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

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

**Types of data**

- Quantitative
- Qualitative

**Examples of data sources**
- Text documents
- Laboratory notebooks
- Books and diaries
- Questionnaires
- Scripts and codebooks
- Audiotapes
- Photographs and films
- Minimization results
- Specimens
- Sample artefacts
- Slides
- Database schemas
- Base contents
- Models
- Algorithms
- Methodologies
- Workflows
- Operating procedures and protocols
- Experimental results
- Metadata
- Other literature review records
- Archives
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 snapshot)
- 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 attributes (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

There are many sources of data.

CESSDA Training Working Group (2017)
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
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 coordinating 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
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 (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)
HOW IT WORKS

1. Get funding
2. Write DMP
3. Gather data during research
4. Finish research and deposit data
5. Publish results
6. Choose repository
7. Inform OpenAIRE
   - the EU-funded Open Access portal

reach more people, have greater impact
avoid duplication of efforts
preserve data for future researchers

Source: EC
Open Access to research data

Importance of research infrastructures / data repositories

Source: EC
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
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)
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)
International Social Survey Programme (ISSP)

- annual programme (started in 1984)
- cross-national collaboration
- rotating thematic modules e.g.
  - Citizenship: 2004 and 2014
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)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Media and social trust</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Politics</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Subjective well-being...</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Gender, Household</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Socio demographics</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Human values</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Immigration</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Citizen involvement</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Health and care</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Economic morality</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Family work and well-being</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

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

Source: ESS
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
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
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)
Factsheet: Accreditation & Data Access Conditions for SLOVENIA

Contact

Name:
Statistiki urad Republike Slovenije (Statistical Office of the Republic of Slovenia, SORS)

Postal Address:
Litostroška cesta 54 SI-1000 Ljubljana (Slovenia)

Contact:
info.stat@spri.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).
Labour Force Survey - 2011

Original Title: Analiza delovne 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 on site 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 adult population, which includes persons that are mainly living in the territory of Republic of Slovenia, regardless of their nationality. Only the population, living in individual households, is covered by the survey. Institutionalized people are not considered as a part of the target 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 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
Some research projects share research data through project websites

http://cwed2.org/

WELCOME

The Comparative Welfare Entitlements Dataset (CWED) contains information about the structure and generosity of social insurance benefits in 33 countries around the world. Since September 2017, an updated version of CWED 2 containing a total of eight household types is available. 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.

Download CWED 2 data here

NEWS

June 22, 2017  Updated list of scientific works and papers is available

More than 200 peer-reviewed scientific works are using CWEDZ data. You can access an updated of all books, papers, and chapters in edited volumes using our dataset here.

June 00, 2017  Forthcoming publications from CWED members
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
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
1. Develop a clear picture of the data you need
2. Locate appropriate data resources
3. Set up a search query and search the data resource
4. Select data candidates
5. Evaluate data quality

CESSDA Training Working Group (2017)
Identify what you need
Four ways we can use archived data

<table>
<thead>
<tr>
<th>Use of methodology</th>
<th>New analysis: one or multiple data sources e.g. combine micro and macro, just secondary data or secondary data combined with primary data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td></td>
</tr>
<tr>
<td>Use of design</td>
<td>Use of study design/methodology (e.g. data collection tools (interview schedules &amp; survey questions) or sampling strategies)</td>
</tr>
<tr>
<td>Teaching</td>
<td>Teaching: Subject-based or research methods, Datasets made for training purposes – e.g. easySHARE</td>
</tr>
</tbody>
</table>
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

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)
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)

SND Swedish National Data Service

Find and order data

Filter/sort

language

Find and order data

political attitudes

247 hits

Socio-political attitudes 1975

Socio-political attitudes 1979
## 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

### ADP CATALOGUE

Date of the last change of the catalog: 06.07.2018

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Study, title, topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVET12</td>
<td>7 EU VET - Study on vocational education in seven European countries  topic: EDUCATION, EDUCATION - vocational education, produced by CER</td>
</tr>
<tr>
<td>EWCG051</td>
<td>European Working Conditions Survey 2005 topic: LABOUR AND EMPLOYMENT, LABOUR AND EMPLOYMENT - employment, produced by EUROSOC</td>
</tr>
<tr>
<td>EWCG051</td>
<td>Working Conditions in European Union Candidate Countries 2001 topic: LABOUR AND EMPLOYMENT, LABOUR AND EMPLOYMENT - employment, produced by EUROSOC</td>
</tr>
<tr>
<td>FMAAT16</td>
<td>Internectivity and Feminist Activism, 2011: Student Feminist Societies in the United Kingdom topic: SOCIAL STRATIFICATION AND GROUPINGS, SOCIAL STRATIFICATION AND GROUPINGS - gender and gender roles, produced by None</td>
</tr>
</tbody>
</table>
Economic Morality (ESS2 2004)
Data/Variables - Round 2 (2004)

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

Source: ESS
Documentation often extensive
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

<table>
<thead>
<tr>
<th>LITERAL QUESTION</th>
<th>Code</th>
<th>Frequency</th>
<th>% of all</th>
<th>% of valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>F17c2. POST CODE: MAIN ACTIVITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values</td>
<td>Categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Paid work</td>
<td>1</td>
<td>25,760</td>
<td>47.1</td>
</tr>
<tr>
<td>2</td>
<td>Education</td>
<td>2</td>
<td>4,704</td>
<td>8.6</td>
</tr>
<tr>
<td>3</td>
<td>Unemployed, looking for job</td>
<td>3</td>
<td>2,978</td>
<td>5.4</td>
</tr>
<tr>
<td>4</td>
<td>Unemployed, not looking for job</td>
<td>4</td>
<td>1,155</td>
<td>2.1</td>
</tr>
<tr>
<td>5</td>
<td>Permanently sick or disabled</td>
<td>5</td>
<td>1,291</td>
<td>2.4</td>
</tr>
<tr>
<td>6</td>
<td>Retired</td>
<td>6</td>
<td>12,819</td>
<td>23.4</td>
</tr>
<tr>
<td>7</td>
<td>Community or military service</td>
<td>7</td>
<td>90</td>
<td>0.2</td>
</tr>
<tr>
<td>8</td>
<td>Housework, looking after children, or other</td>
<td>8</td>
<td>4,796</td>
<td>8.8</td>
</tr>
<tr>
<td>9</td>
<td>Other</td>
<td>9</td>
<td>520</td>
<td>1.0</td>
</tr>
<tr>
<td>66</td>
<td>Not applicable</td>
<td>66</td>
<td>66</td>
<td>1.0</td>
</tr>
<tr>
<td>77</td>
<td>Refusal</td>
<td>77</td>
<td>39</td>
<td>0.1</td>
</tr>
<tr>
<td>88</td>
<td>Don't know</td>
<td>88</td>
<td>105</td>
<td>0.2</td>
</tr>
<tr>
<td>99</td>
<td>No answer</td>
<td>99</td>
<td>416</td>
<td>0.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>54,573</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Source: ESS; Dataset: ESS6-2012, ed.2.4
Crosstabs – frequency distribution of two variables

**Socio-demographics**

- 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

**Politics**

- Placement on left right scale
- How satisfied with life as a whole
- How satisfied with present state
- How satisfied with the national
- How satisfied with the way democracy works
- State of education in country
- State of health services in country
- Government should reduce differences in income levels
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?

<table>
<thead>
<tr>
<th>Main activity, last 7 days. All respondents. Post coded</th>
<th>Median</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Standard deviation</th>
<th>Sum</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paid work</td>
<td>7.5</td>
<td>7.0</td>
<td>0.0</td>
<td>10.0</td>
<td>2.2</td>
<td>179,423.0</td>
<td>25,653</td>
</tr>
<tr>
<td>Education</td>
<td>8.0</td>
<td>7.7</td>
<td>0.0</td>
<td>10.0</td>
<td>1.9</td>
<td>35,885.0</td>
<td>4,676</td>
</tr>
<tr>
<td>Unemployed, looking for job</td>
<td>5.6</td>
<td>5.4</td>
<td>0.0</td>
<td>10.0</td>
<td>2.7</td>
<td>16,028.0</td>
<td>2,956</td>
</tr>
<tr>
<td>Unemployed, not looking for job</td>
<td>5.7</td>
<td>5.6</td>
<td>0.0</td>
<td>10.0</td>
<td>2.7</td>
<td>6,347.0</td>
<td>1,139</td>
</tr>
<tr>
<td>Permanently sick or disabled</td>
<td>5.6</td>
<td>5.5</td>
<td>0.0</td>
<td>10.0</td>
<td>2.8</td>
<td>7,037.0</td>
<td>1,278</td>
</tr>
<tr>
<td>Retired</td>
<td>7.0</td>
<td>6.5</td>
<td>0.0</td>
<td>10.0</td>
<td>2.5</td>
<td>83,123.0</td>
<td>12,725</td>
</tr>
<tr>
<td>Community or military service</td>
<td>7.6</td>
<td>6.9</td>
<td>0.0</td>
<td>10.0</td>
<td>2.6</td>
<td>618.0</td>
<td>90</td>
</tr>
<tr>
<td>Housework, looking after children, others</td>
<td>7.6</td>
<td>6.7</td>
<td>0.0</td>
<td>10.0</td>
<td>2.4</td>
<td>5,899.0</td>
<td>517</td>
</tr>
<tr>
<td>Other</td>
<td>7.6</td>
<td>7.0</td>
<td>0.0</td>
<td>10.0</td>
<td>2.4</td>
<td>3,859.0</td>
<td>517</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>7.30</td>
<td>6.76</td>
<td>0.00</td>
<td>10.00</td>
<td>2.40</td>
<td>363,529.00</td>
<td>53,801</td>
</tr>
</tbody>
</table>

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

nesstar_gesis

Advanced Search

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

• very good
• 2 good
• 3 fair
• 4 poor
• 5 very poor
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
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
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

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

**Metadata ("data about data")**
- descriptors that facilitate cataloguing data and data discovery.

**Documentation**
- user guides, survey questionnaires, interview schedules and fieldwork notes
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
Now finally, I’ve found some great data, how to I get it?

- Licenses
- Access process
- Getting started
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 (= an 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)
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 2.1. 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 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

**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


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.
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.
Data Management Expert Guide

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

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

You can download the full text for your personal study offline. More for every single chapter are also available for being printed as handouts for training.

Target audience and mission

This guide is written for social science researchers who are in an early stage of practicing research data management. With this guide, CESSDA wants to contribute to professionalism in data management and increase the value of research data.
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)