How to find and access data

## Trainer notes

The *How to Find and Access Data in Europe* workshop materials have been developed to introduce types and sources of secondary data useful for social science research. Designed for anyone interested in secondary data analysis, the materials

• introduce key data sources including CESSDA and European data archives

• discuss tools and strategies for finding data

• cover how to evaluate the quality and usefulness of data

• explain key information about licenses and conditions of access

## Notes

This document outlines the overall content of the materials including some approximate times for different sections. Trainers can find detailed content notes in the slides (PowerPoint).

The materials were designed to form a 2-2.5 hour session. However, trainers can remove or extend adapt sections and allow more or less time for activities and discussion. Some example schedules are included at the end of this document.

The practical activities included are based around an example research question. The example research question can be changed to reflect interests of participants and/or participants can use their own research questions. Activities do not need to be prescriptive; participants might prefer just to browse.

## Introduction to session including introduction to CESSDA

|  |  |  |
| --- | --- | --- |
| This section briefly introduces the session with some background on CESSDA, opens science and national data services. Keep this brief and consider what background information is needed.  Time: 5-10 min | | |
| Today’s workshop | Overview of session: 1) 5 main sections 2) lecturers with some interactive/practical exercises 3) assumes no prior knowledge but you may come with lots; opportunity for discussion. | Time: 1-3min |
| Introduction to CESSDA/ National data Services | Background: 1) CESSDA, activities and open science and 2) What and who are national data services | Time: 4-8 min |

## Part 1 Data types and sources

|  |  |  |
| --- | --- | --- |
| This section introduces types and sources of data useful for social and economic research. Activity 1 acts as a warmer and gives insight into participant’s background knowledge.  Idea: Get participants to quickly introduce themselves as part of the feedback. For shorter session, avoid the feedback session taking too long.  Time: 25-30min | | |
| Data Landscape | Many of parts of the data landscape are easy to navigate but the volume of data and different locations of data can make it difficult. This session aims to help. | 1-2 minutes |
| Activity 1 | Aim is to get participants talking, identify some key data sources/types, and assess participants’ knowledge. See slide for details. | Time:12 min  (3 mins activity time, 7-9 min feedback time) |
| Types of data | Thinking about different types of data can help you work out what you need and how to find it among the volume of data available  Three ways to differentiate types of data available (1. Quantitative/qualitative 2. Level of analysis and 3. Time)  \*\* will be basic to some but knowledge of different data types/classifications is a precursor to developing effective data searches. | Time: 5 min  Adjust speed/depth to reflect knowledge/experience of group (as the activity as revealed) |
| Sources of data | This section introduces some major sources of data (1) European social science data archives 2) Major cross-national studies, 3)Longitudinal studies 4) key data providing organisations (Eurostat- case study) 3) Project/researcher websites 4) Selected non-European archives and 5) Repositories (and Re3data) | Time 10 min |

## Part 2 Identifying data needs

|  |  |  |
| --- | --- | --- |
| This section focuses on identifying the type of data needed to address a research question as a precursor to a data search  Time 25-30 min | | |
| Four ways we can reuse data | IDEA: get participants to make suggestions – and/or refer back to activity 1  New analysis, replication, use of study design/methodology and teaching | Time: 2-3 minutes |
| Identifying data needs (for using data in research) | 1. Identifying what you need is a big part of finding data for your research question 2. What is the ideal dataset for a research question? (May need to compromise later but useful to have some basis to evaluation what you find). 3. How? key concepts [example political behaviour], how to operationalise [examples = religiosity, social class], the Ideal Population, Geographical coverage, Time, Units of analysis, Study design and sample | Time:7-10 min |
| Activity 2 | In groups…Participants determine an ideal dataset for a research question. See slide and handout. Get feedback from groups.  The next slide includes some possible answers, which can help focus the next activity.  NB: This exercise could be developed around any research question. You could also ask participants to do it for their own research question. | Time: Give 10 min (explain it is a quick exercise to generate ideas)  Up to 5 min for feedback |

## Part 3 Searching data archives

|  |  |  |
| --- | --- | --- |
| This section address how to search the data catalogues of data archives; the CESSDA data catalogue will becomes the focus once fully operational.  Time: 25 min | | |
| Searching data archives | Starts with 3 common search types and then introduces data catalogues (including the new CESSDA catalogue) and NESSTAR [could extend session with demo]. It then looks at the reality of finding data with reference to common problems and solutions including ELSST as a tool for selecting search terms. | Time: 10 min |
| Activity 3 | Two tasks 1) Search of data archive and 2) use ELSST to select and refine search terms (in other languages)  NB: Can split group into sub-groups and get each to focus on a specific task(s)  Can relate searches to an example e.g. searching for data to examine cultural backlash or give participants option to do their own searches | Time: 15 minute activity including feed back |

## Part 4 Evaluating data: quality and usefulness

This section looks at how a potential user might assess the quality and usefulness of a dataset using metadata and documentation and with reference to their understanding of the idea dataset.

Time: 25 min approx.

|  |  |  |
| --- | --- | --- |
| Evaluating data using metadata and documentation | Discusses the need to consider data collection and processing in secondary data analysis introducing metadata and other data documentation.  How to assess quality (list of questions to ask).  Judging if the data is useful for you with reference to the ideal dataset. | 7 min |
| Activity 4 | Evaluate the usefulness of a data collection  Use an example research question [example included continues the cultural backlash example] or participants can assess data for their own research question | 15 min  + feedback |

## Part 5: Accessing data

This section focuses on accessing data with reference to licenses, conditions of access and common access arrangements and then looks at how user need to reference the documentation to make sense of data

Time: 20min (extend by including A Guide to Getting Started)

|  |  |  |
| --- | --- | --- |
| Data access arrangements | 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)  CESSDA general principal  Some common arrangements across data services include | 10 |
| Getting started | Discuss how documentation key to understanding how to use the date  [Additional extra - A GUIDE TO GETTING STARTED (WITH THE EUROPEAN SOCIAL SURVEY) – Give to participants for later reference or integrate into session] | 7-10  [10min] |
| Data Citation | Why and how to cite data (DOI) | 1 min |

## Ending – Further resources

|  |  |  |
| --- | --- | --- |
|  | [Additional extra - A GUIDE TO GETTING STARTED (WITH THE EUROPEAN SOCIAL SURVEY) |  |
| CESSDA Training website | Highlight   * Events calendar * Data Management Guide – chapter on Data Discovery * Resources section | 3 min |

## Example schedules [time in brackets=cumulative]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2.5hour | 2 hour | 2 hour | notes |
| Introduction to session | 5min [5min] | 5min [5min] | 5min [5min] |  |
| 1 Data types and sources including activity | 30 min [35min] | 25 min [30min] | 25min [30min] | * Keep activity feedback brief * Go through data types and sources briefly |
| 2 Identifying data needs | 30min [1hr 5min] | 25 [55min] | 25 [55min] |  |
| 3 Searching data archives | 30-25min [1 30 hour] – break here | 25 [1 hr 20min] | 25 min [1 hr 20min] | Participants will need access to PCs (which may affect time needed) |
| 4 Evaluating data: quality and usefulness | 25min [1hr55min] | 20 [1hr40] | Combine 4+5, drop activity 4 but allow time at the end for exploring data sources/how to access.  20 min content [1hr 40min]  +20min activity/questions [2hrs] |  |
| 5 Accessing data | 20min [1hr15]  Questions | 15 min [1hr55]  5min questions[2hours] |