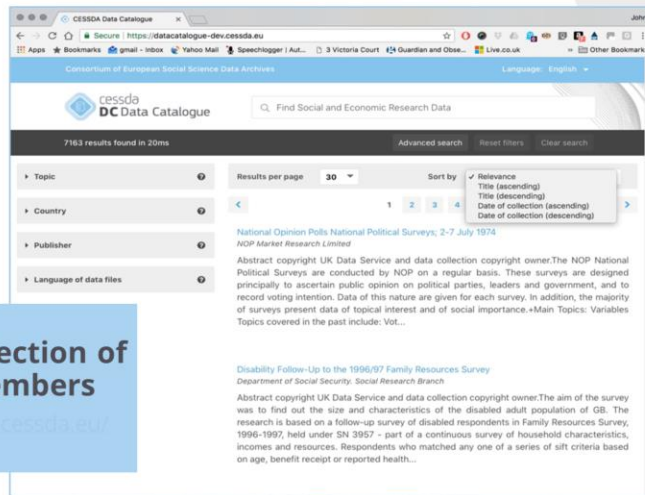
cessda

In CESSDA catalogue you can also search with the help of different filters such as topic, country, publisher, …

This new tool allows you to search across all CESSDA member archives, with features search as a free text search and ways to sort and filter results. The results will include a summary of the data and link to the relevant archive.

As a result you get list of surveys with metadata (study description) but if you want to access the data you have to click on go to study link and it will re-derect you to on page of archive containing this survey.

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
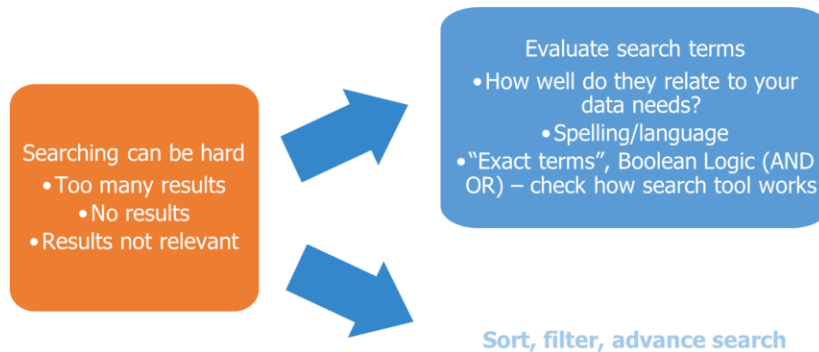
Searching archives can be demanding; the volume of data available, while useful, makes finding data for a specific purpose difficult

Alternatively, you might struggle to get results or al least results that seem relevant.

What you can do? Two general strategies.

1. Think about your search terms
How do your search terms relate to your data needs
- e.g. interested in doing research on a specific groups such as fathers many social surveys may ask about children and family life but not necessarily index the data using the term fathers
- lots of useful data for examining migration topics might come from studies not specifically focused on migration. For example, for migrant labour market outcomes, you may want to focus on identifying data on employment and then checking if you can identify migration status.

Using exact terms can reduce number of results/make them more relevant e.g.

Spelling can be an issue – American and English spelling – ageing and aging

2. Sort results and use filters – refer to your ideal data to refine by data types, date and geography
Check settings on search tool – are you searching titles, subjects, variables etc.

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.

Search for
concepts

ELSST covers the core
social science disciplines

Enter your search term and select
a language to browse ELSST

elsst.ukdataservice.ac.uk

cessda

One tool that can help is ELSST.

ELSST is a multilingual thesaurus for the social sciences. It contains concepts which aim to be language- and country-neutral, i.e. relevant to all languages and member counties.
Concepts are linked through hierarchical and non-hierarchical relations. Users can exploit these relationships to broaden or narrow their search to possibly more relevant concepts.
ELSST also supports multilingual information retrieval. When embedded into a search engine, it allows a user to query the system in one language and retrieve data indexed with the term in another language of ELSST.
The English language version contains over 3,000 concepts, including Preferred Terms that are used to index data and Non-preferred terms that are not used by indexers but lead indexers to the Preferred Terms. 90 % of the Preferred Terms used to index data have been translated into all the other languages, with several languages having 100% coverage.
It may be useful to consider alternative spellings of key search terms.

A suggested idea for ELSST demo
- Use the term NURSES, which has a Use For term, plus Broader,

Narrower and Related terms (see the online guide available by clicking 'ELSST guide' on the top navigation bar https://elsst.ukdataservice.ac.uk/elsst-guide for an explanation of these terms).

- You can view NURSES in its hierarchy, and its equivalents in other languages here:

https://elsst.ukdataservice.ac.uk/thesaurus-search/view-concept/?id=52f955ef-0871-4372-9763-43c9247c810c&lang=EN

- Alternatively, you can view it as a visual graph (click on the nodes to expand): https://elsst.ukdataservice.ac.uk/thesaurus-search/view-concept/?id=52f955ef-0871-4372-9763-43c9247c810c&lang=EN#/tab-visual-graph

- Some terms need a Scope note to explain their meaning. An example is PARAMEDICAL PERSONNEL:

https://elsst.ukdataservice.ac.uk/thesaurus-search/view-concept/?id=e39bbfe2-0491-480a-b741-0d083d82295a&lang=EN#/tab-current-version

Activity: Searching for data

Task
Search for data using a data catalogue
• **Any national data service**
• **See CESSDA for links:** www.cessda.eu/Consortium

cessda

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

⊙ cessda

Metadata and documentation are key to making sense of the data.

- Metadata or "data about data" are descriptors that facilitate cataloguing data and data discovery.
- Documentation might include user guides, survey questionnaires, interview schedules and fieldwork notes.

Archives produce catalogue records containing metadata. Documentation is often accessible directly from the catalogue record (without needing to register), usually as fully searchable PDF documents (but format and quality varies, especially for older studies).
The quality and accessibility of metadata data and documentation can vary. A core aim of CESSDA is that data collections have documentation that allows use without recourse to the data creator.
Work by CESSDA and its' service providers as well as many other research and data professionals is improving the way data is documented e.g. metadata standards such as consistent categories and researchers being trained in research data management.
Some services (where resources allow) offer support to help users understand data through helpdesks and 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?

Asking some simple questions about the data can help determine its quality.

[Trainer note: for more interaction, you can get participants to think of list – then show to confirm]

Answers to many questions on the catalogue page/website
For others, you may need to review supporting documentation such as User Guides, Technical Reports

If you cannot answer these questions easily, you need to consider if you can use this data meaningfully. If you have unanswered questions, you may be able to ask the data archive or data owners for further information.

But is it useful?

Compare:
· key concepts
· population
· geographical area
· time period
· units of analysis
· study/sample design
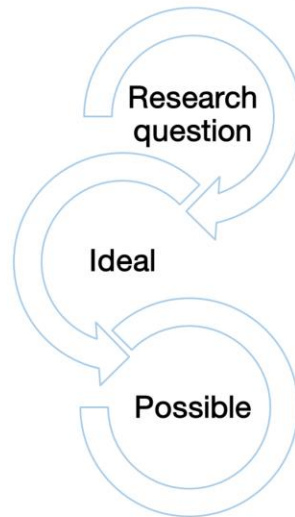
Evaluate the usefulness of data for a specific purpose using metadata and documentation – i.e. comparing a dataset to the ideal dataset identified previously.

- units of analysis
- access
- time period(s)
- geographical coverage
- include the correct characteristics/concepts of interest
- nature of the sample

You may need to make some compromises – in terms of data and research questions.

60

Accessing data

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

· **Licenses**
· **Access process**
· **Getting started**

Once you've found useful data, how do you access it and get started? In this final session, we cover

- Licenses and conditions of access (what you can and cannot do with data; different levels of access (e.g. Open data, registered users/uses))
- access processes (registering, downloading/requesting data, data formats)

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)

cessda

Open - Some services support access to open data collections that any type of user can download without registering (but they may need to acknowledge the source).

Registration - you usually need to register with the data service before accessing data.

- University researchers may be able to register using their institutional user name and password.
- You usually need to provide an e-mail address (and possibly institutional details).
- You may also have to wait for a username and password before you can access data.
- You may need to register the use of data.

Terms and conditions - You will usually be agreeing to terms and conditions as part of the registration and process or when accessing data.

Common terms and conditions include not trying to identify individuals, households or organisations in the data and not distributing data to others. Some restrictions about use of data are common e.g. "data is for non-commercial use only" or for "use in research or teaching" only.

Download - Often you download data directly from the catalogue. However, some services ask users to complete a form to request data.

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.

cessda

- Additionally, access to some data collections may require permission from the data owners (and therefore often an additional stage in the access procedure).
- More strict applications generally operate for any data collections considered to contain sensitive or confidential data and some services provide access through dedicated safe rooms or remote settings.
- Access to data, documentation and metadata is generally free. Charges can apply for commercial use and for supplementary services such as burning data to a CD or distributing data by post.
- You should find access conditions displayed clearly in the catalogue for each data collection. If you are unsure however, 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)

cessda

When using existing data, it is good practice to cite the data. Data citations gives credit to the data creators and allows other researchers to find the data.

Many services provide recommended citations for each data collection and in general, a citation should include enough information so that the exact version of the data being cited can be located.

Many services give persistent identifies for their data such as a Digital Object Identifier (DOI), which link to the data used (even if the location of the data changes).

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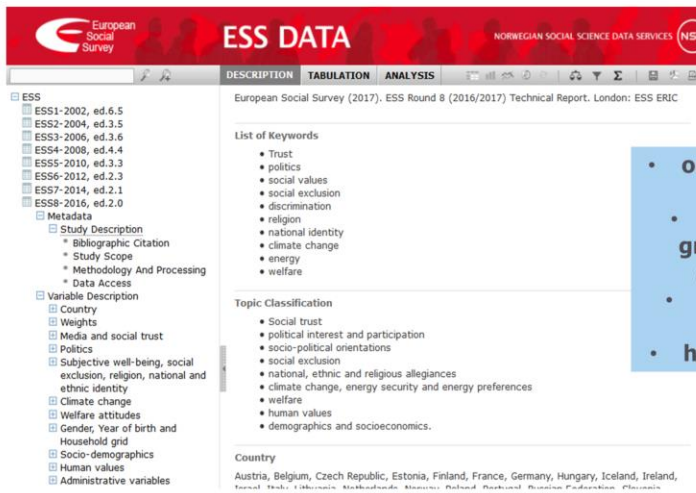
Questions regarding data or documentation, please contact essdatasupport@nsd.no

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

cessda

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

cessda

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.

cessda

www.cessda.eu/DMEG

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](#)

cessda

Questions

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

cessda.eu   @CESSDA_Data