

cessda eric

Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Research Data Management

International Summer School
in Uganda

Dr. Anja Perry
Oliver Watteler

gesis

Leibniz Institute
for the Social Sciences



This work is licensed under [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

CESSDA ERIC

Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Introductions

Who we are

Oliver Watteler

Anja Perry



Image by See-ming Lee (CC-by-sa).

Where we come from

German Data Archive for the Social Sciences at GESIS

Established in 1960 (since 1986 a part of GESIS)

Our data

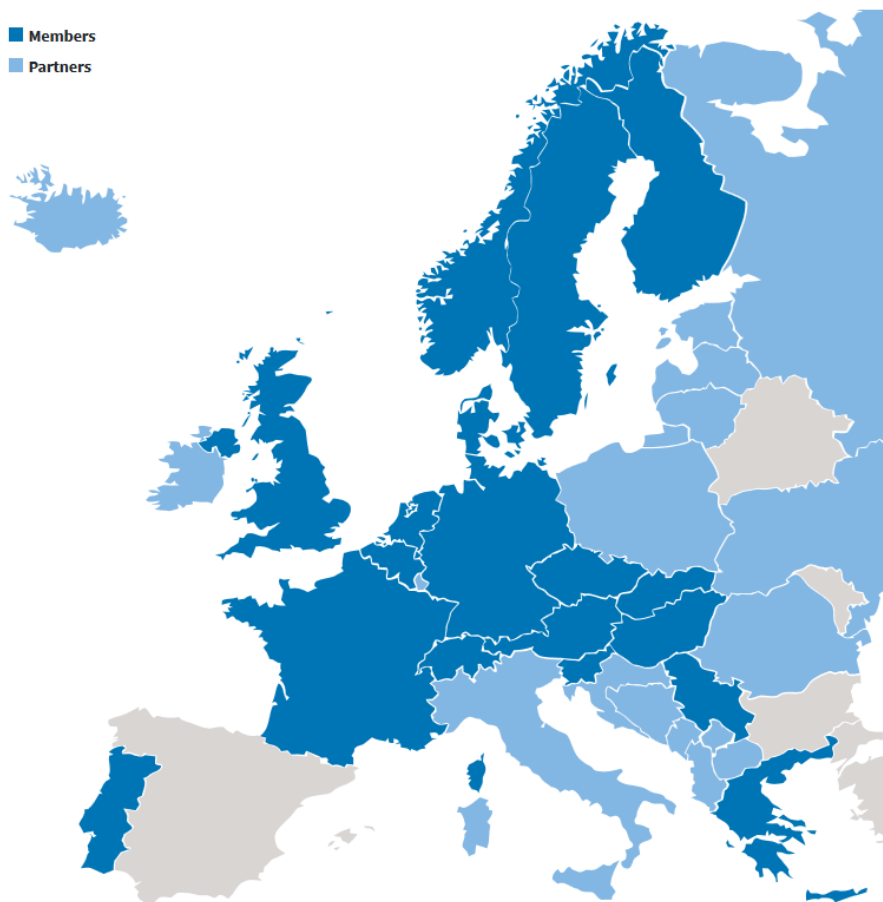
- about 6.200 national and international studies
- 3 research data centers

Our services

- software development
- processing, documenting and archiving quantitative data, such as ISSP, CSES, EVS
- consultation and support
- ...



What is CESSDA?



- Network of social science data archives across Europe
- Sustainable research infrastructure
- Enables research community to conduct high-quality research

The floor is yours!

- What's your name?
- Where do you work? (name, city, country)
- What's your research about? (in 20 words or less)
- Why are you here?

Preliminaries

- Feel free to ask questions at any time
- We want to make this workshop as interactive as possible
- Slides and other materials are available through ILIAS

Outline of the Workshop

Mornings

8:30 am Theoretical Session

10:30 am Coffee Break

11 am Theoretical Session

12:30 pm Lunch

Afternoons

1:30 pm Practical Session (work on your own)

3 pm Coffee Break

3:30 pm Practical Session (work on your own)

5 pm Evening events or free time

Outline of the Workshop

This week's schedule

Monday Introduction to RDM and Data Management Plans

Tuesday Data Collection: Ethical aspects, Data Protection, Informed Consent

Wednesday Data Processing and Anonymization

Thursday Data Documentation and Metadata

Friday Data Storing, Archiving, and Sharing

Certificate of accomplishment



Image: pixabay (CC-0)

- Assignments throughout the week
- Work on your own in the afternoons
- We are here in case you have any questions, but feel free to work wherever you wish
- Submit your results on Friday, Sept. 20th, at 11:59 pm

CESSDA ERIC

Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Introduction to Research Data Management

What is Research Data Management?

RDM is like “health care” for your data

- keeps them safe from harm
- makes them usable and discoverable

All strategies, processes and measures to maintain

- data quality
- interpretability of research results
- (re-)usability of research data

Specifically addresses legal and ethical issues (e.g. informed consent, ownership, licensing)



Image: pixabay (CC-0)

Benefits: teams, researchers, & scientific community

- Saves resources
- Correct results
- Better workflow

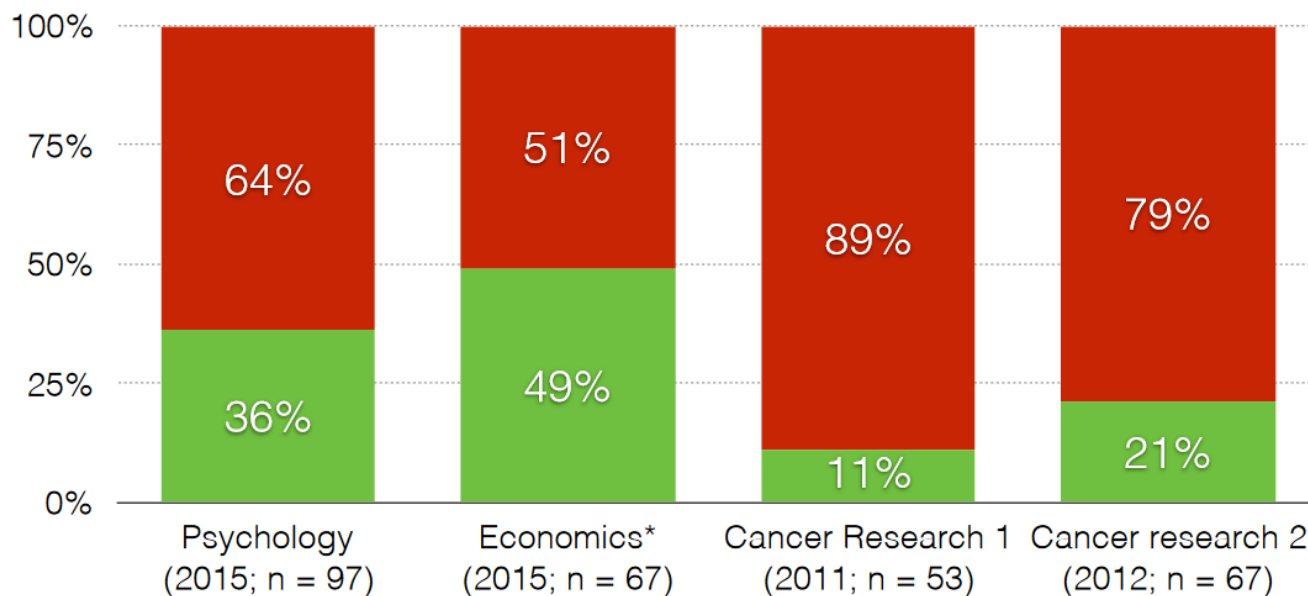


- Proof of transparent & valid conduct
- Facilitates re-use of own data

- Better access to data
- Best use of public funds
- ...



Replication of published research findings



Open Science Collaboration (2015);
 Chang & Li (2015);
 Begley, C. G., & Ellis, L. M. (2012);
 Prinz, F., Schlange, T., & Asadullah, K. (2011)

* The data on economics is about *reproducibility*; i.e. the attempt to get the same results if you apply the original data analysis on the original data set.

RDM may be required by...

- Your **institute**, e.g. by your work contract or **project** agreements
- **Funding agencies**, e.g. to ensure reusability
- **Journals**, claiming your data before publishing your article
- Your **supervisor**...

⇒ check for such conditions!



Image: A. Herrema & H. Bouwteam (CC-by)

Goals of Research Data Management

1. Quality assurance

readable, authentic data;
avoidance of data loss

2. Replicability

comprehensibility; enable
reproducibility of research
results; proof of validity

3. Reusability

using the data in new
contexts; answering new
research questions





Data Management Plans

The Data Management Plan (DMP)

A DMP is an important tool to structure RDM of your project

- defines strategies, measures and responsibilities for
 - processing and validating
 - storing and protecting
 - preserving and sharing your data
- throughout the data lifecycle



Image: pixabay (CC-0)

How to write a DMP

A DMP uniquely relates to your research project

- start right at the beginning of your research project
- do not copy and paste!
- document what you did and why
- frequently update and adapt it
- A DMP is not just a plan, it is the implementation of a (research) plan

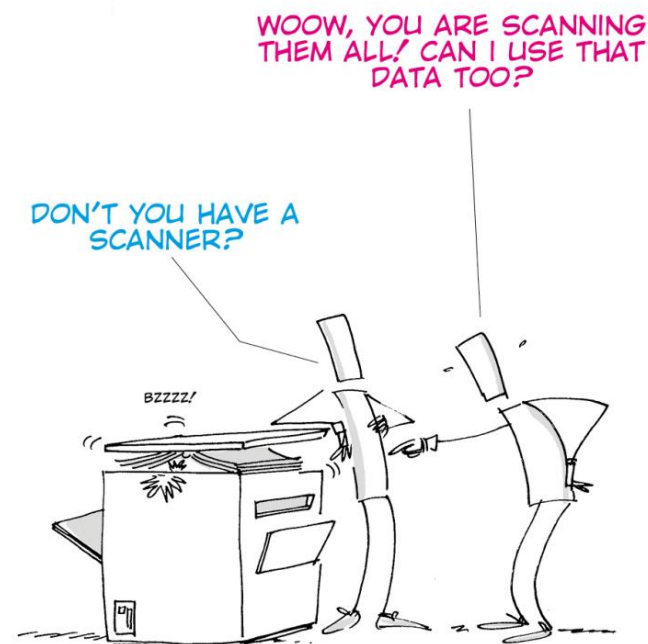


Image by P. Hochstenbach (CC-BY)

How to get started?

- **Rely on the data lifecycle**
- **Various online templates and tools available, e.g.**
 - CESSDA (2018). Adapt your DMP – list of questions:
https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf
 - DCC (2013). Checklist for a Data Management Plan:
http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf
 - Tools: DMPTool (<https://dmptool.org/>) or DMP Online (<https://dmponline.dcc.ac.uk>)

Have a first look!

- **Have a look at the DMP template provided**
- **Familiarize yourself with its sections and subsections**
- **Consider ...**
 - ...parts you would be able to fill in off the top of your head
 - ...parts you would have to look for further information
 - ...obstacles or problems you encountered in the past



Image: pixabay (CC-0)

 individual work -  time: about 15 minutes

Structure of a DMP

Consists of various parts

- Cover page and general information on the project
- Seven sections (with subsections):
 1. Overview
 2. Organize and document ⇒ **day 4 and 5**
 3. Process ⇒ **day 3**
 4. Store ⇒ **day 5**
 5. Protect ⇒ **day 2**
 6. Archive and publish ⇒ **day 5**
 7. Discover ⇒ **day 1**



cessda eric

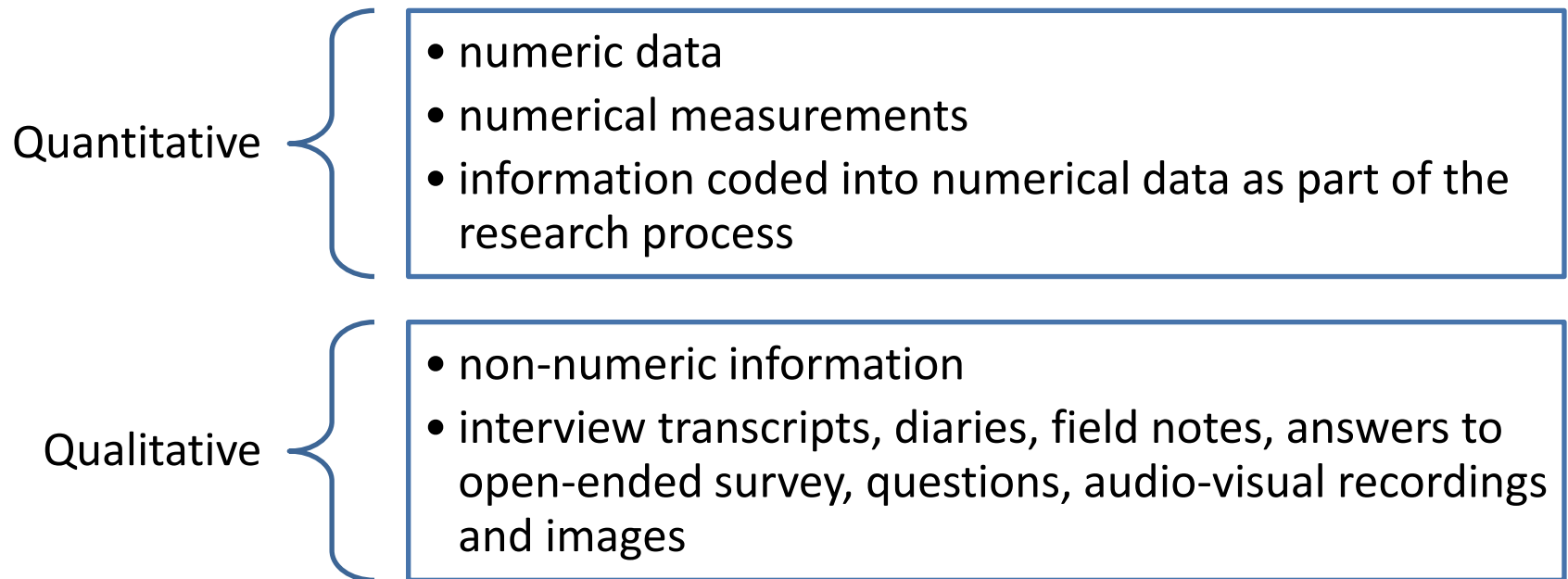
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Data Discovery

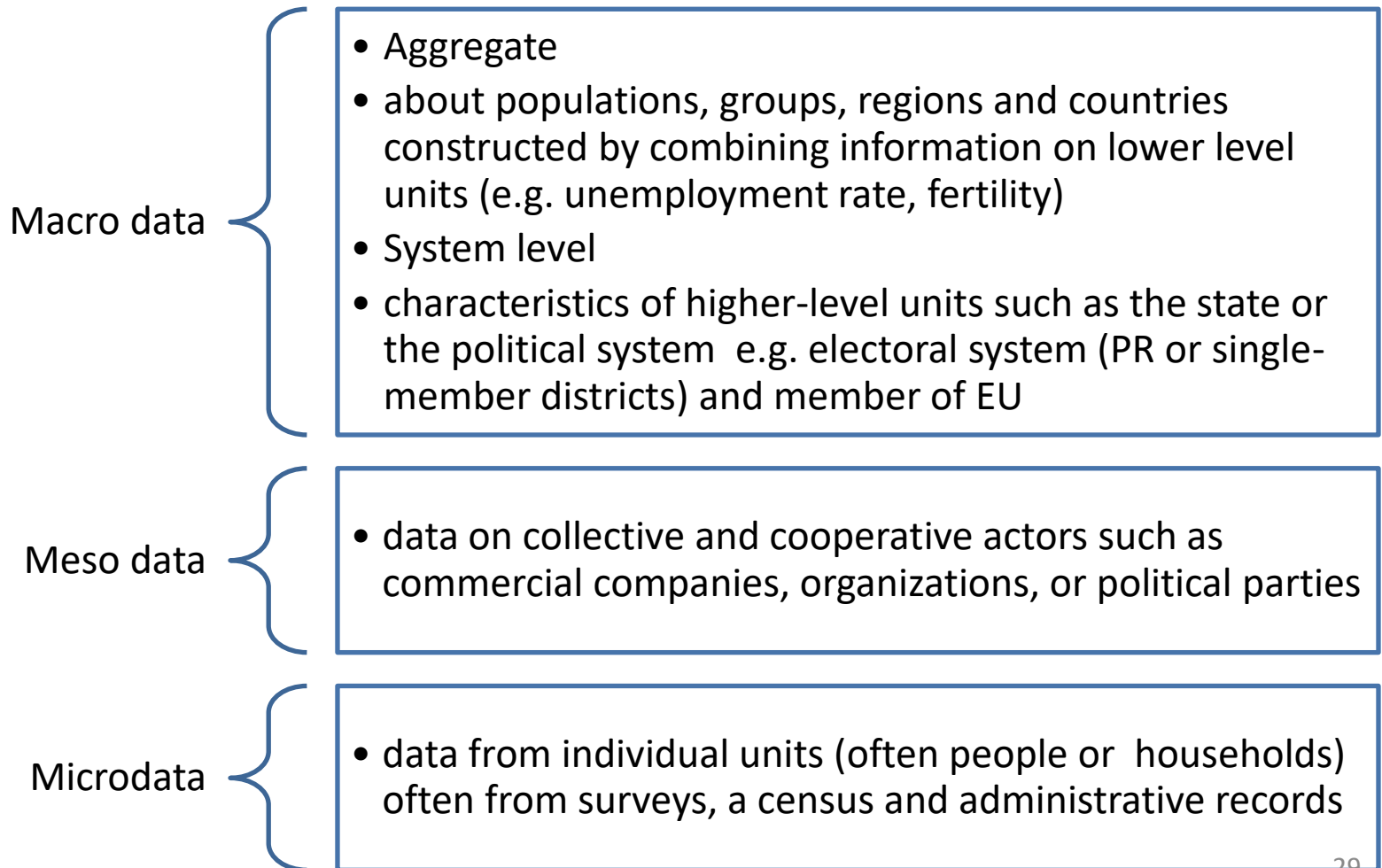
Why search for data?

- Reuse saves costs and time
- Compare results and make replications
- Reuse verified elements of research design
- Enhance data quality and foster innovation

Types of data: quantitative and qualitative



Types of data: level of analysis



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
- data points may come from sample surveys, e.g. unemployment rate 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'

Identifying data needs



Research Question

- What is the ideal dataset for addressing this question?



Key concepts

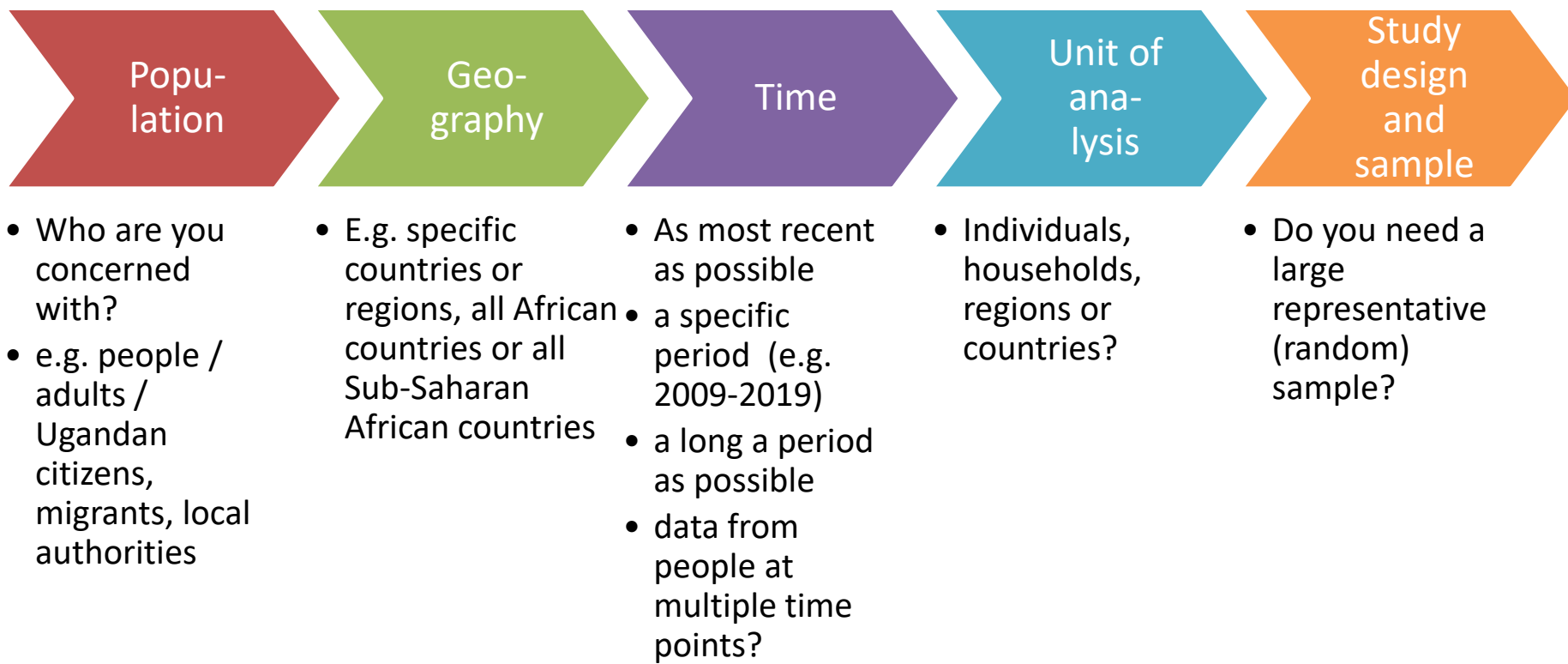
- Key features
- Multidimensional
- Groups of people
- Dependent/
independent variables



How to operationalise?

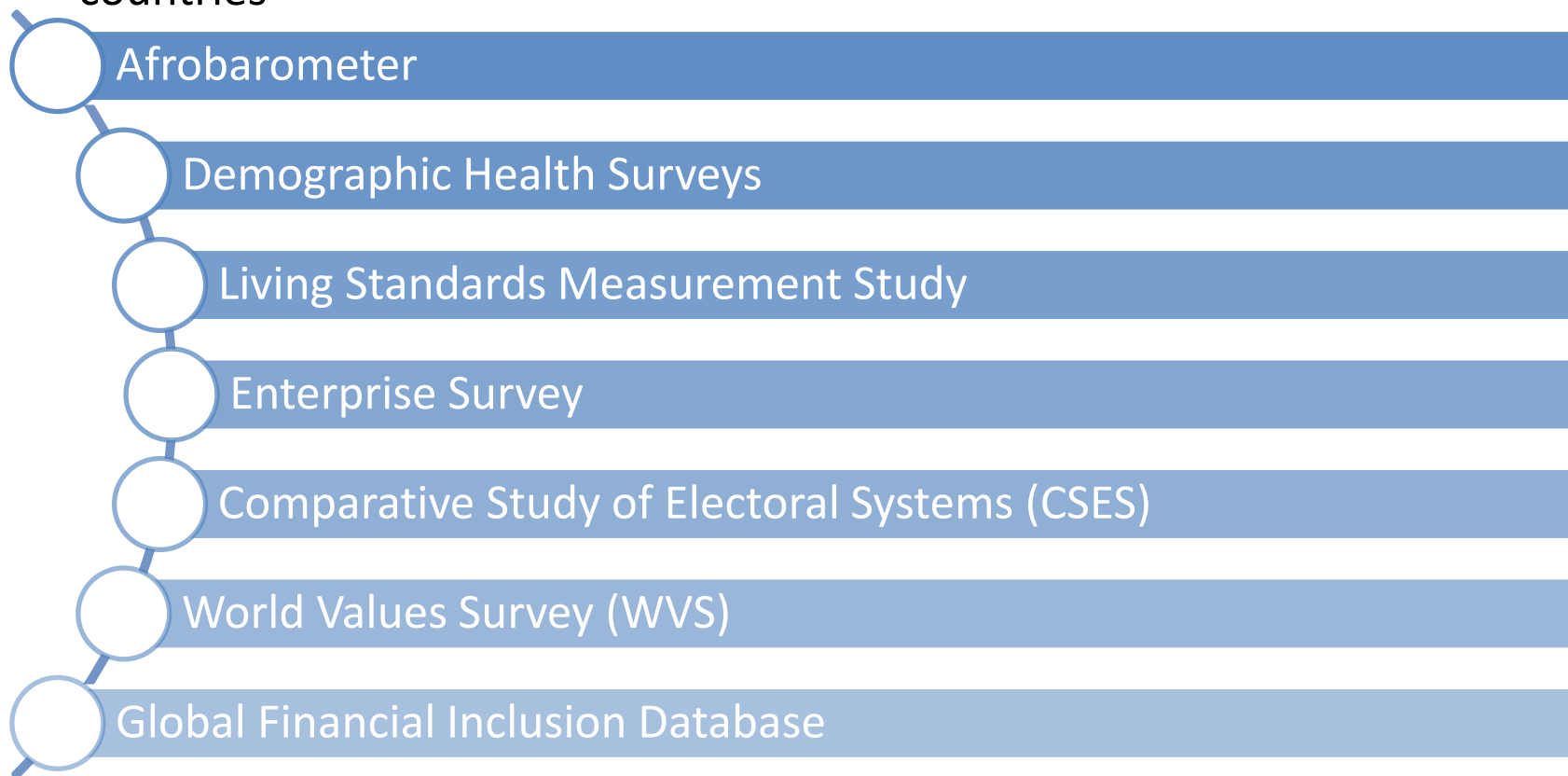
- What key features
- Multiple variables?
- Comparable/established measures (e.g. Schwarz Human Values)
- Standardised (e.g. ILO unemployment, ISCO, ESeC)

Identifying data needs



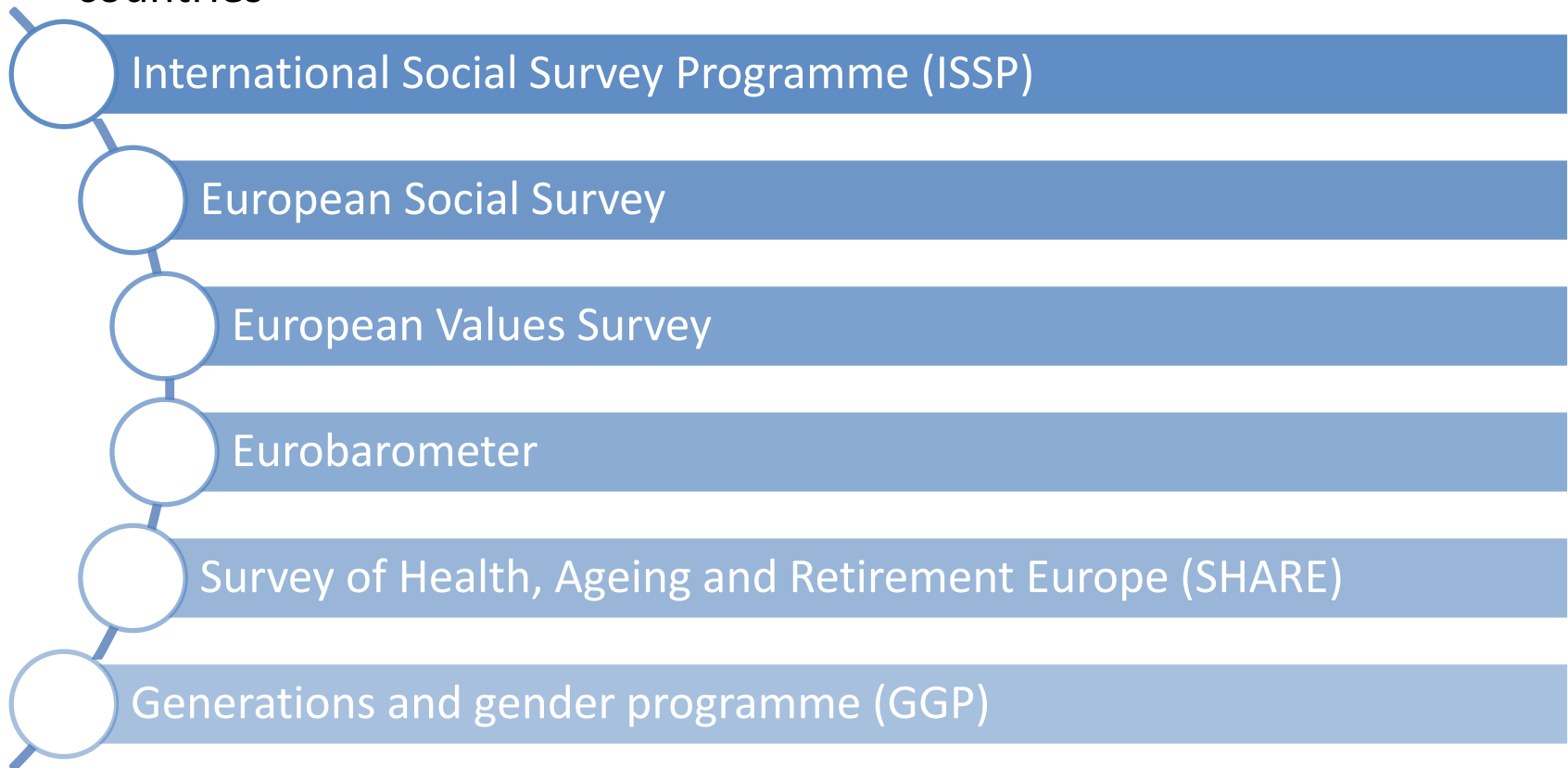
Cross-national studies

International survey research programmes that include many African countries

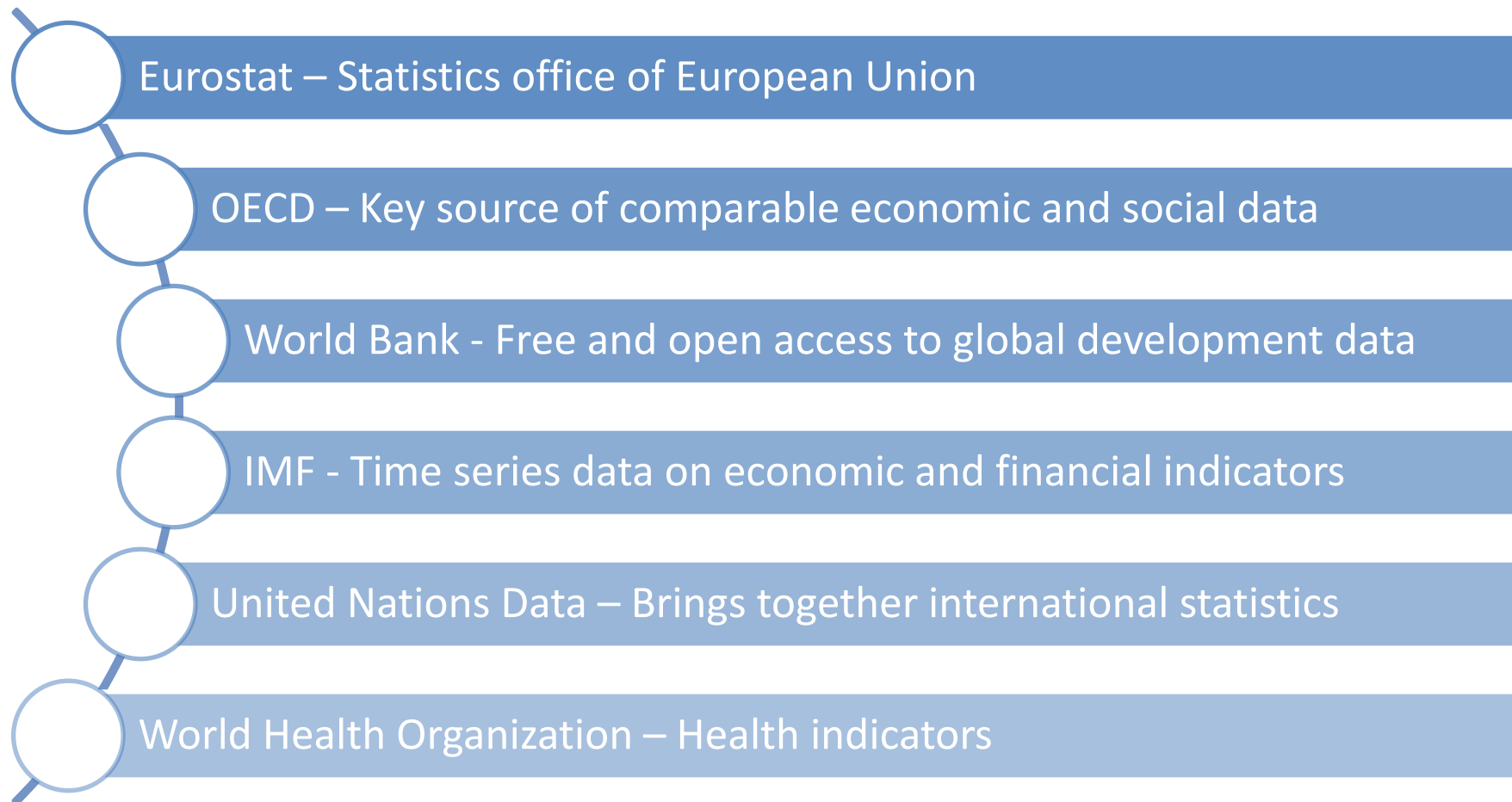


Cross-national studies

International survey research programmes that include many European countries



Key data providing organisations



Data from different sources

BP123: Governance and democracy attitudes in higher performing African countries

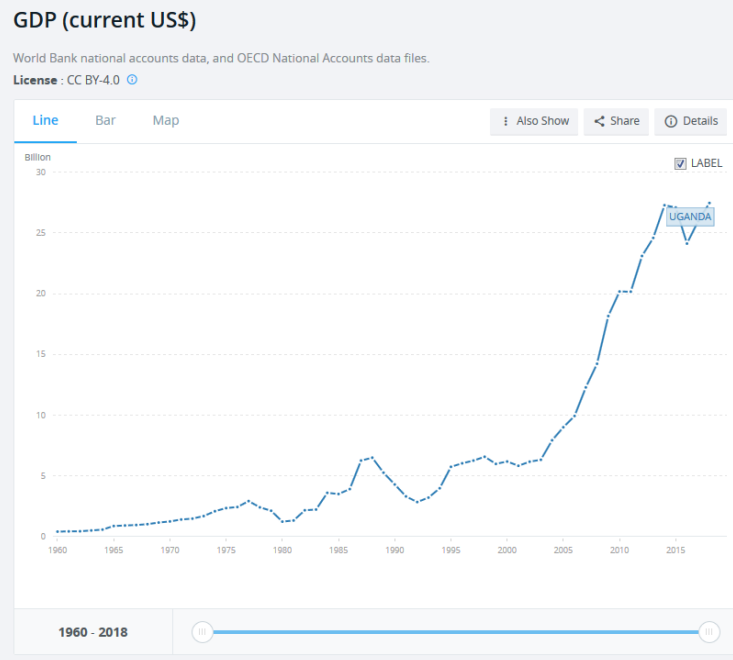
Namibia is usually regarded as one of the best performing democracies in Africa. Using the Afrobarometer Round 5 survey, this paper compares public attitudes that are central to democratic life across high performing countries in Africa. Several countries usually at the top of democracy ratings will be included in the comparison of democratic attitudes. Small population democracies such as Botswana and Namibia, and Botswana, and Botswana are the most comparable to Namibia, and Ghana are also included because of comparable democracy ratings.

Briefing papers

2013

Download now 

<https://afrobarometer.org/publications/bp123-governance-and-democracy-attitudes-in-higher-performing-african-countries>



<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2018&locations=UG&start=1960&view=chart>

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 -

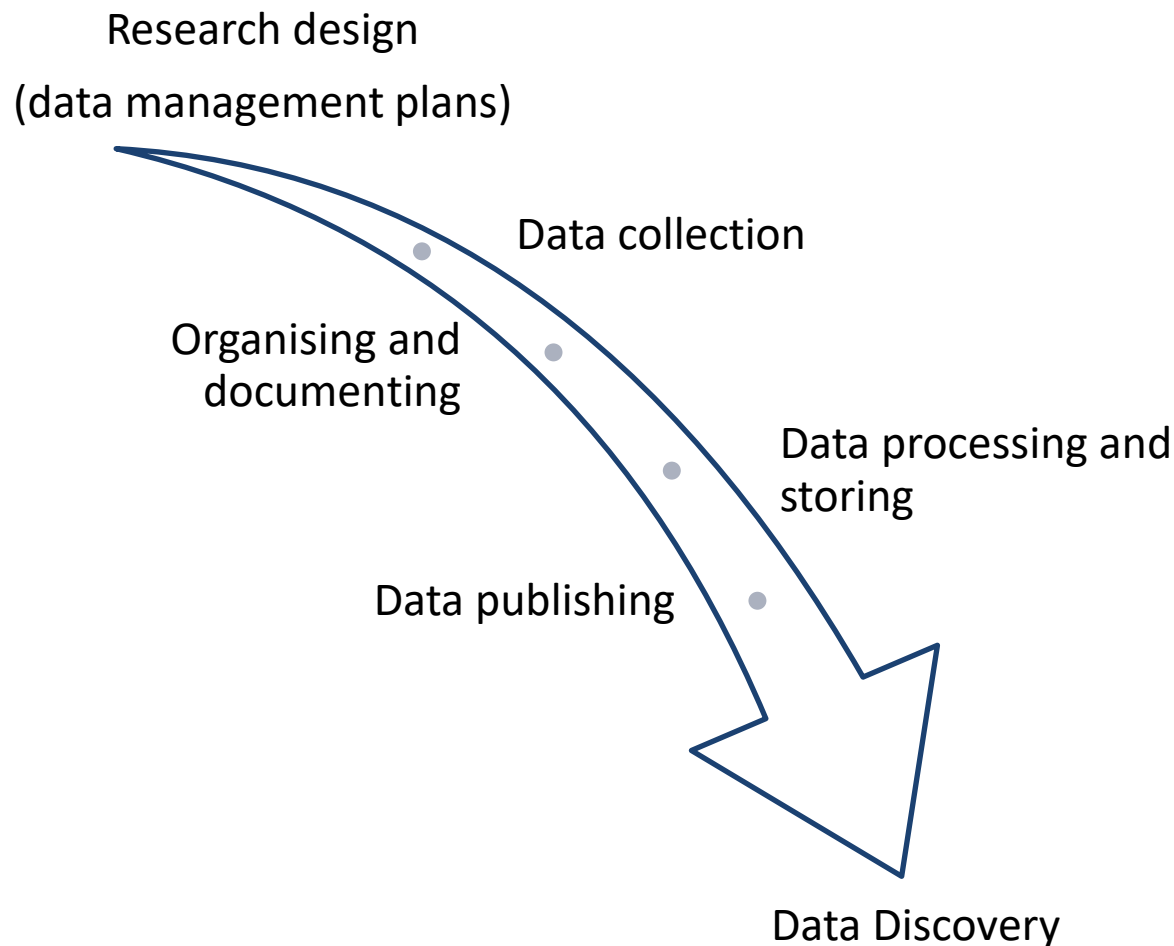
Institutional research
data repositories

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

Steps in data discovery

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 candidates
5. Evaluate the data

The longer process



Consider the previous processes to assess data quality and suitability

- What information was collected, from whom, when and where?
- What has been done to the resulting data?



**Thank you for your
attention!**

anja.perry@gesis.org
oliver.watteler@gesis.org



Your task

- Define a research question!
- Either use your own research question or pick one of our examples:
 - How does single parenting affect a child?
 - Gender roles in a family
 - Impact of foreign aid on national economies
 - Youth culture in Uganda
 - What effects voting behavior?

Your task

- Consider the bullet points in your exercise booklet to identify data needs to explore theories of your research question.
- Search for the data and describe how you approached your search.
- Have you found a good candidate? Evaluate the data.
- Now, let's imagine you cannot find the data that fits your needs and you plan to collect data yourself. Start setting up a Data Management Plan (DMP) and fill in everything you can already answer. We will return to your DMP throughout the week.

Possible search results for reasearch questions

- Search e.g. for „[concept]“ + Uganda + data.
- Or search for entire research questions.

Questions	
How does single parenting affect a child?	Uganda Household Surveys
The gender roles in family	Oxfam
Impact of foreign aid on national economies	OECD
Youth culture in Uganda	International Alert Youth Policy Press
What effects voting behavior?	Afrobarometer