cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Open Science and Open Data

Summer School Short Course

FOSTER

seriss SYNERGIES FOR EUROPE'S RESEARCH INFRASTRUCTURES IN THE SOCIAL SCIENCES

gesis Leibniz Institute for the Social Sciences

Dr. Anja Perry
Dr. Kristi Winters

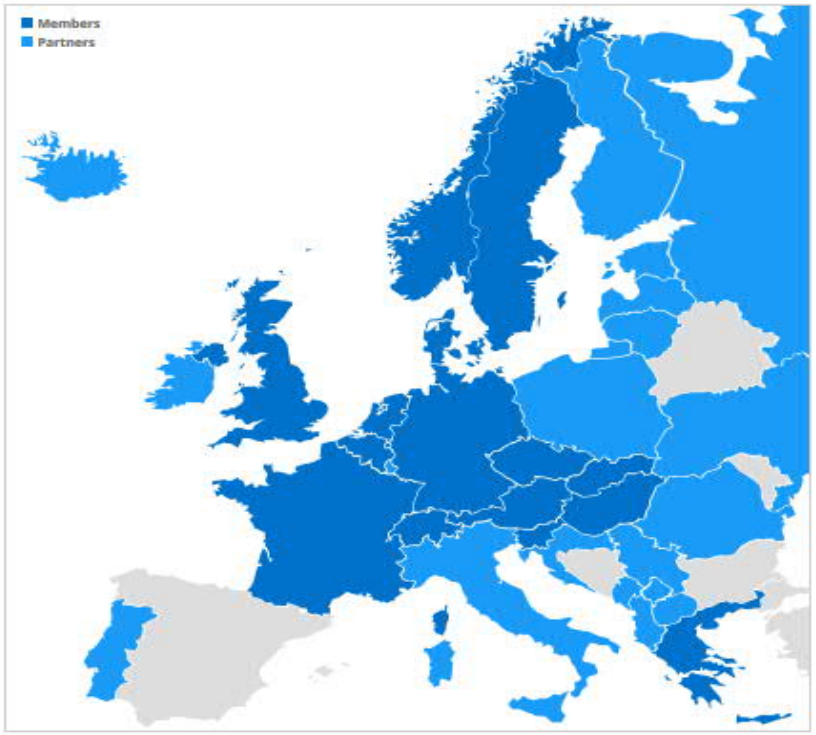
gesis Leibniz Institute for the Social Sciences

- Member of CESSDA
- Research infrastructure that covers services across the research lifecycle
- Hosts the Data Archive for the Social Sciences
- About 6.000 studies and 3 research data centers

cessda eric

- Consortium of European Social Science Data Archives
- Brings together Social Science Data Archives across Europe
- Supports and promotes results of Social Science research

cessda eric



- Austria
- Belgium
- Czech Republic
- Denmark
- France
- Finland
- Germany
- Greece
- Hungary
- Netherlands
- Norway
- Portugal
- Slovakia
- Slovenia
- Sweden
- Switzerland
- UK

gesis Leibniz Institute for the Social Sciences

cessda eric

FOSTER

- **F**acilitate **O**pen **S**cience **T**raining for **E**uropean **R**esearch
- Training support network with 15 partner countries
- Supports culture change towards OS
- Delivers f2f training, blended, e-learning
- Provides Open Science resources
- European Union’s Horizon 2020 ID: 741

gesis Leibniz Institute for the Social Sciences

cessda eric

FOSTER

Open Licensing

Licensing your research outputs is an important part of practicing Open Science. In this course, you will:

- know what licenses are, how they work, and how to apply them&nbs...

Open Science and Innovation

This course helps you to understand open business models and responsible research and innovation (RRI) and illustrates how these can foster innovation. By the end of the course, you will:

...

Open Peer Review

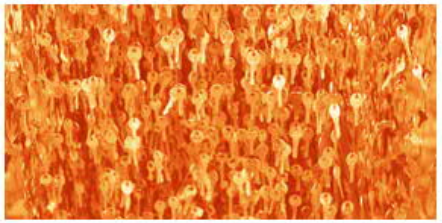
This course introduces you to open peer review (OPR), an emerging practice which is gaining momentum as part of Open Science. Upon completing this course, you will:

- understand ...

Sharing preprints

This course shows you how sharing preprints can improve your research and support Open Science. By the end of the course, you will:

- know what preprints are
- be able t...



Open Access Publishing

This course helps you to become skilled in Open Access (OA) publishing in the context of Open Science. By the end of the course, you will:

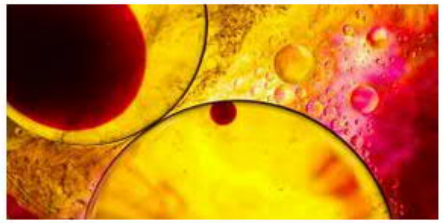
- understand how to publish your work o...

Data Protection and Ethics

This course covers data protection in particular and ethics more generally. It will help you understand the basic principles of data protection and introduces techniques for implementing data pr...

Open Source Software and Workflows

This course introduces Open Source Software (OSS) management and workflow as an emerging but critical component of Open Science. The course explains the role of software sharing and sustainabili...



Managing and Sharing Research Data

Data-driven research is becoming increasingly common in a wide range of academic disciplines, from Archaeology to Zoology, and spanning Arts and Science subject areas alike. To support good rese...

https://www.fosteropenscience.eu/courses

gesis Leibniz Institute for the Social Sciences

cessda eric

- Brings together leading European Research Infrastructures
- Addresses key challenges for cross-national data collection
- Provides specific training on products developed in SERISS
- European Union's Horizon 2020 ID: 654221

gesis Leibniz Institute for the Social Sciences

cessda eric

seriss SYNERGIES FOR EUROPE'S RESEARCH INFRASTRUCTURES IN THE SOCIAL SCIENCES

- Deliverables
  - WP2: Representing the population
  - WP3: Maximising equivalence through translation
  - WP4: Interactive tools for cross-national surveys
  - WP5: Training and Dissemination
  - WP6: New forms of data – legal, ethical and quality issues
  - WP7: A survey future online
  - WP8: A coding module for socio-economic survey questions

gesis Leibniz Institute for the Social Sciences

cessda eric

Who we are:

Anja Perry
(Open Data and data discovery)

Kristi Winters
(FAIR and data harmonization)

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

Organizational matters

gesis Leibniz Institute for the Social Sciences

cessda eric

Agenda

Day 1

1. Introduction to Open Science
2. What is Open Data?
3. How to produce FAIR Data?

Day 2

4. How to discover data?
5. How to harmonize data?
6. How to store data
7. Wrap Up

And …
… who are you???
… where are you from???
??
… what do you do???
… what is your interest in Open Science???
Image: www.pixabay.com (CC-0).

Source: Jennifer Jacquemart (CC-BY)
https://www.flickr.com/photos/eusocialmediadude/37319355556/in/dateposted-public/

"… there is a revolution happening in the way science works. Every part of the scientific method is nowadays becoming an open, collaborative and participative process."

Carlos Moedas, October 12th, 2015

gesis Leibniz Institute for the Social Sciences

cessda eric

Discussion

Exercise Booklet:
Open Science and Open Data,
August 20 & 21, 2017, GESIS Cologne

gesis Leibniz Institute for the Social Sciences

Exercise 1: What is Open Science?

- Work in 2-4 groups
- Time: about 20 minutes
- At the end, we will discuss your results with the class.

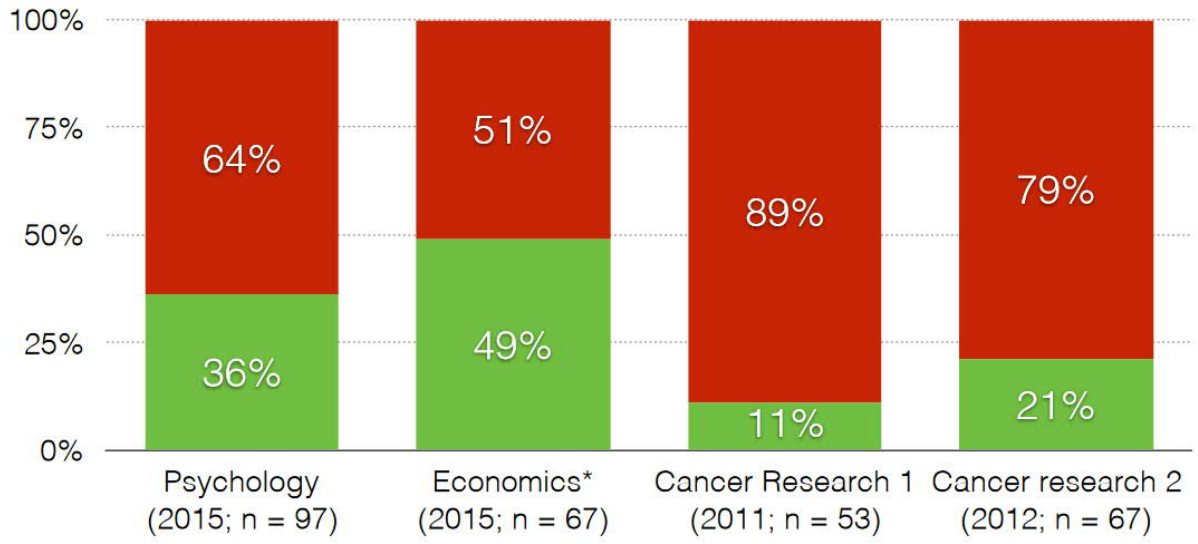
Discuss the following topics in your group:

- What is your impression of Open Science?
- What are the trends you see in Open Science?
- What are your concerns about Open Science?

What is Open Science?

Word cloud prepared by Cord Wiljes

Replication of published research findings



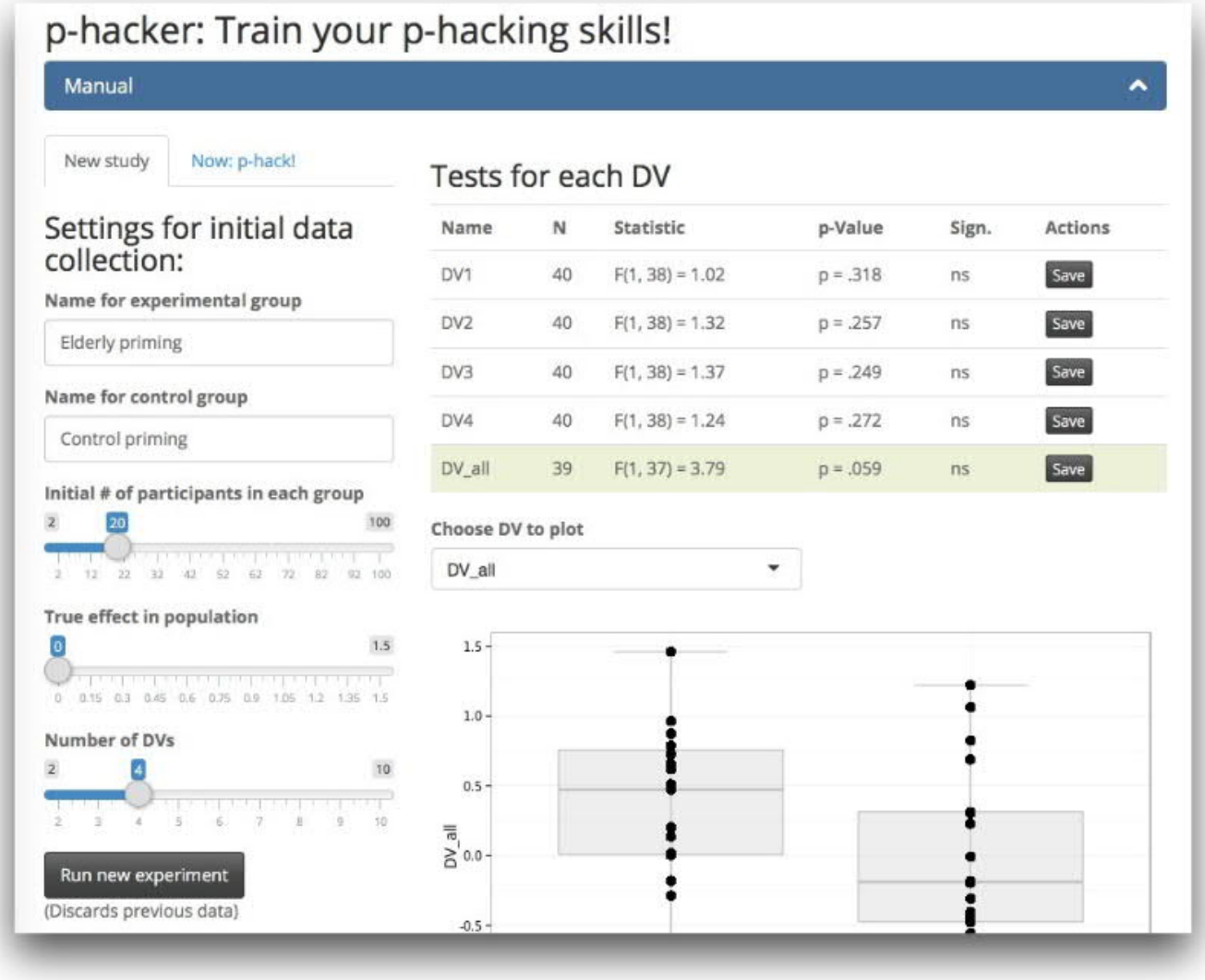
*The data on economics is about *reproducibility*; i.e. the attempt to get the same results if you apply the original data analysis on the original data set.

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

Open Access to Research Data? Why?

Image: http://ottersandsciencenews.blogspot.ca/2017/03/fake-science-how-sting-operation.html

p-hacking



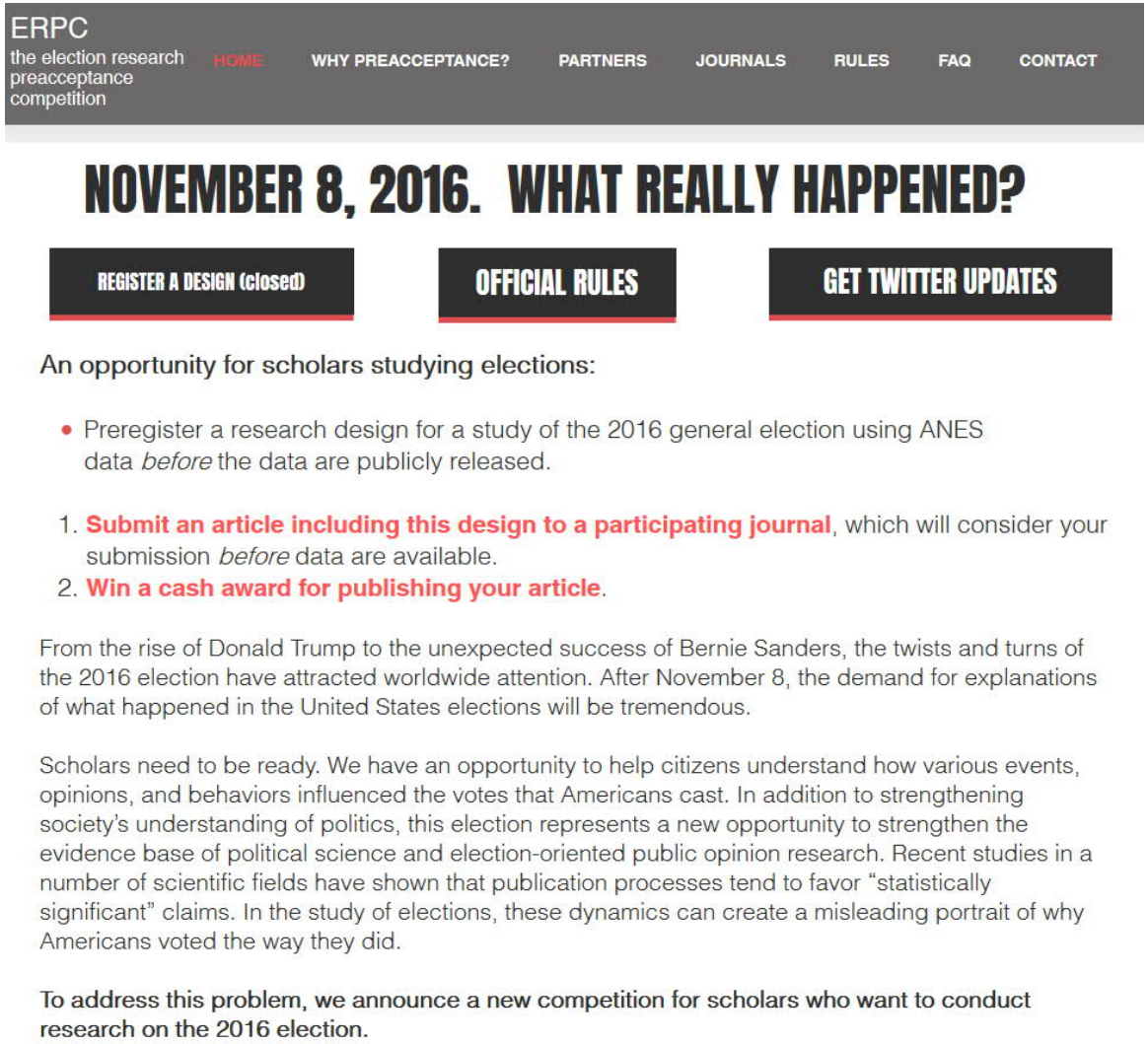
http://shinyapps.org/apps/p-hacker/
Felix Schönbrodt (2017). *p*-hacking: What it is, how to prevent it

Researchers are not rewarded for being right, but rather for publishing a lot.

Nelson, Simmons, & Simonsohn (2012); Nosek, Spies, Motyl (2012); Munafo (2016)

What about Pre-Registration?

Example for secondary research



ERPC
the election research preacceptance competition

HOME | WHY PREACCEPTANCE? | PARTNERS | JOURNALS | RULES | FAQ | CONTACT

NOVEMBER 8, 2016. WHAT REALLY HAPPENED?

REGISTER A DESIGN (closed) | OFFICIAL RULES | GET TWITTER UPDATES

An opportunity for scholars studying elections:

- Preregister a research design for a study of the 2016 general election using ANES data *before* the data are publicly released.

1. **Submit an article including this design to a participating journal**, which will consider your submission *before* data are available.
2. **Win a cash award for publishing your article**.

From the rise of Donald Trump to the unexpected success of Bernie Sanders, the twists and turns of the 2016 election have attracted worldwide attention. After November 8, the demand for explanations of what happened in the United States elections will be tremendous.

Scholars need to be ready. We have an opportunity to help citizens understand how various events, opinions, and behaviors influenced the votes that Americans cast. In addition to strengthening society's understanding of politics, this election represents a new opportunity to strengthen the evidence base of political science and election-oriented public opinion research. Recent studies in a number of scientific fields have shown that publication processes tend to favor "statistically significant" claims. In the study of elections, these dynamics can create a misleading portrait of why Americans voted the way they did.

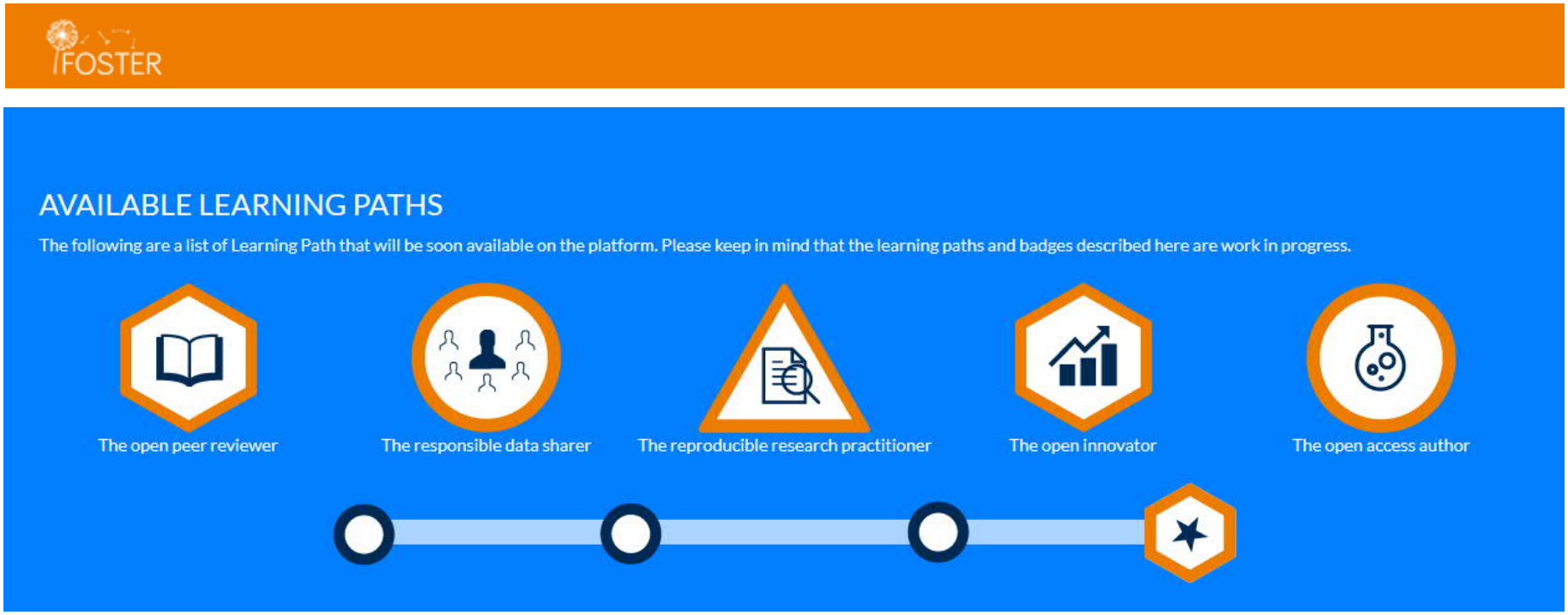
To address this problem, we announce a new competition for scholars who want to conduct research on the 2016 election.

https://www.erpc2016.com/

gesis Leibniz Institute for the Social Sciences

cessda eric

Dive into the Open Science concepts



https://www.fosteropenscience.eu/

gesis Leibniz Institute for the Social Sciences

cessda eric

Open Science MOOC

About Modules Blog Resources Contact People

OPEN SCIENCE MOOC

FREE | OPEN | LEARNING

Welcome to the home of the Open Science MOOC!

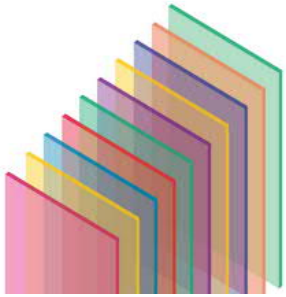
This website is aimed to provide information about our MOOC on Open Science principles and practices, its rationale, the current state of the project, and the people behind it.

CURRENT STATUS: **IN DEVELOPMENT**

The first draft of Module 5 has been completed and awaiting feedback from the wider community (including you!)

This project was started in early 2017 after a barcamp at the Open Science Conference in Berlin. Soon, more than 30 people contributed and a first draft curriculum was made. Now, already more than 100 volunteers have agreed to share their knowledge about Open Science and to contribute to what they see as an extremely important issue in nowadays and future science. Concomitantly, the European Commission published its report Providing researchers with the skills and competencies they need to practice Open Science, supporting the importance of the topic and thereby the necessity to explain, teach and support researchers to gain the necessary skills.

We are excited by the support we got so far and we would like to invite everybody to create, comment, contribute and share!



OPEN SCIENCE MOOC

FREE | OPEN | LEARNING

OPEN ADVOCACY

Twitter

Follow @OpenScienceMOOC

Tweets by @OpenScienceMOOC

Open Science MOOC Retweeted

Egon Willigh@gen
@egonwillighagen

two weeks left to improve the Open Science Monitor

1h

Open Science MOOC Retweeted

J. Colomb, @pen
@j_colomb

hi twitterverse. I am looking for stories of data analysis facilitated or impeded by good/bad research data management #RDM #promote_rdm.

too long for a tweet, please check https: rdmpromotion.rbind.io/contribute to see how you can reach out.
(pict CC-BY, aukeherrema.nl)

https://opensciencemooc.github.io/site/

Further sources

Data Colada

Thinking about evidence and vice versa

02.06.18

by Uri Simonsohn

[69] Eight things I do to make my open research more findable and understandable

It is now common for researchers to post original materials, data, and/or code behind their published research. That's obviously great, but open research is often difficult to find and understand.

In this post I discuss 8 things I do, in my papers, code, and datafiles, to combat that.

Paper

1) Before all method sections, I include a paragraph overviewing the open research practices behind the paper. Like this:

Transparent Reporting

For each study we set sample size before any data were collected, always at 50+ per cell. We pre-registered Studies 2, 3, and 4a. The supplement [http://bit.ly/example] includes materials, data and code for all studies, and links to all pre-registrations. Because we conducted our studies on MTurk, we included several attention and comprehension checks based on which we carried out exclusions. All exclusions took place before random assignment, and we detail them in Supplement 1. The sample sizes presented below are for the net number of participants

http://datacolada.org/69

Further sources

Blog

Open Science

Research

Software

Talks & Workshops

Publications

About me (CV)

→ shinyapps.org

→ researchtransparency.org

Search the nicebread blog

Visit bloggers

Hiring Policy at the LMU Psychology Department: Better have some open science track record

June 25, 2018

In 2015, the psychology department at LMU Munich for the first time announced a professorship position with an "open science statement" (see original job description here):

> *"Our department embraces the values of open science and strives for replicable and reproducible research. For this goal we support transparent research with open data, open materials, and study pre-registration. Candidates are asked to describe in what way they already pursued and plan to pursue these goals."*

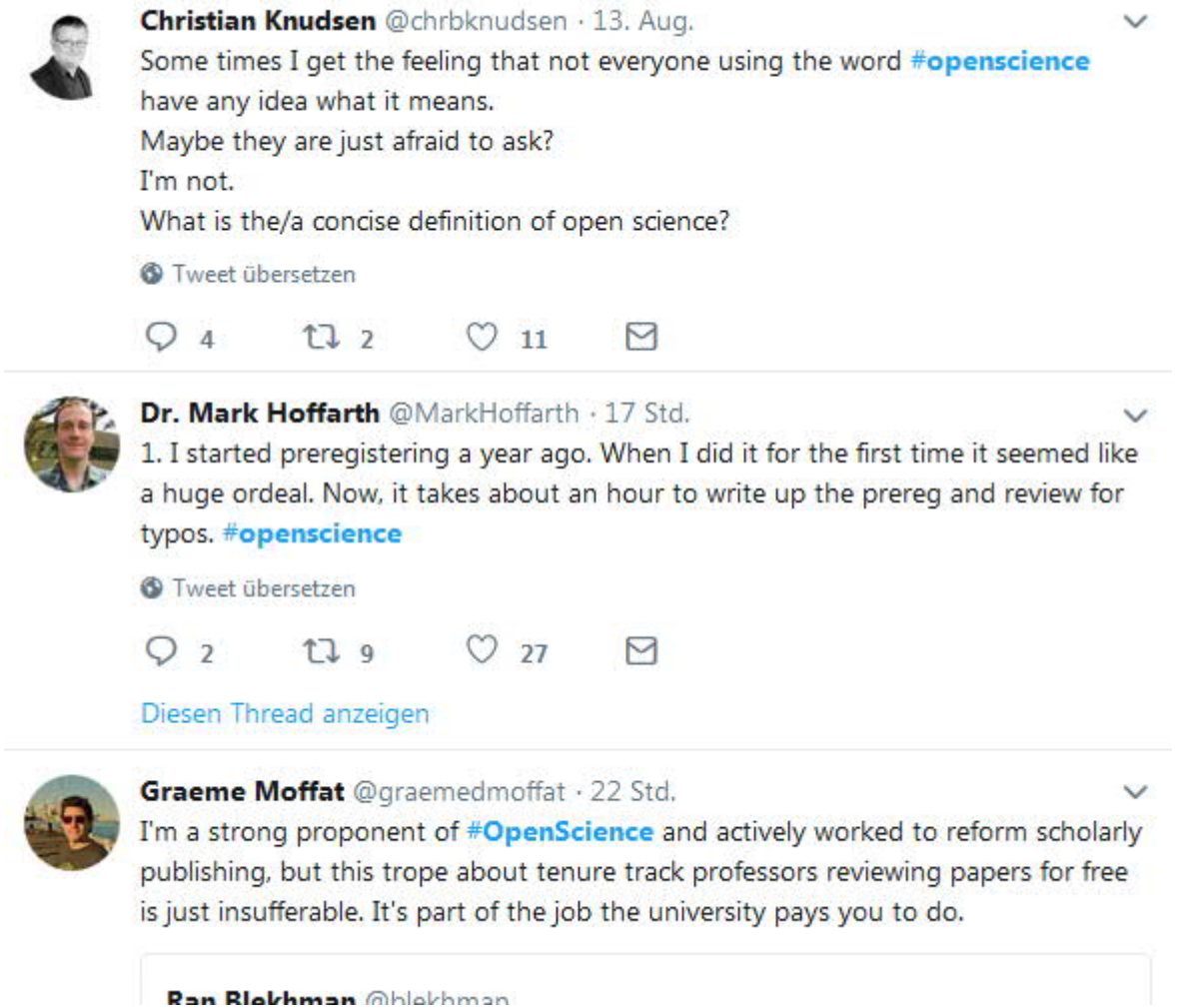
Since then, every professorship announcement contained this paragraph (and we made good experiences with it).

I am very happy to announce that my department now turned this implicit policy into an explicit hiring policy, effective since May 2018: **The department's steering committee unanimously voted for an explicit policy to always include this (or a similar) statement to all future professorship job advertisements.**

It is the task of the appointment committee to value the existing open science activities as well as future commitments of applicants appropriately. By including this statement, our department aims to communicate

http://www.nicebread.de/

Further sources



Christian Knudsen @chrbknudsen · 13. Aug.

Some times I get the feeling that not everyone using the word **#openscience** have any idea what it means.
Maybe they are just afraid to ask?
I'm not.
What is the/a concise definition of open science?

Tweet übersetzen

4 2 11

Dr. Mark Hoffarth @MarkHoffarth · 17 Std.

1. I started preregistering a year ago. When I did it for the first time it seemed like a huge ordeal. Now, it takes about an hour to write up the prereg and review for typos. **#openscience**

Tweet übersetzen

2 9 27

Diesen Thread anzeigen

Graeme Moffat @graemedmoffat · 22 Std.

I'm a strong proponent of **#OpenScience** and actively worked to reform scholarly publishing, but this trope about tenure track professors reviewing papers for free is just insufferable. It's part of the job the university pays you to do.

Ran Blekhman @blekhman

https://twitter.com/

2. WHAT IS OPEN DATA?

What is Open Data?

"Open Data are **online, free of cost, accessible data** that can be used, reused and distributed provided that the data source is attributed."

https://www.fosteropenscience.eu/taxonomy/term/6

What is Open Data?

"Openness means **access on equal terms for the international research community** at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based."

OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007

gesis Leibniz Institute for the Social Sciences

cessda eric

The Data Spectrum

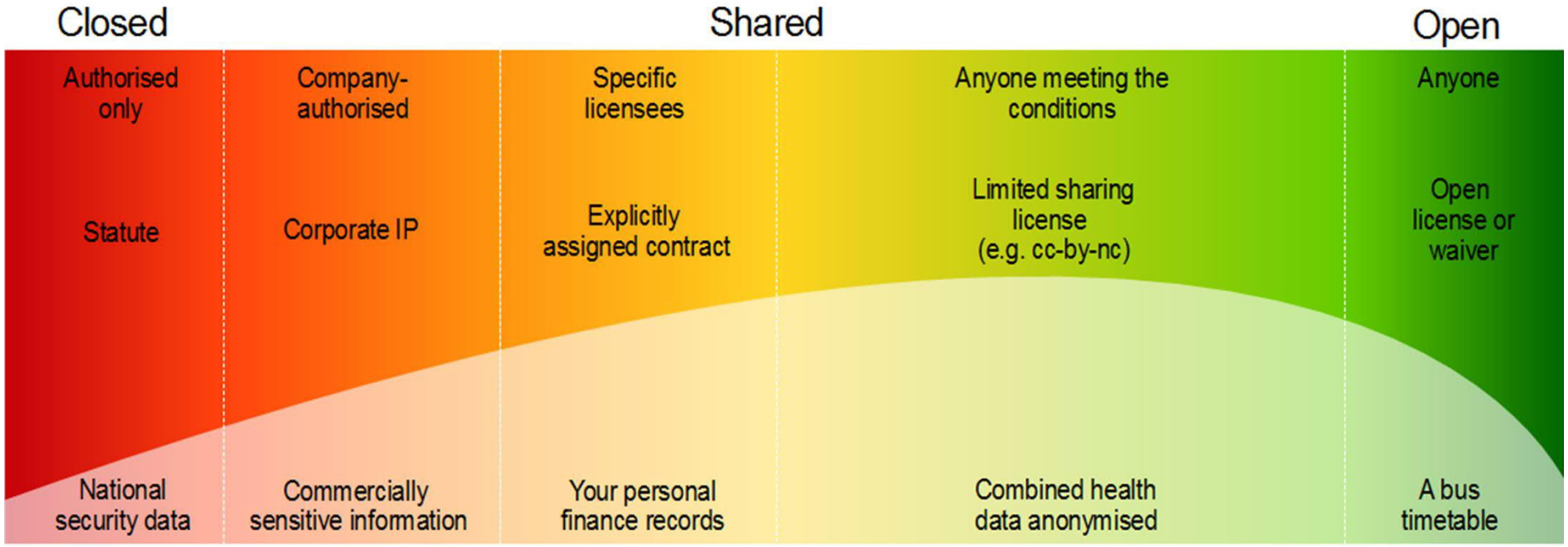
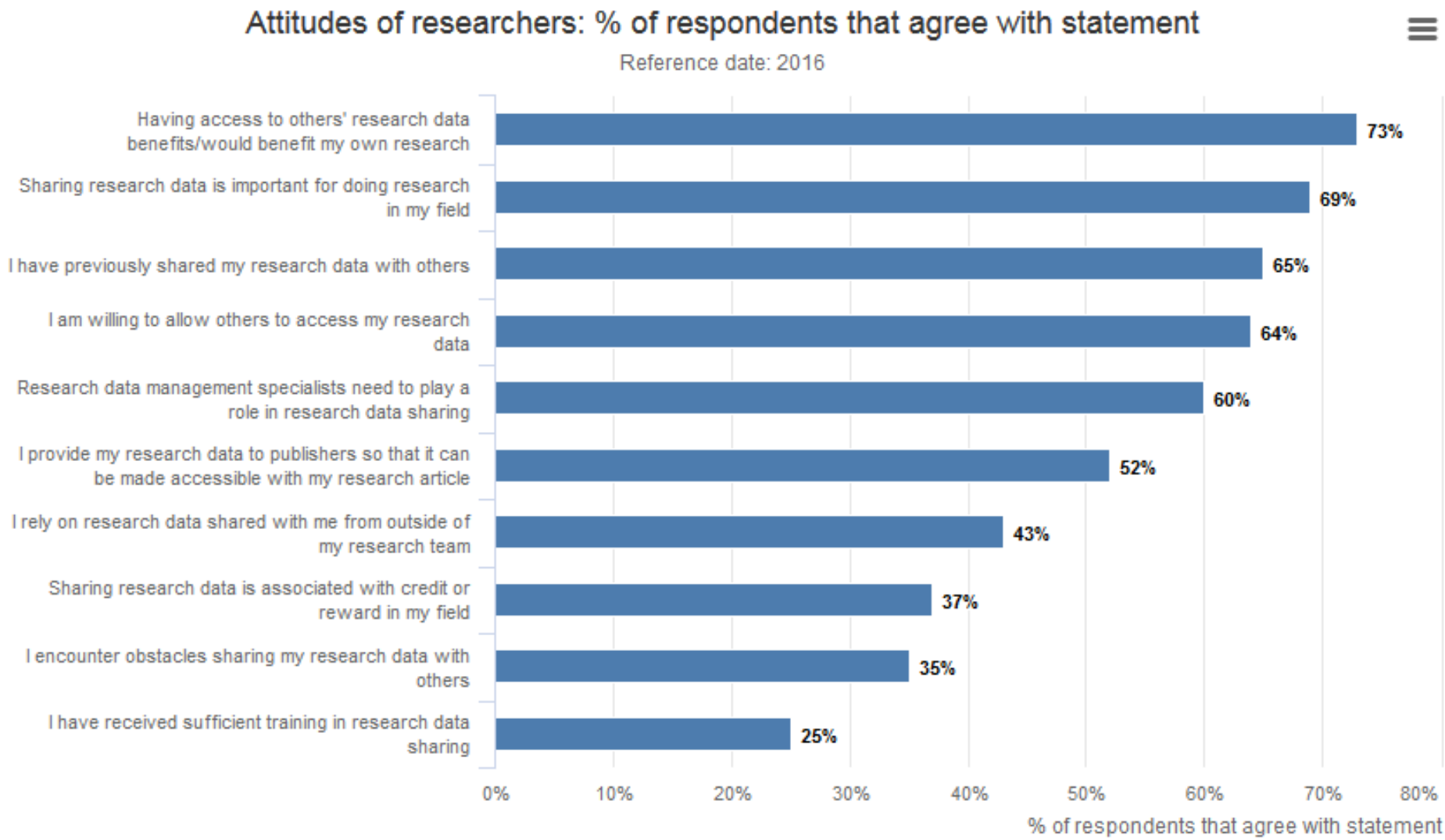
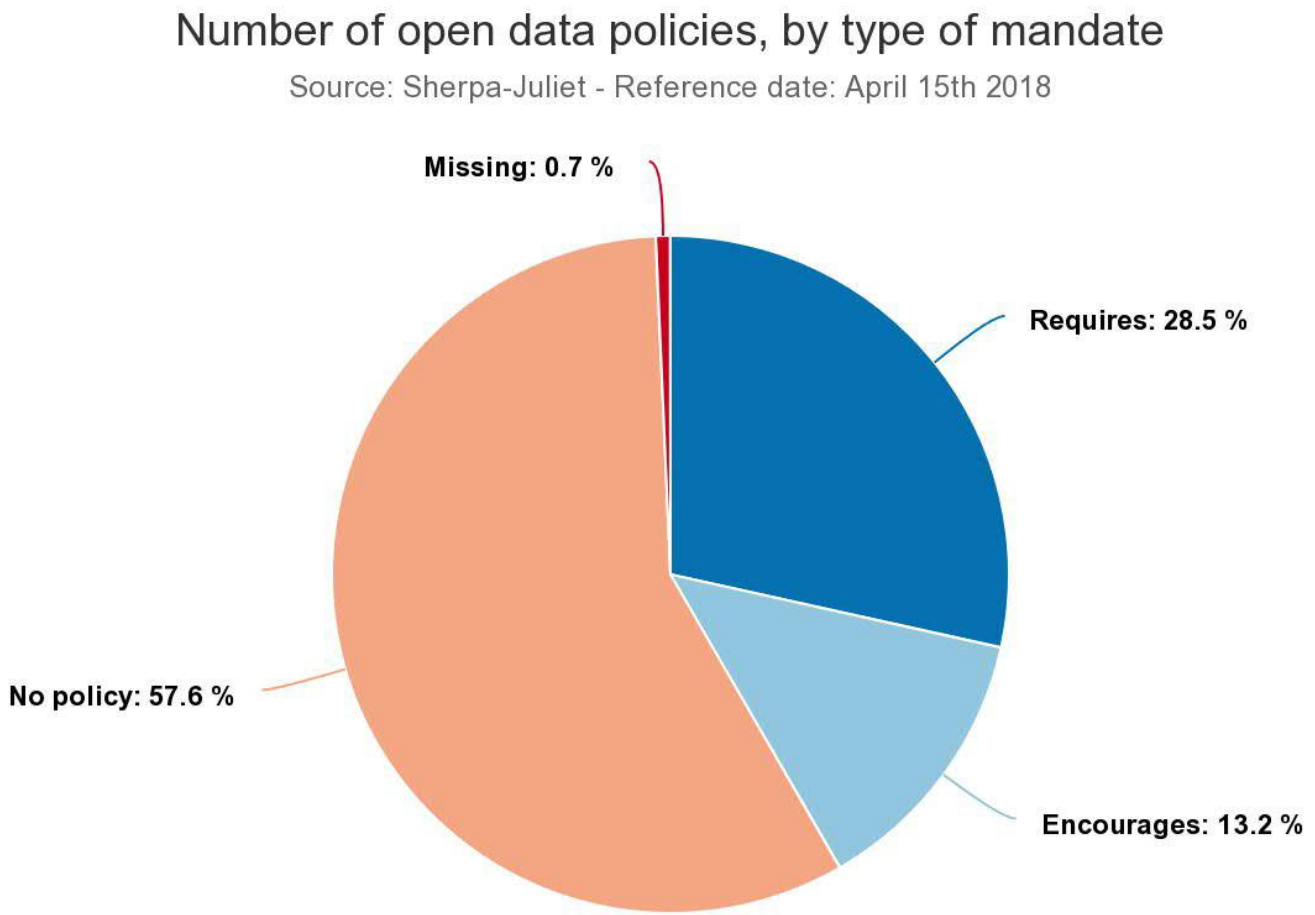


Image: “The Data Spectrum”. Adapted from http://theodi.org/data-spectrum (Open Data Institute, cc-by)



https://ec.europa.eu/info/open-science/open-science-monitor/facts-and-figures-open-research-data_en

Funders´ policies



https://ec.europa.eu/info/open-science/open-science-monitor/facts-and-figures-open-research-data_en

gesis Leibniz Institute for the Social Sciences

cessda eric

Open Access to Research Data? Why?

*Possibility "for third parties to **access**, **mine**, **exploit**, **reproduce** and **disseminate** (…) the data"*

- enabling reproducibility
- fostering research
- using public money efficiently
- boosting reputation

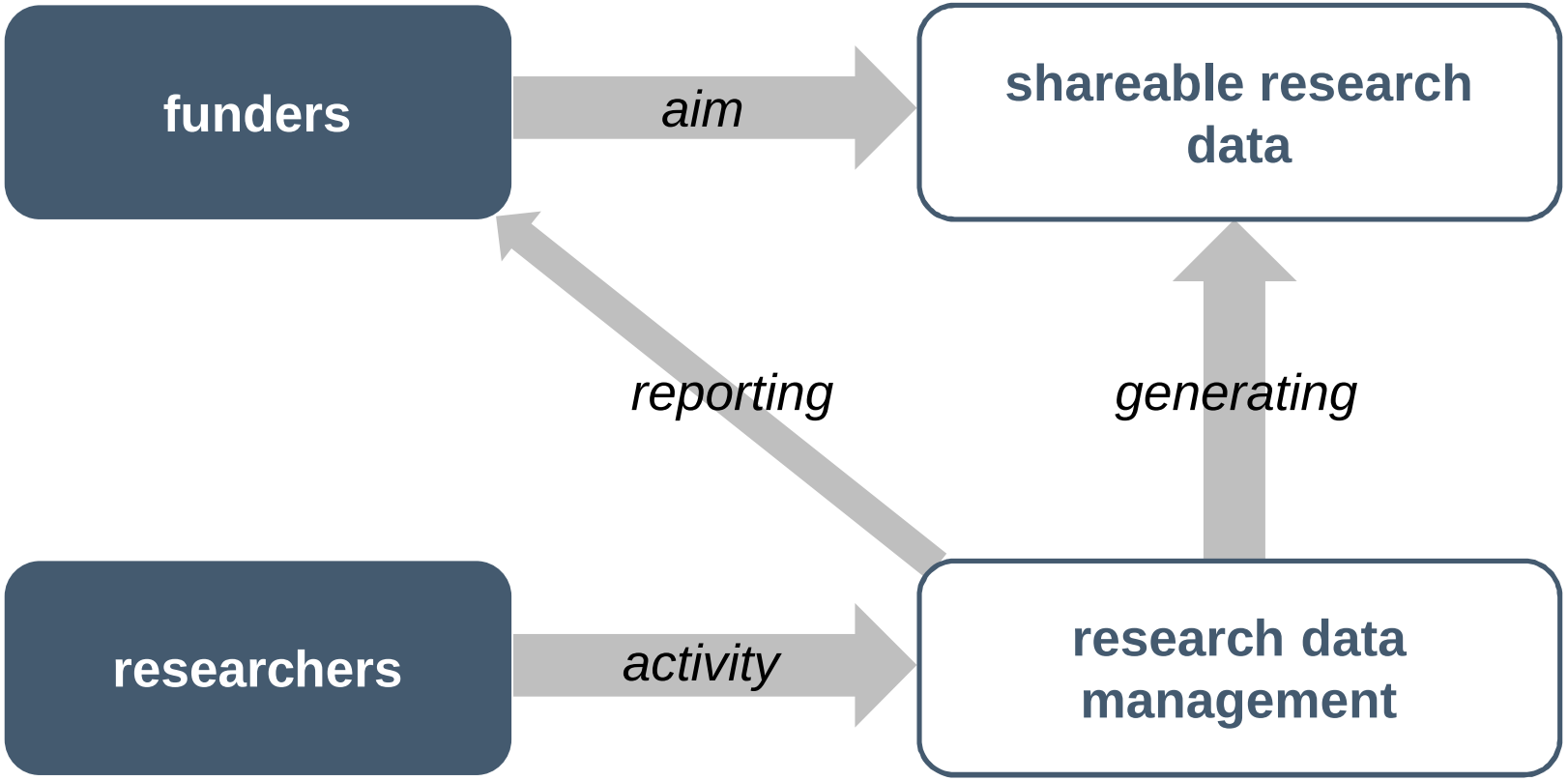
etc.

Horizon 2020: Grant Agreement, Art. 29.3.
Image: Herrema, A. (2014): FOSTER Cartoon: Data Sharing. FOSTER-Project (EU-Funding 612425). Available at: https://www.fosteropenscience.eu/foster-taxonomy/research-data-management. (CC-by).

gesis Leibniz Institute for the Social Sciences

cessda eric

Funders´ requirements



gesis Leibniz Institute for the Social Sciences

cessda eric

Horizon 2020: The FAIR-Principles

"*Horizon 2020 beneficiaries make their research data* ***findable, accessible, interoperable and reusable*** *(**FAIR**), to ensure it is soundly managed.*"

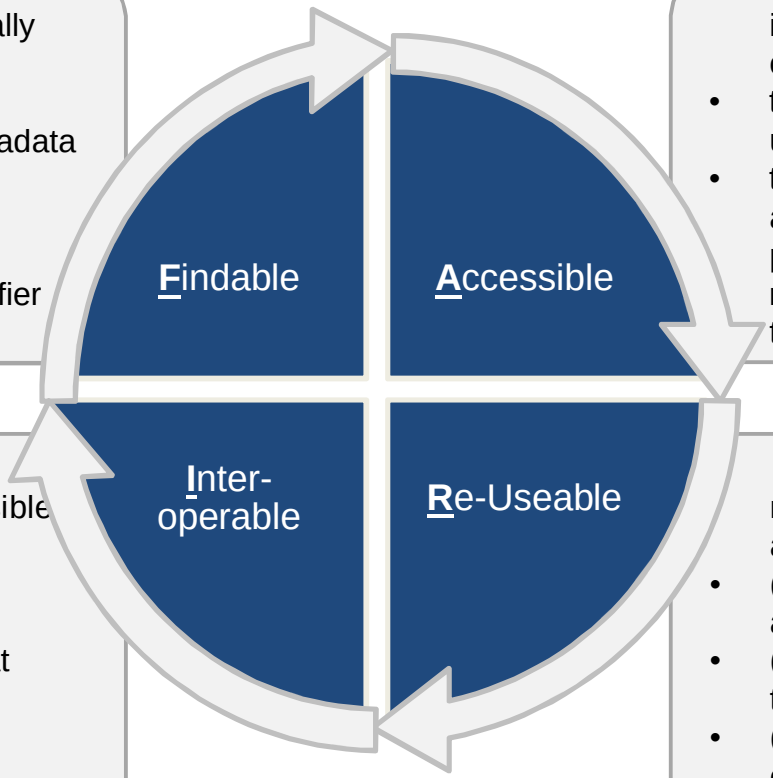
EU-Horizon 2020 Programme: Guidelines on FAIR Data Management in Horizon 2020.

gesis Leibniz Institute for the Social Sciences

cessda eric

The FAIR-Principles

- (meta)data are assigned a globally unique and eternally persistent identifier
- data are described with rich metadata
- (meta)data are registered or indexed in a searchable resource
- metadata specify the data identifier



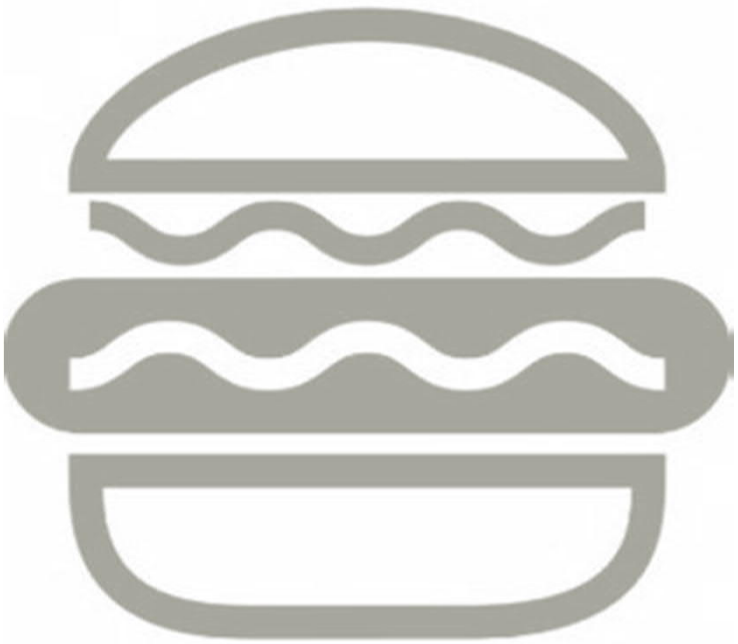
- (meta)data are retrievable by their identifier using a standardized communications protocol
- the protocol is open, free, and universally implementable
- the protocol allows for an authentication and authorization procedure, where necessary
metadata are accessible, even when the data are no longer available

- (meta)data use a formal, accessible shared, and broadly applicable language for knowledge representation
- (meta)data use vocabularies that follow FAIR principles
- (meta)data include qualified references to other (meta)data

meta(data) have a plurality of accurate and relevant attributes

- (meta)data are released with a clear and accessible data usage license
- (meta)data are associated with their provenance
- (meta)data meet domain-relevant community standards

www.force11.org/group/fairgroup/fairprinciples.

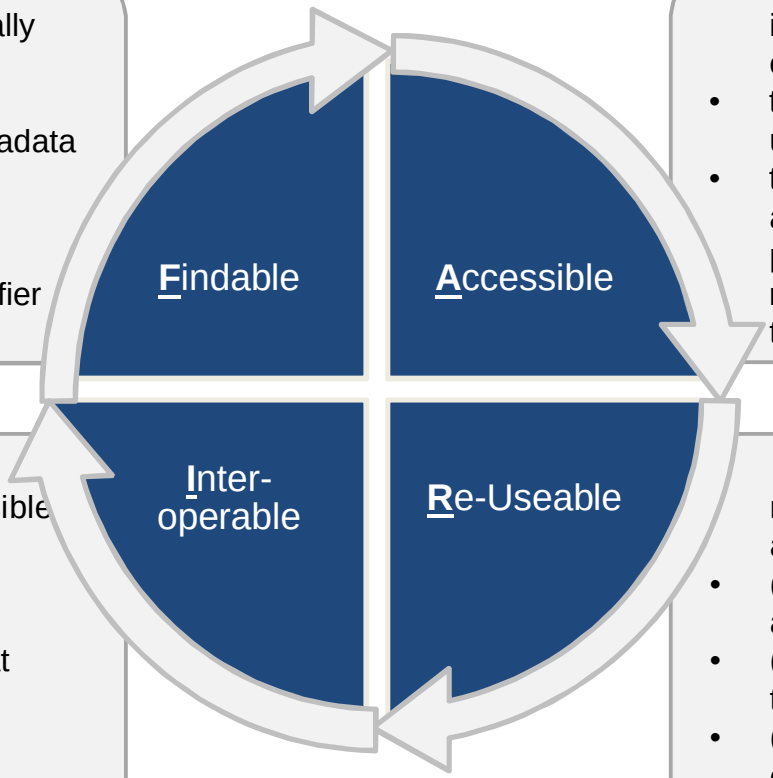


gesis Leibniz Institute for the Social Sciences

cessda eric

The FAIR-Principles

- (meta)data are assigned a globally unique and eternally persistent identifier
- data are described with rich metadata
- (meta)data are registered or indexed in a searchable resource
- metadata specify the data identifier



- (meta)data are retrievable by their identifier using a standardized communications protocol
- the protocol is open, free, and universally implementable
- the protocol allows for an authentication and authorization procedure, where necessary metadata are accessible, even when the data are no longer available

- (meta)data use a formal, accessible shared, and broadly applicable language for knowledge representation
- (meta)data use vocabularies that follow FAIR principles
- (meta)data include qualified references to other (meta)data

meta(data) have a plurality of accurate and relevant attributes

- (meta)data are released with a clear and accessible data usage license
- (meta)data are associated with their provenance
- (meta)data meet domain-relevant community standards

an archive´s perspective, but …

www.force11.org/group/fairgroup/fairprinciples.

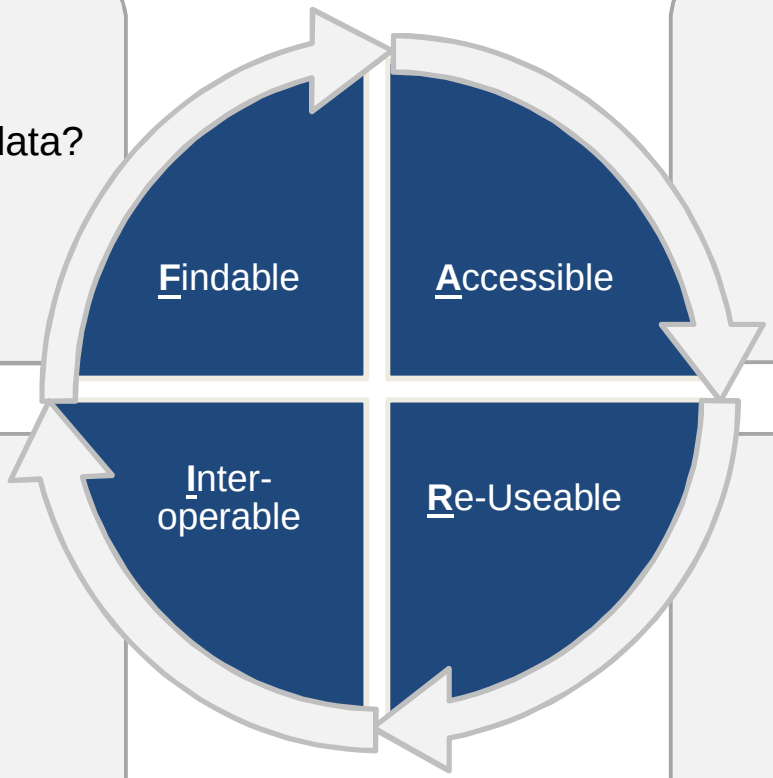
gesis Leibniz Institute for the Social Sciences

cessda eric

… FAIR is FAIR

How do others discover the data?

How do others get access to the data?



Can others technically re-use the data?

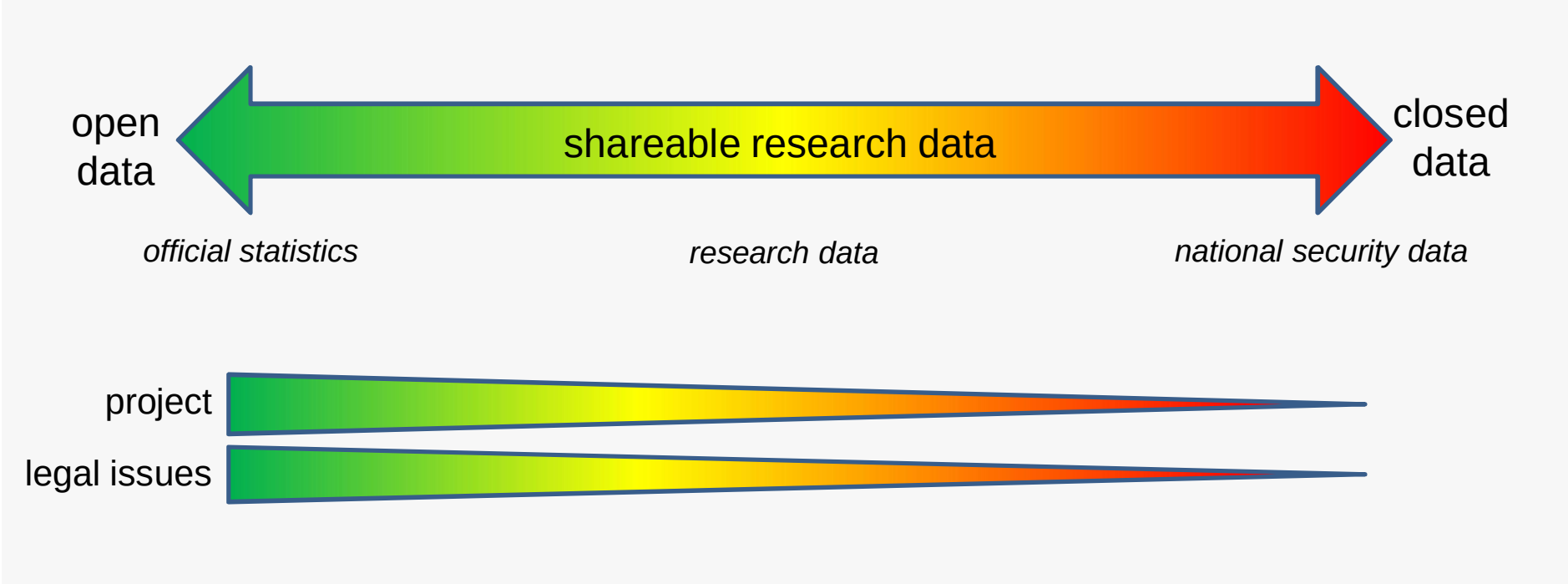
Can others analytically re-use the data?

3. HOW TO PRODUCE FAIR DATA

gesis Leibniz Institute for the Social Sciences

cessda eric

Accessibility and Openness of Research Data



⇒ "*as open as possible, as closed as necessary*"

EU-Horizon 2020 Programme: Guidelines on FAIR Data Management in Horizon 2020.

‘FAIR National Election Studies: How Well Are We Doing?’

- FAIR National Election Studies: How Well Are We Doing? Eder and Jedinger (2018)
- Researchers depend on data documentation quality, data quality is the foundation of credible statistical analyses and conclusion.
- Faniel et al. (2016) found that data re-users’ satisfaction is positively related to documentation quality.

Eder, C. & Jedinger, A. (2018). ‚FAIR national election studies: How well are we doing?’ European Political Science. https://link.springer.com/article/10.1057/s41304-018-0194-3

‘FAIR National Election Studies: How Well Are We Doing?’

- Data quality “cannot be assessed independent of … data consumers” (Strong et al., 1997: 104).
- *FAIR Guiding Principles for Scientific Data Management and Stewardship* (Wilkinson et al., 2016).
- Used these principles to develop a scheme to assess election studies’ accessibility, data and documentation quality.

‘FAIR National Election Studies: How Well Are We Doing?’

- A prerequisite to conducting research is an open and transparent data landscape
- “Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based” (OECD, 2007: 15).
- Transparency: “Information on research data and data-producing organisations, documentation on the data and specifications of conditions attached to the use of these data should be internationally available in a transparent way, ideally through the Internet.” (OECD, 2007: 15).

‘FAIR National Election Studies: How Well Are We Doing?’

- Effective and “open” exchange of data that meet the FAIR criteria:
  - Findability
  - Accessibility
  - Interoperability
  - Reusability of data (and documents).

‘FAIR National Election Studies: How Well Are We Doing?’

“…elements of the FAIR Principles are related, but independent and separable. The Principles define characteristics that contemporary data resources, tools, vocabularies and infrastructures should exhibit to assist discovery and reuse by third-parties.

By minimally defining each guiding principle, the barrier-to-entry for data producers, publishers and stewards who wish to make their data holdings FAIR is purposely maintained as low as possible.”

- Wilkinson et al. (2016: 4)

‘FAIR National Election Studies: How Well Are We Doing?’

- Eder and Jedinger “customized” the FAIR elements and adapted them according to the needs of
  - primary investigators carrying out election studies, and
  - secondary users employing these data for their research.

‘FAIR National Election Studies: How Well Are We Doing?’

- ***Findability***
- To be findable means data and descriptive metadata are easily discovered by humans and computers (Wilkinson et al., 2016: 4).
- By obtaining a permanent identifier such as a DOI or URN, election studies can be easily cited and referenced.

gesis Leibniz Institute for the Social Sciences

cessda eric

‘FAIR National Election Studies: How Well Are We Doing?’

Table 1: Findability criterion: concept and metrics

Criterion	Coding
FINDABILITY	
Provided via data archive?	0 = no, 1 = yes
Persistent identifier (DOI/URN)?	0 = no, 0.5 = archive specific, 1 = yes
Study ID available?	0 = no, 1 = yes
Versioning?	0 = no, 1 = yes
Are information on errata potentially available?	0 = no, 1 = yes
Findable through search engine?	0 = no, 1 = yes
Own webpage?	0 = no, 1 = yes
Is a citation recommendation provided?	0 = no, 1 = yes

‘FAIR National Election Studies: How Well Are We Doing?’

- **Accessibility**
- The ease with which data and descriptive metadata are available and downloadable by humans and computers.
- Data should be stored for the long term to allow stable access in the future.
- Can the datasets and their documentation both be downloaded without restrictions or is a registration required? If so, how quick and easy is it?

gesis Leibniz Institute for the Social Sciences

cessda eric

‘FAIR National Election Studies: How Well Are We Doing?’

Table 2: Accessibility criterion: concept and metrics

Criterion	Coding
ACCESSIBILITY	
Document access without registration?	0 = no, 1 = yes
Data access without registration?	0 = no, 1 = yes
Registration easy and free of charge?	0 = no, 1 = yes
Data download option?	0 = no, 1 = yes
Provision free of charge?	0 = no, 1 = yes
Are there variables excluded from the PUF?	0 = no, 1 = yes
Can they be obtained by the researcher?	0 = no, 1 = yes
Are data provided in major formats (Stata and SPSS)?	0 =no, 0.5 = partly, 1 = yes
Are questionnaires available in English?	0 = no, 1 = yes
Are study reports (field report, methods description, information on errata etc.) available in English?	0 = no,1 = yes
Is a codebook available in English?	0 = no, 1 = yes

‘FAIR National Election Studies: How Well Are We Doing?’

- ***Interoperability***
- How easily structured metadata on methodological aspects can be exchanged (Wilkinson 2016).
- Structured metadata means that methodological aspects of the study are prepared according to a standardized scheme.

‘FAIR National Election Studies: How Well Are We Doing?’

Table 3: Interoperability criterion: concept and metrics

Criterion	Coding
INTEROPERABILITY	
Information on primary investigators provided?	0 = no, 1 = yes
Information on funding provided?	0 = no, 1 = yes
Information on population provided?	0 = no, 1 = yes
Information on geographic coverage provided?	0 = no, 1 = yes
Information on target population provided?	0 = no, 1 = yes
Information on sampling frame provided?	0 = no, 1 = yes
Information on sampling procedure(s) provided?	0 = no, 1 = yes
Information on sample size provided?	0 = no, 1 = yes
Information on mode of data collection provided?	0 = no, 1 = yes
Information on date of data collection provided?	0 = no, 1 = yes
Information on data collector provided?	0 = no, 1 = yes
Is the study part of a larger comparative survey collection (across several countries)?	0 = no,1 = yes

‘FAIR National Election Studies: How Well Are We Doing?’

- **Reusable**
- Details on the study’s methodology should be well-described and the data error-free. Research data should include documentations such as:
  1. methodology reports,
  2. copies of questionnaires, and
  3. codebooks (see also Vardigan and Granda, 2010).
- A methodological report (aka field report/technical report) provides information about the
  - target population,
  - sampling design,
  - data collection modes,
  - response rates,
  - and weighting procedures of the survey.

‘FAIR National Election Studies: How Well Are We Doing?’

- Ideally, the codebook comprises an overview of all variables contained in a dataset, sorted by theme. A codebook lists
  - all variables and their values are named (variable names, variable labels and value labels),
  - filters
  - missing values are defined
  - comments on the variables are included.
  - weighted/unweighted marginal distributions and descriptive statistics.
  - derived and new variables calculated from other variables and their construction should be documented
- Special survey instruments and scales as well as their sources should be documented.

‘FAIR National Election Studies: How Well Are We Doing?’

- Paradata
  - interviewer observations,
  - interviewer characteristics,
  - aggregate data,
  - administrative data
- The questionnaire documentation
  - original questionnaire used in field work,
  - advance letters,
  - show cards,
  - consent forms, and
  - any other relevant material.

‘FAIR National Election Studies: How Well Are We Doing?’

- Regardless of the specific scheme employed, election study metadata should information on:
  - principal investigator(s),
  - geographic coverage,
  - target population,
  - sampling,
  - mode of data collection,
  - survey organization,
  - date(s) of collection,
  - and condensed information on survey content.

gesis Leibniz Institute for the Social Sciences

cessda eric

‘FAIR National Election Studies: How Well Are We Doing?’

Table 4: Reusability criterion: concept and metrics

Criterion	Coding
REUSEABILITY	
Are unique respondent numbers (id’s) available?	0 = no,1 = yes
Are the data free of wild codes and out-of-range values?	0 = no, 0.5 = partly, yes = no
Are variables labels assigned?	0 =no, 0.5 = partly, 1 = yes
Are value labels assigned?	0 =no, 0.5 = partly, 1 = yes
Are missing values defined?	0 =no, 0.5 = partly, 1 = yes
Are weighting factors documented?	0 =no, 0.5 = partly, 1 = yes, -1 = not applicable (no weights)

Discussion

Exercise Booklet:
Open Science and Open Data,
August 20 & 21, 2017, GESIS Cologne

gesis Leibniz Institute for the Social Sciences

Exercise 2: FAIR Data

- Work in 2-4 groups
- Time: about 20 minutes
- At the end, we will discuss your results with the class.

Discuss the following topic in your group:

Thinking about studies you worked with / created, what are your good or bad experiences with FAIR Data documentation?

CESSDA Expert Tour Guide

cessda eric About Consortium Projects Research Infrastructure Contact

Home / Research Infrastructure / Training / Expert Tour Guide on Data Management

Expert Tour Guide on Data Management

About this expert tour guide

This tour guide by CESSDA ERIC (the Consortium of European Social Science Data Archives European Infrastructure Consortium) aims to put social scientists like yourself at the heart of making their research data findable, understandable, sustainably accessible and reusable.

You will be guided by 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 17 CESSDA social science data archives. With this guide and the training events being held across Europe, we want to accompany and inspire you in your journey through the research data life cycle.

Expert Tour Guide on Data Management

1. Plan
2. Organise & Document
3. Process
4. Store
5. Protect
6. Archive & Publish

Search this guide Search

Target audience and mission

This tour guide was written for social science researchers who are in an early stage of practising research data management. With this tour guide, CESSDA wants to contribute to increased professionalism in data management and to improving the value of research data.

https://www.cessda.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management

CESSDA User Guide on RDM

CESSDA User Guide RDM

Name	Description	Type	Size
1 Research data management		PDF	109 KB
2 Research data management plan		PDF	103 KB
3 Data sharing		PDF	101 KB
4 Documentation and metadata		PDF	112 KB
5 File formats		PDF	104 KB
6 Data security		PDF	139 KB
7 Intellectual property rights		PDF	106 KB
8 Data consent and ethics		PDF	123 KB
9 Research data management check-list		PDF	99 KB

https://www.cessda.eu/Research-Infrastructure/Training/Research-Data-Management

DataWiz

https://datawiz.leibniz-psychology.org/DataWiz/

4. HOW TO DISCOVER DATA?

gesis Leibniz Institute for the Social Sciences

cessda eric

Types of data: quantitative and qualitative

Quantitative

- numeric data
  - numerical measurements
  - information coded into numerical data as part of the research process

Qualitative

- non-numeric information
  - interview transcripts, diaries, field notes, answers to open-ended survey, questions, audio-visual recordings and images

gesis Leibniz Institute for the Social Sciences

cessda eric

Types of data: level of analysis

Macro data

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

Meso data

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

Microdata

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

Types of data: time

Cross-sectional

- one-point of time (a snap shot)
- usually information on multiple cases and variables

Repeated cross sectional

- cross-sectional surveys repeated with new samples
- data from the different samples allows analysis of trends

Time series

- series of data points in time order (often equally spaced in time)
- aggregate macro data are often time-series data.
- time points may come from sample surveys e.g. unemployment from labour force surveys

Longitudinal

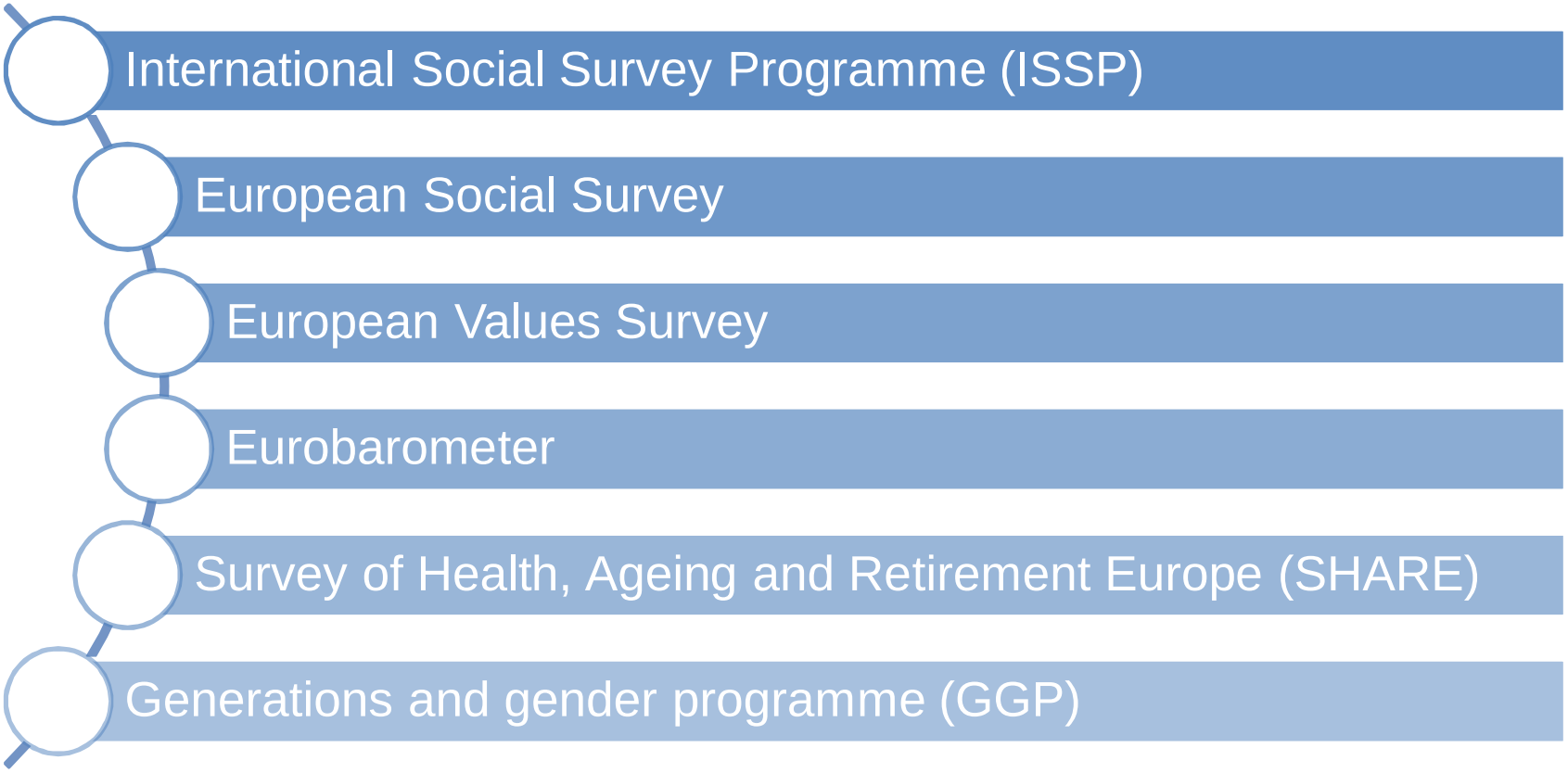
- follow the same units over time e.g. household panel studies collect information from a sample of households in regular ‘waves’

gesis Leibniz Institute for the Social Sciences

cessda eric

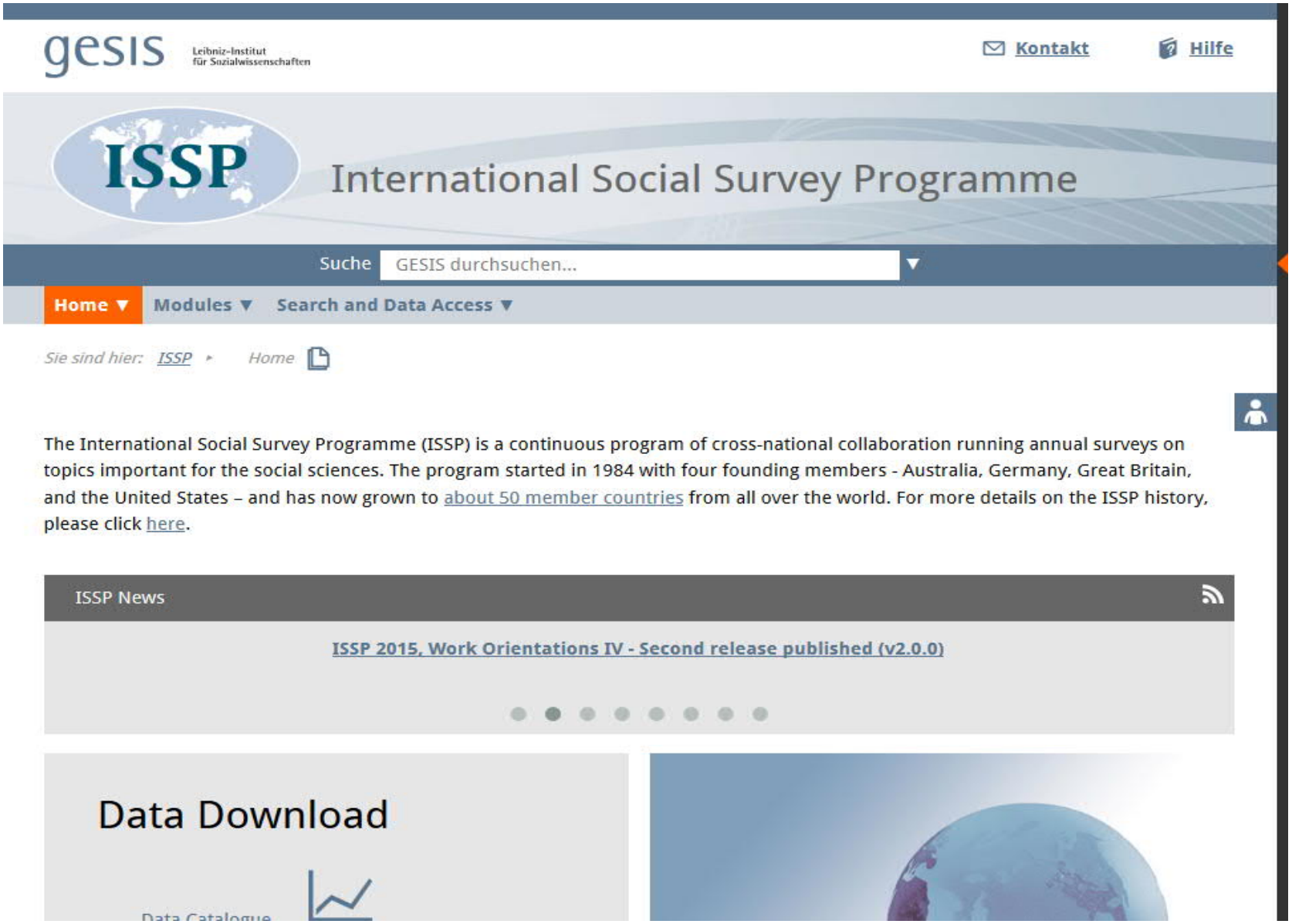
Cross-national studies

International survey research programmes include many European countries



International Social Survey Programme (ISSP)

- annual programme
- 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



https://www.gesis.org/issp/home/

gesis Leibniz Institute for the Social Sciences

cessda eric

Survey of Health, Ageing and Retirement Europe (SHARE)

- longitudinal study
- more than 123,000 individuals aged 50 and older
- 27 European countries and Israel
- health, socio-economic status and social and family networks

SHARE
Survey of Health, Ageing and Retirement in Europe
50+ in Europe

Survey of Health, Ageing and Retirement in Europe

Newsletter

Latest Newsletter

Subscribe to the SHARE newsletter by sending an e-mail to info@share-project.org

Twitter Facebook

SHARE - Survey of Health, Ageing and Retirement in Europe

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of approximately 123,000 individuals aged 50 or older (more than 293,000 interviews). SHARE covers 27 European countries and Israel.

The data are available to the entire research community free of charge. For a summary overview and research results of SHARE, you can download the SHARE brochure or our latest First Results Book: Ageing in Europe - Supporting Policies for an Inclusive Society.

Search

search

News

Wave 6 data released!

SHARE is very happy to announce the release of Wave 6 data! Wave 6 not only includes Croatia as a...

read more...

Launch of the Executive Master's Programme in Management of Research Infrastructures

RItrain, the Research Infrastructure Training Programme, is launching its Executive Master's...

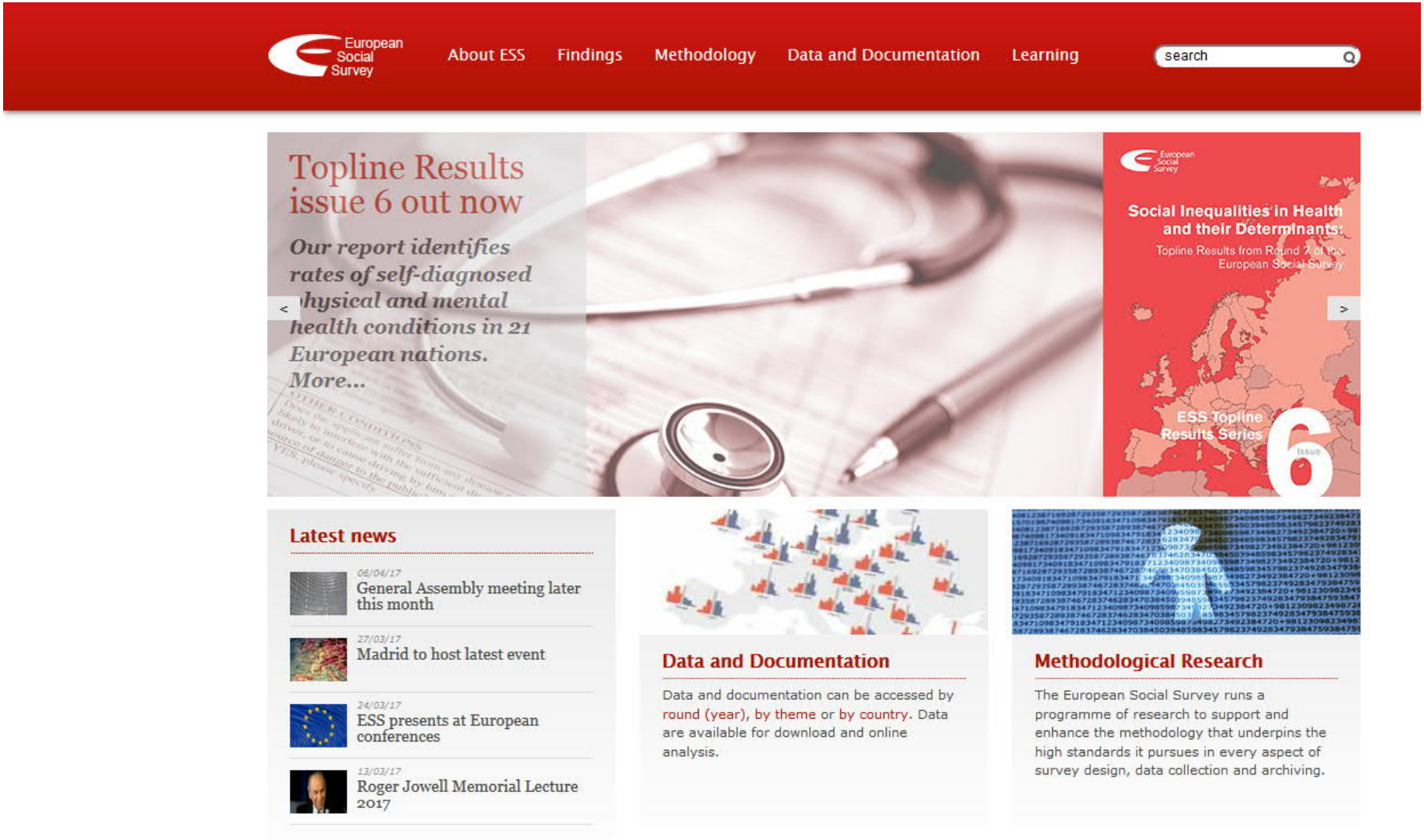
read more...

SHARE Newsletter No. 20

Please enjoy reading our Newsletter No. 20. In this edition, we present to you three of our new...

read more...

http://www.share-project.org/



http://www.europeansocialsurvey.org/

gesis Leibniz Institute for the Social Sciences

cessda eric

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

SOEP

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

Schweizer Haushalt-Panel
Panel suisse de ménages
Swiss Household Panel

gesis Leibniz Institute for the Social Sciences

cessda eric

Large Scale Assessments

https://www.gesis.org/en/en/piaac/piaac-home/

gesis Leibniz Institute for the Social Sciences

cessda eric

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 -

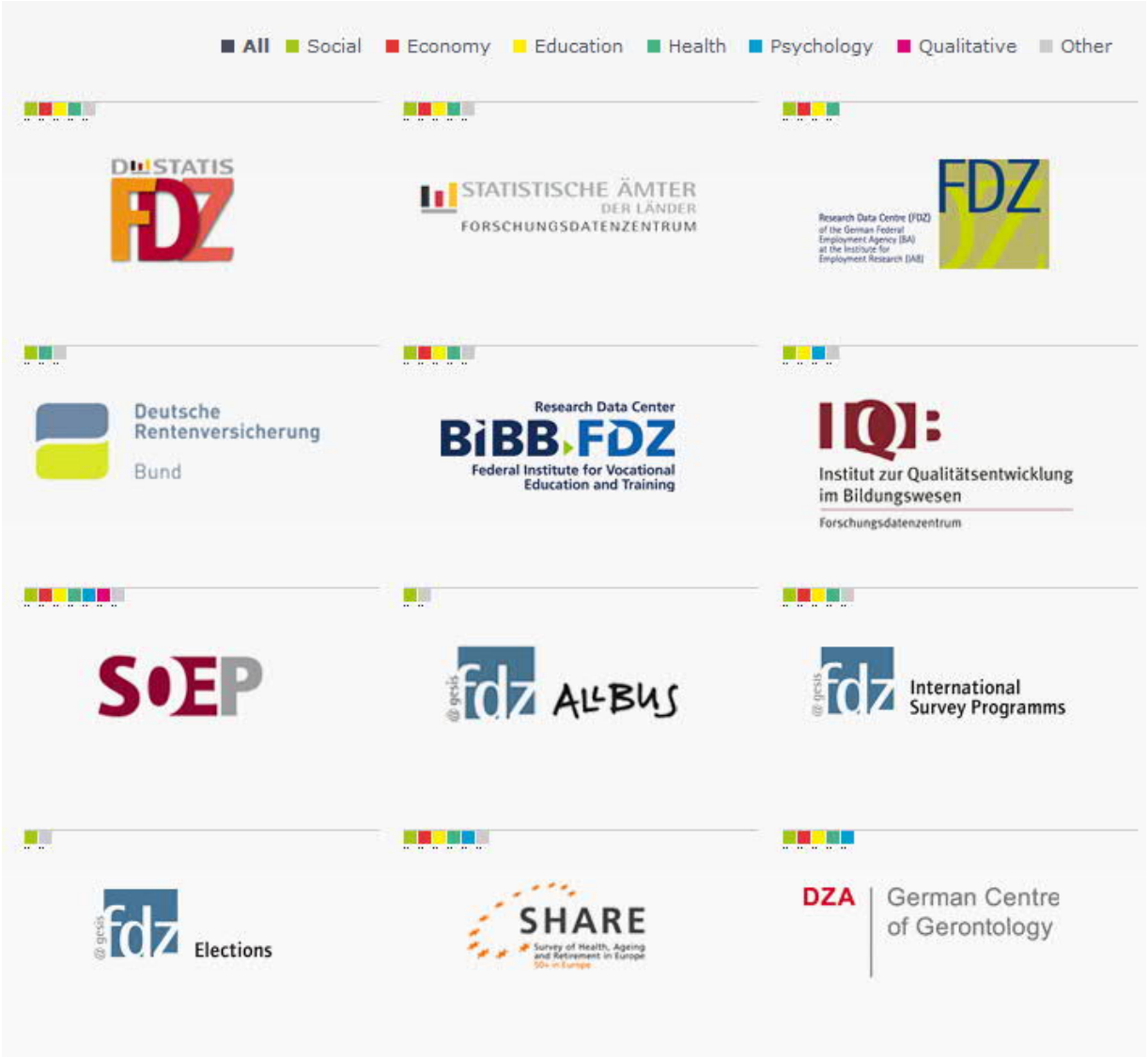
Institutional research data repositories

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

gesis Leibniz Institute for the Social Sciences

cessda eric

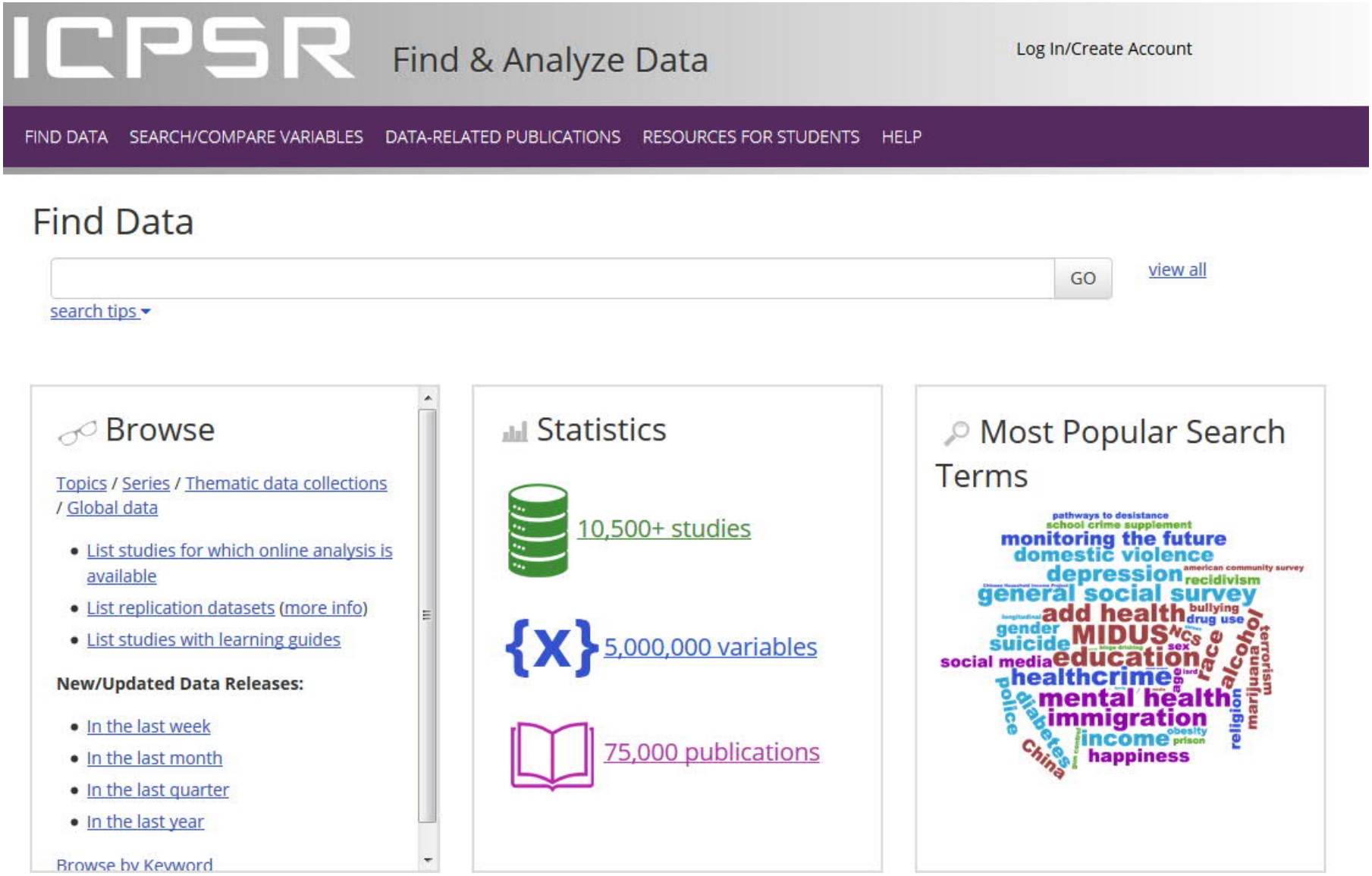
German Data Forum (RatSWD)



https://www.ratswd.de/en/data-infrastructure/rdc

gesis Leibniz Institute for the Social Sciences

cessda eric



https://www.icpsr.umich.edu/icpsrweb/ICPSR/

gesis Leibniz Institute for the Social Sciences

cessda eric

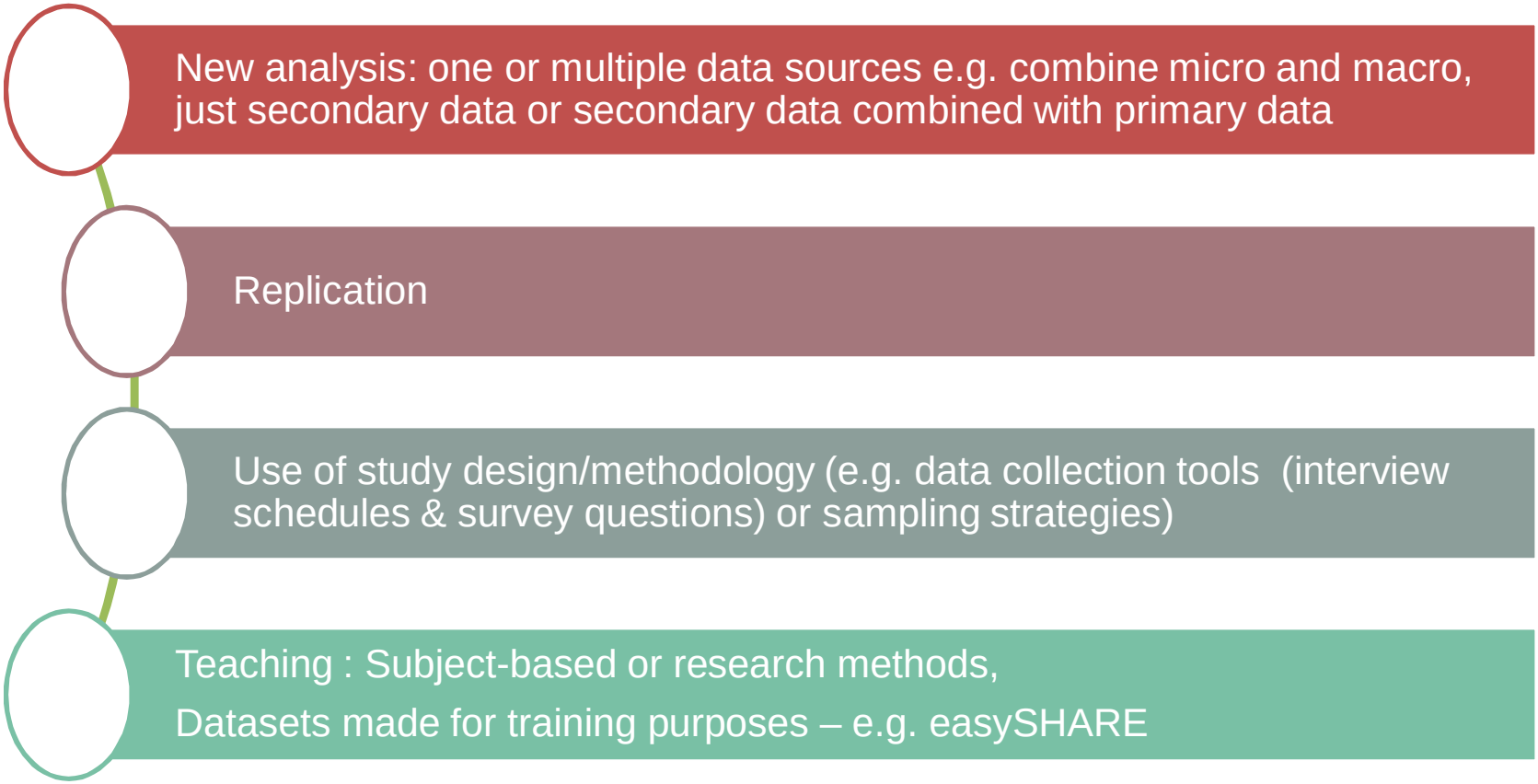
Four key data providing organisations

- Eurostat – Statistics office of European Union
- 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

gesis Leibniz Institute for the Social Sciences

cessda eric

Four ways we can use archived data



gesis Leibniz Institute for the Social Sciences

cessda eric

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)

gesis Leibniz Institute for the Social Sciences

cessda eric

Identifying data needs

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

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

- 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?

- Individuals, households, regions or countries?

- Do you need a large representative (random) sample?

gesis Leibniz Institute for the Social Sciences

cessda eric

Exercise

Exercise Booklet:
Open Science and Open Data,
August 20 & 21, 2017, GESIS Cologne

gesis Leibniz Institute for the Social Sciences

Exercise 3: Data Discovery

We discuss the task below with the class
Time: about 30 minutes

Rising voter support for parties blending populist and authoritarian appeals has disrupted mainstream party competition in many European societies – and had major consequences worldwide. What explains this phenomenon? We theorize that an important part of any explanation lies in perceived cultural threats, where rapid and profound value change in post-industrial societies during recent decades have affected core feelings of social identity, wrapped around values of family, faith and nation. These developments have generated a 'cultural backlash' activating authoritarian values and voting support for populist parties with authoritarian policy positions, especially among older and non-college educated citizens. (Norris and Inglehart 2017)

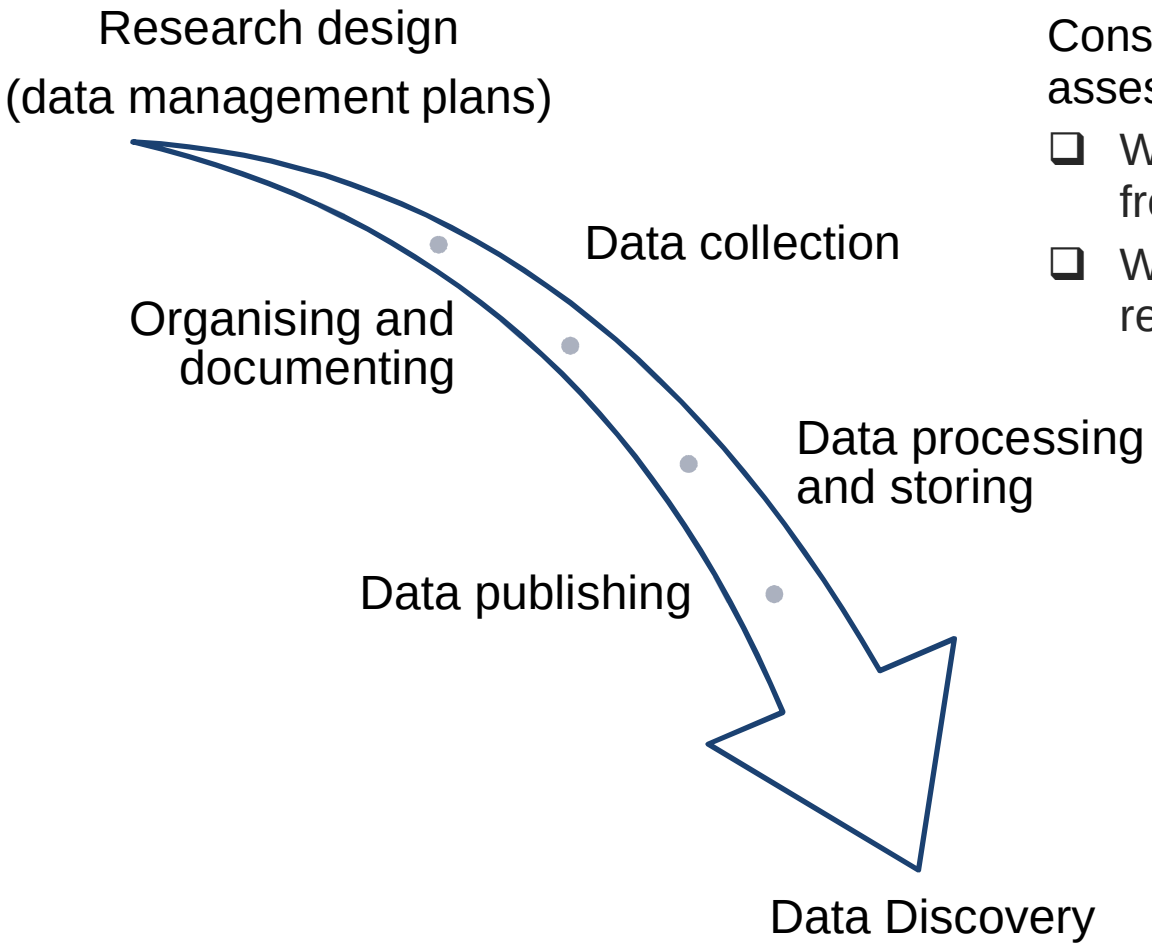
As the quote above shows, Norris and Inglehart propose a theory of a cultural backlash to explain increased support for politicians with populist-authoritarian policy positions.

Consider the bullet points below to identify data needs to explore theories of cultural backlash and/or this research question: "What attracts ordinary people to support parties emphasizing populist-authoritarian policy positions?"

gesis Leibniz Institute for the Social Sciences

cessda eric

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?

Metadata and documentation

CESSDA Training Working Group (2017)

- descriptors that facilitate cataloguing data and data discovery.

Documentation

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

gesis Leibniz Institute for the Social Sciences

cessda eric

Exercise

Exercise Booklet:
Open Science and Open Data,
August 20 & 21, 2017, GESIS Cologne

gesis Leibniz Institute for the Social Sciences

Exercise 4: Evaluate the data

- Work on your own
- Time: about 10 minutes
- Results will be discussed with all participants.

Use documents

- ISCO08_structure.pdf and/or
- ISCO08_definitions.pdf

in ILIAS or download the ISCO Classification Structure:
http://www.ilo.org/public/english/bureau/stat/isco/isco08/

Find a code for the profession stated in an answer to an open question in a survey:
"I work in a hospital as a nurse and I take care of newborns".

What to look for when assessing quality?

CESSDA Training Working Group (2017)

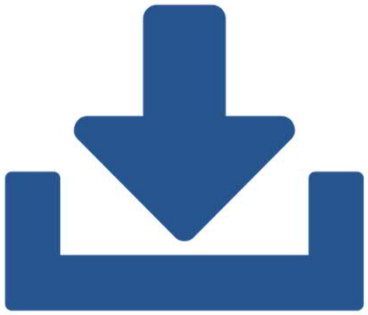
Can I establish

- 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?

gesis Leibniz Institute for the Social Sciences

cessda eric

Data access arrangements



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)

gesis Leibniz Institute for the Social Sciences

cessda eric

Data access arrangements

- 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.

Licensing

= to give permission to someone else to do something

Licenses define

- conditions of access
- conditions of (re-)use

But first:

- Don‘t just assume you are the owner of your data
- Clarify ownership

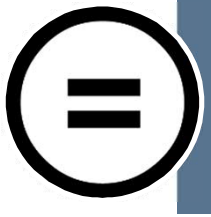
gesis Leibniz Institute for the Social Sciences

cessda eric

Creative Commons Licenses

Attribution (by): copy, distribute, display, perform and derivate work as long as the author or licensor is cited in the manner specified

Non-commercial (nc): copy, distribute, display, perform and derivate work as long as the work is used only for non-commercial purposes



No Derivative Works (nd): copy, distribute, display, perform but no derivative works based on it

Share-alike (sa): copy, distribute, display, perform and derivate work as long as the derivative works is licensed identical to the license that governs the original work

Images: www.creativecommons.org (CC-by).

gesis Leibniz Institute for the Social Sciences

cessda eric

Exercise

Exercise Booklet:
Open Science and Open Data,
August 20 & 21, 2017, GESIS Cologne

gesis Leibniz Institute for the Social Sciences

Exercise 5: Licenses and Data Re-Use

- Work in 2-4 groups
- Time: about 30 minutes
- Select one member of your group to briefly present the results of the discussion in the end. Results will be discussed with all participants.

In your research project you examine reasons for drop-outs at three universities in your country. You use data about individuals from various sources. The datasets include all relevant information and can be harmonized for your research project. However, they carry different licenses:

1. Dataset: students from university A, licensed under CC BY;
2. Dataset: students from university B, licensed under CC BY-SA;
3. Dataset: students from university C, licensed under CC BY-NC.

In addition, you received the following information form the universities' administration: number of enrolled students, the drop-out rate as well as the rate of students successfully completing their final degrees, the age structure and gender ratio at each of the three universities. These data are for the use in your research project exclusively. Further license agreements for any of this administrative data do not exist.

gesis Leibniz Institute for the Social Sciences

cessda eric

Getting started: using project-level and data-level documentation

What you might need to know?

- Folder structures and file names
- Version
- Case summaries
- Variables lists
- Exact question wording
- Questionnaire routing
- Manipulations to data and coding
- Weights
- Missing data
- Contextual details

Where to look?

- In a data file (e.g. summary of interview details before transcript, variable and value labels)
- Data lists (e.g. case summaries, variables lists)
- A user guide
- Fieldwork notes
- Technical report
- Readme files/version notes

gesis Leibniz Institute for the Social Sciences

cessda eric

And finally...remember to cite data

Why?

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

How?

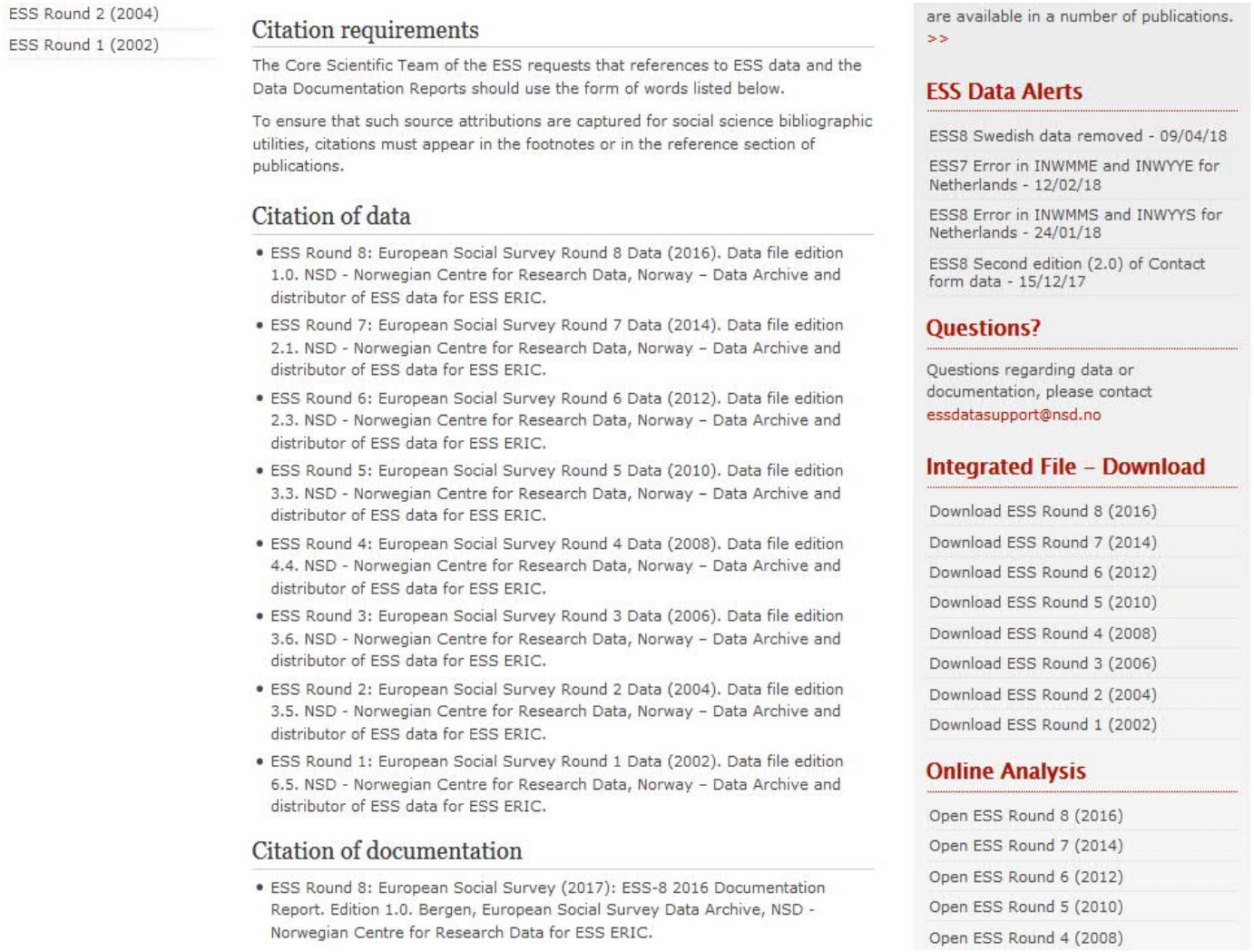
- Give enough information to locate the exact version of the data
- Look for recommended citation
- Use persistent identifiers (Digital Object Identifier (DOI))

CESSDA Training Working Group (2017)

gesis Leibniz Institute for the Social Sciences

cessda eric

And finally...remember to cite data



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

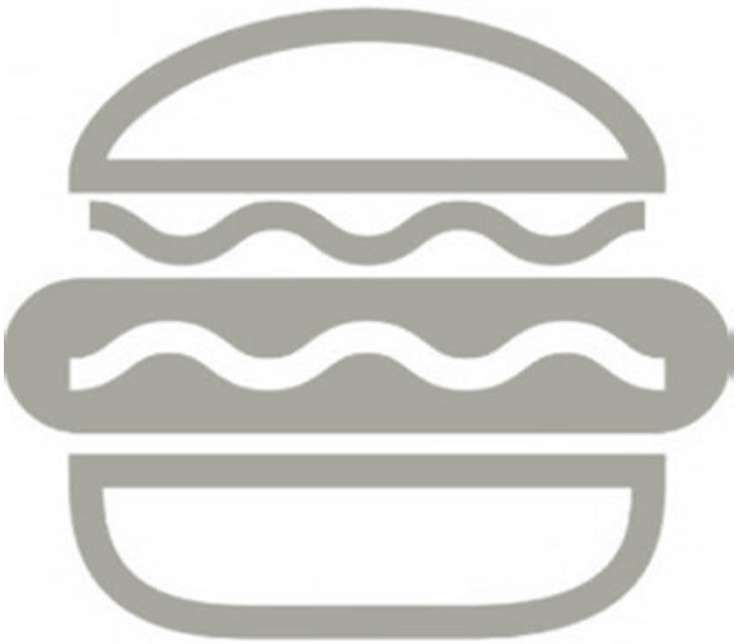
Integrated File – Download

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)

https://www.europeansocialsurvey.org/data/conditions_of_use.html



5. HOW TO HARMONIZE DATA?

gesis Leibniz Institute for the Social Sciences

cessda eric

Digital data harmonization

- Introduction to CharmStats
- What is metadata and how is it used in CharmStats?
- Interface workflow
- Practical example:
  - Finding metadata
  - Creating a project
  - Generating outputs (Report, Graph, Recoding syntax)

6. HOW TO STORE DATA?

gesis Leibniz Institute for the Social Sciences

cessda eric

gesis datorium

- Easy upload of research data free of charge in a secure environment
- Based on Dspace
- Licenses
- Researchers determine the conditions of data (re-)use

Access categories

Free access	With or without registration.
Restricted access	Users have to apply for permission to download the data by contacting the depositor.
Embargo	Data files can be published with an embargo date of one year maximum. The metadata though will be published in any case.

In a nutshell

Documentation
&
Upload

Storage
&
Review

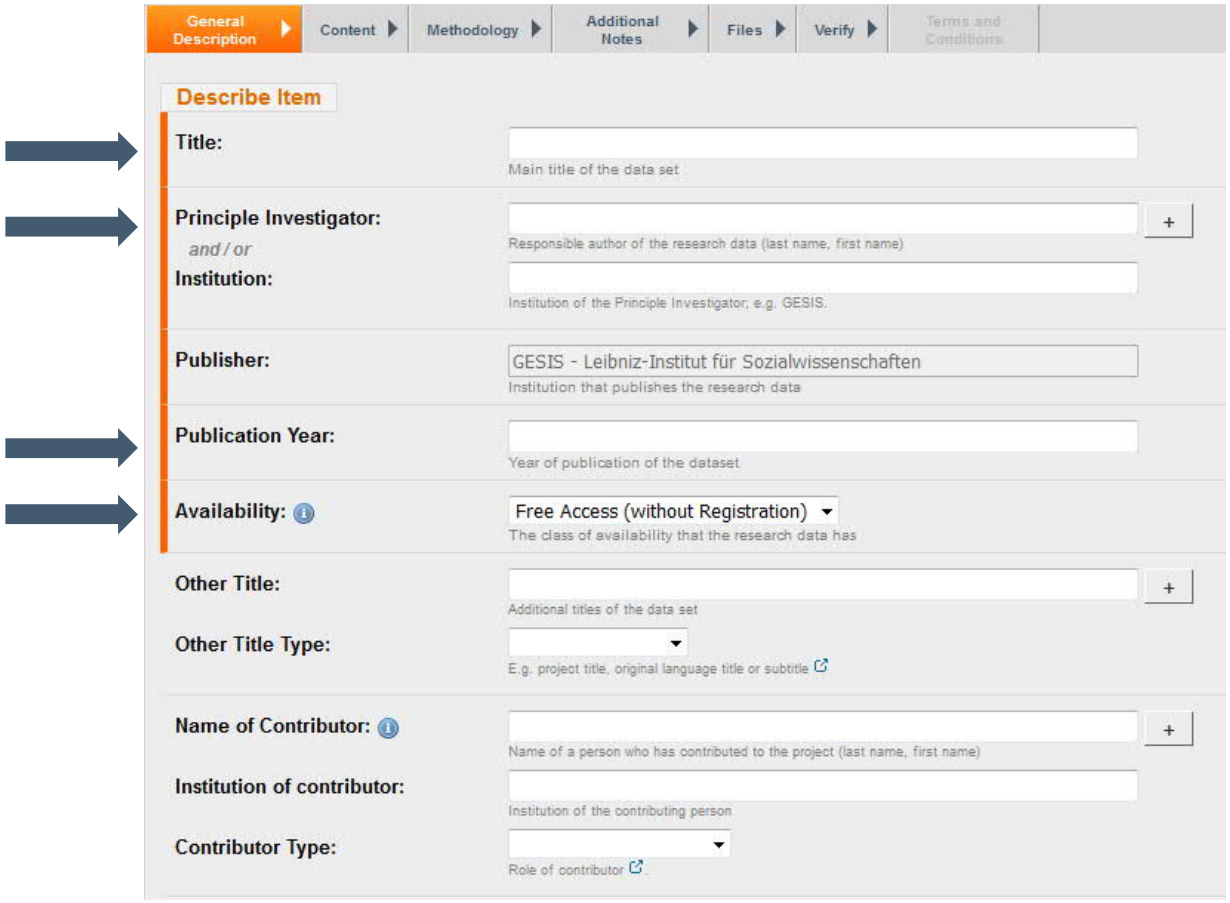
Publication

Search
&
Download

gesis Leibniz Institute for the Social Sciences

cessda eric

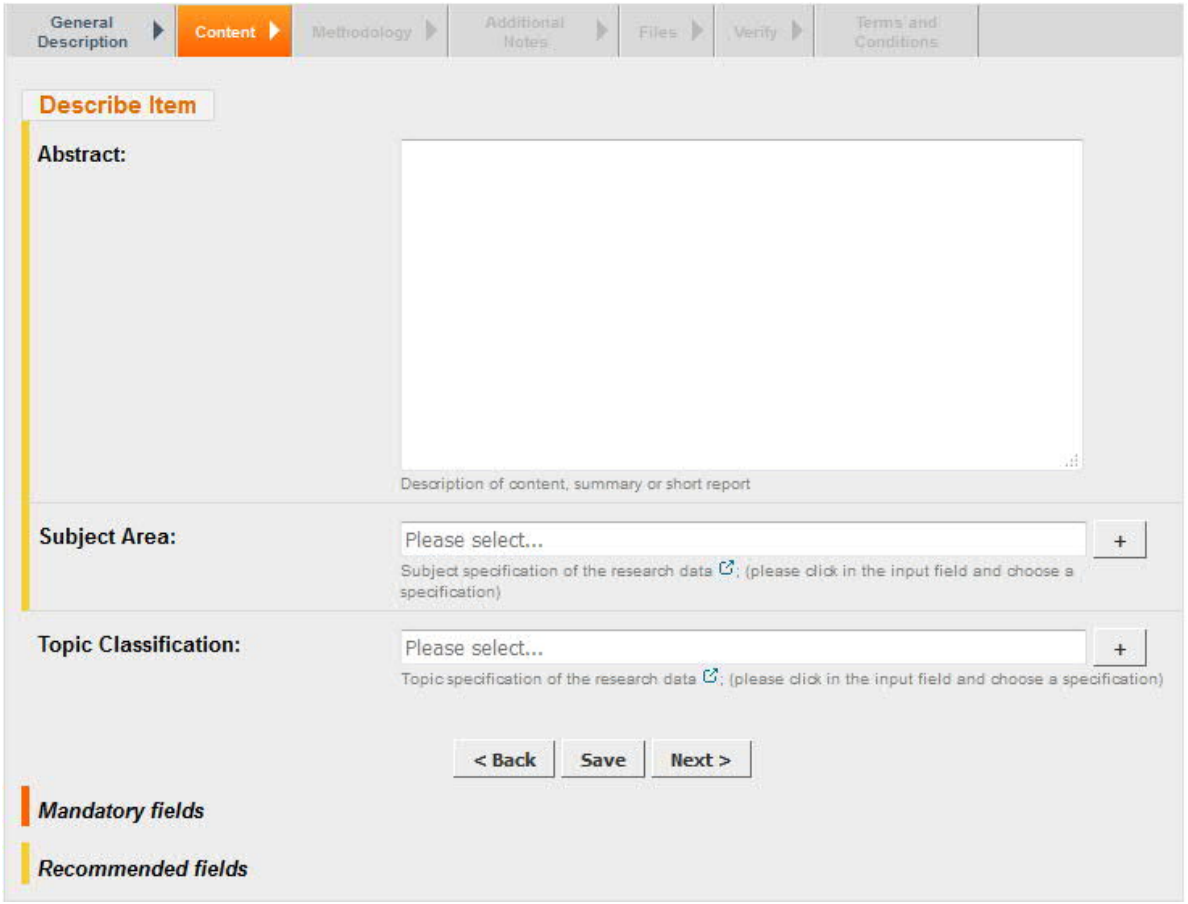
Only 4 mandatory fields.



gesis Leibniz Institute for the Social Sciences

cessda eric

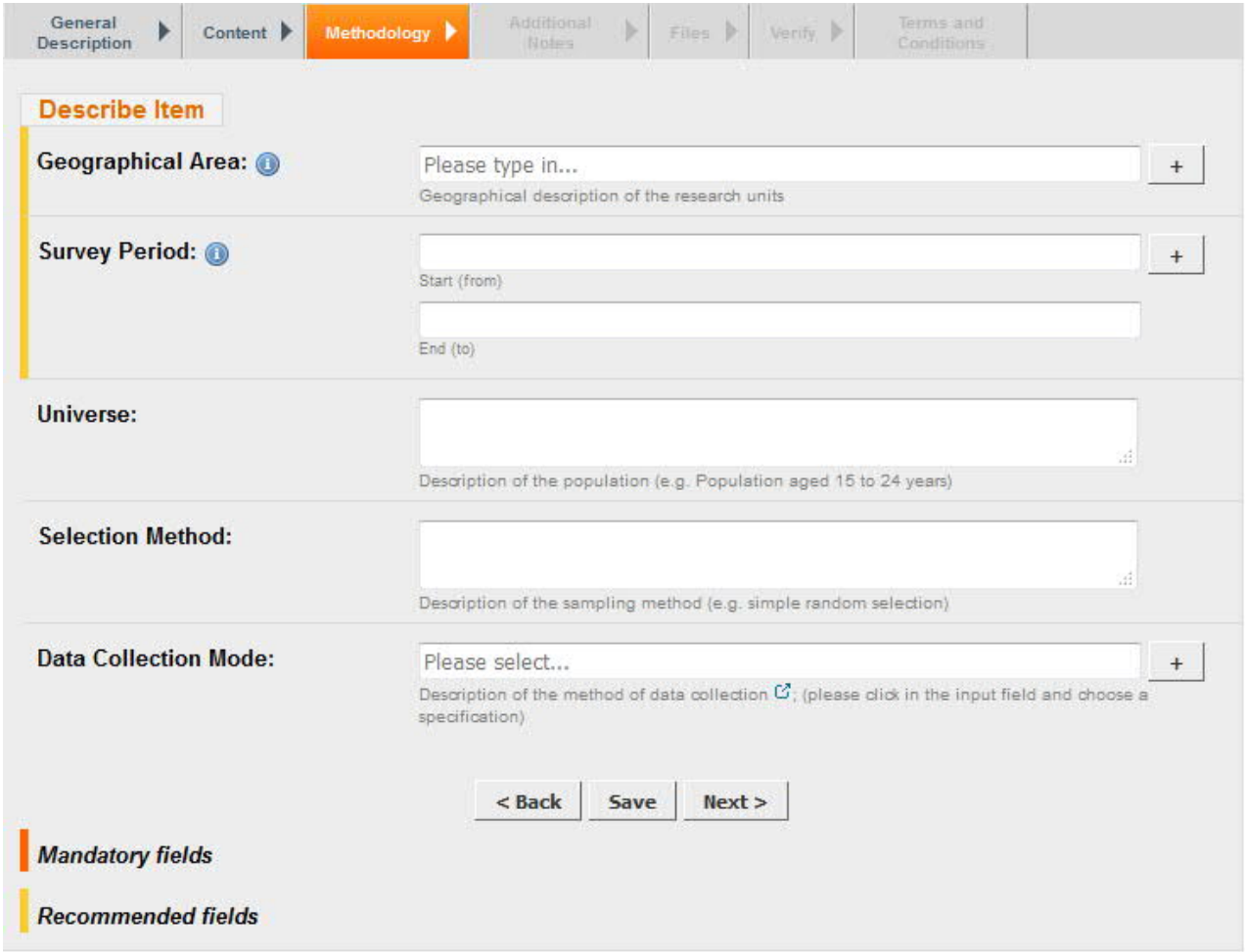
However, it makes sense for…



gesis Leibniz Institute for the Social Sciences

cessda eric

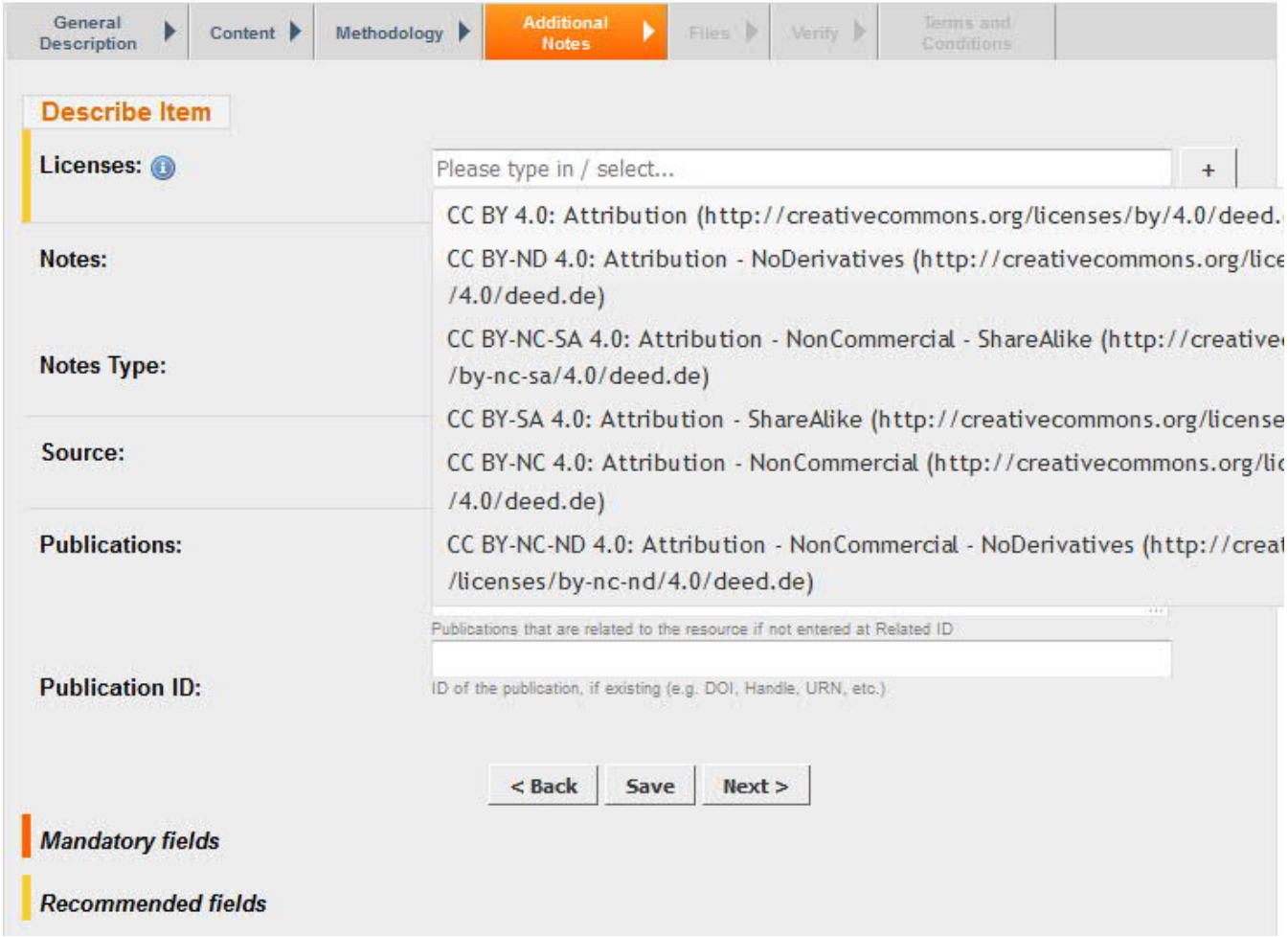
…a better understandig to…



gesis Leibniz Institute for the Social Sciences

cessda eric

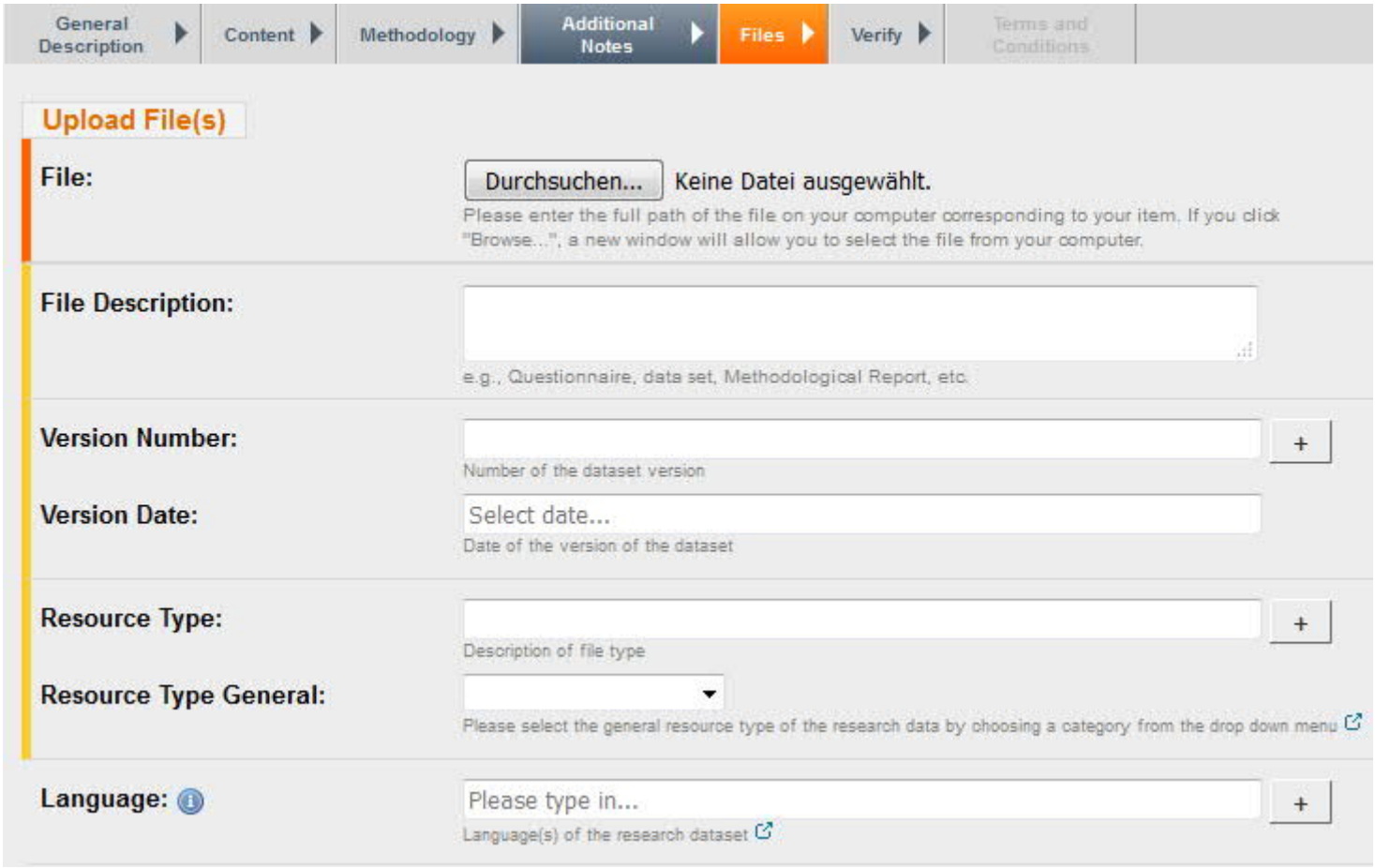
…add some more metadata.



gesis Leibniz Institute for the Social Sciences

cessda eric

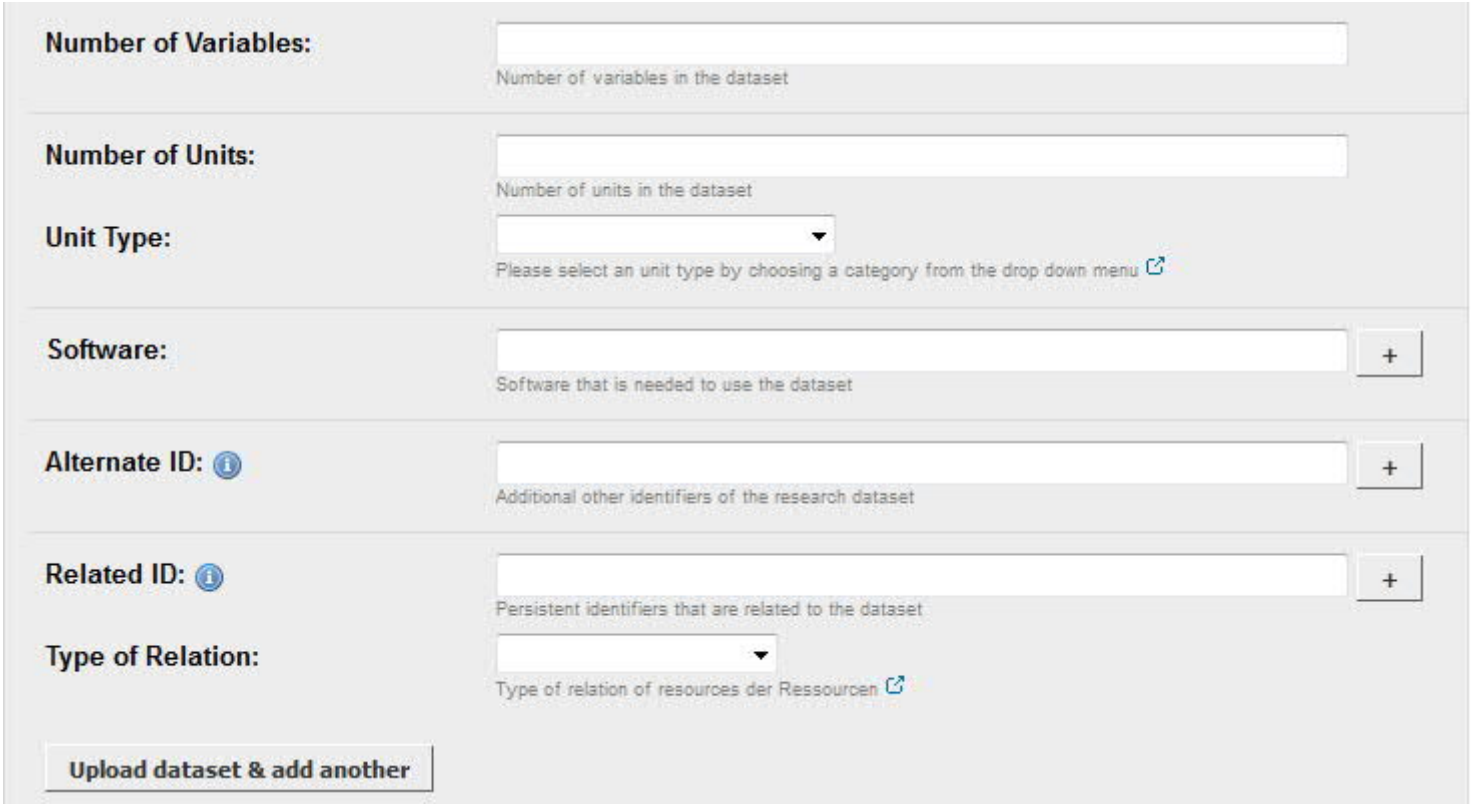
It is also possible to describe…



gesis Leibniz Institute for the Social Sciences

cessda eric

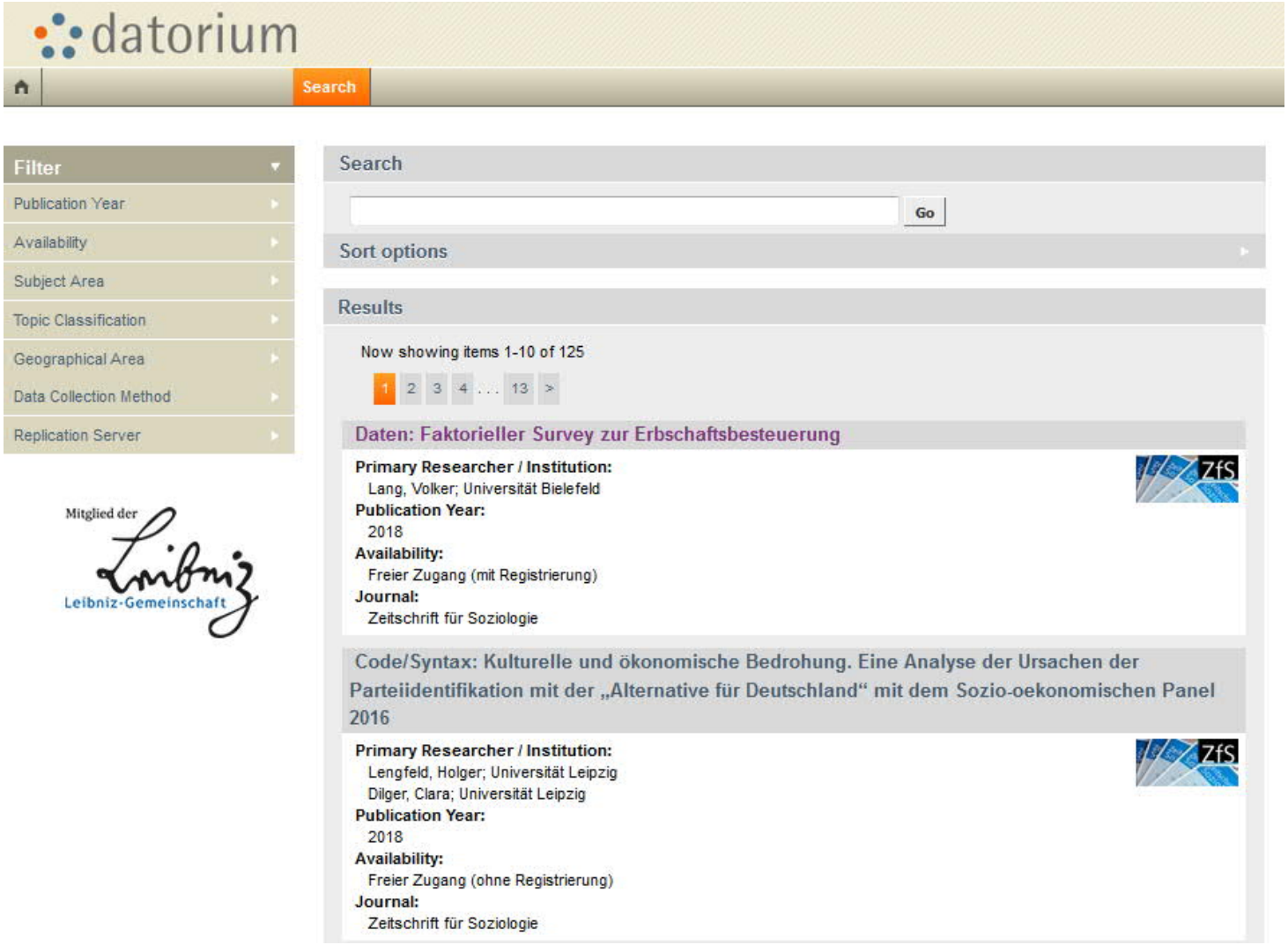
...single files.



gesis Leibniz Institute for the Social Sciences

cessda eric

Data search



gesis Leibniz Institute for the Social Sciences

cessda eric

datorium

View Item

Search

Add data

About datorium

Mitglied der Leibniz-Gemeinschaft

General Description

Back

Title	Daten: Faktorieller Survey zur Erbschaftsbesteuerung
URI	http://dx.doi.org/10.7802/1640
Primary Researcher	Lang, Volker; Universität Bielefeld
Publication Year	2018
Availability	Free Access (with Registration)
Contributor	Groß, Martin; Universität Tübingen; Projektleitung (Project Leader)
Journal Name	Zeitschrift für Soziologie
Volume	47
Issue Number	3
Subject Area	Soziologie Sozialpolitik Erhebungstechniken und Analysetechniken der Sozialwissenschaften
Topic Classification	Sozialpolitik Politische Fragen (Issues) Sparen, Geldanlagen, Vermögensbildung
Abstract	Faktorieller Survey zur Erbschaftsbesteuerung mit besonderem Fokus auf die Vererbung von Firmenkapital, CAWI-Survey auf Basis eines kommerziellen Samplingpools
Geographical Area	Deutschland / DE
Universe	Deutsche Wohnbevölkerung ab Alter 16 Jahre
Selection Method	Zufallsauswahl, quotiert nach Geschlecht und Alter der deutschen Wohnbevölkerung
Data Collection Mode	CAWI (Computerunterstütztes Web-Interview) - Interaktiver Selbstausfüller
Survey Period	2015-12; 2015-12
Licenses	CC BY-NC-SA 4.0
Publications	Groß, Martin; Lang, Volker, 2018: Warum Bürger gegen die Erhebung von Erbschaftssteuern sind – auch wenn sie keine zahlen müssen: Ergebnisse einer Vignettenstudie. Zeitschrift für Soziologie. Band 47, Heft 3 (August 2018); https://doi.org/10.1515/zfsoz-2018-1014
Reference Publication	Groß, Martin; Lang, Volker, 2018: Warum Bürger gegen die Erhebung von Erbschaftssteuern sind – auch wenn sie keine zahlen müssen: Ergebnisse einer Vignettenstudie. Zeitschrift für Soziologie. Band 47, Heft 3 (August 2018): https://doi.org/10.1515/zfsoz-2018-1014

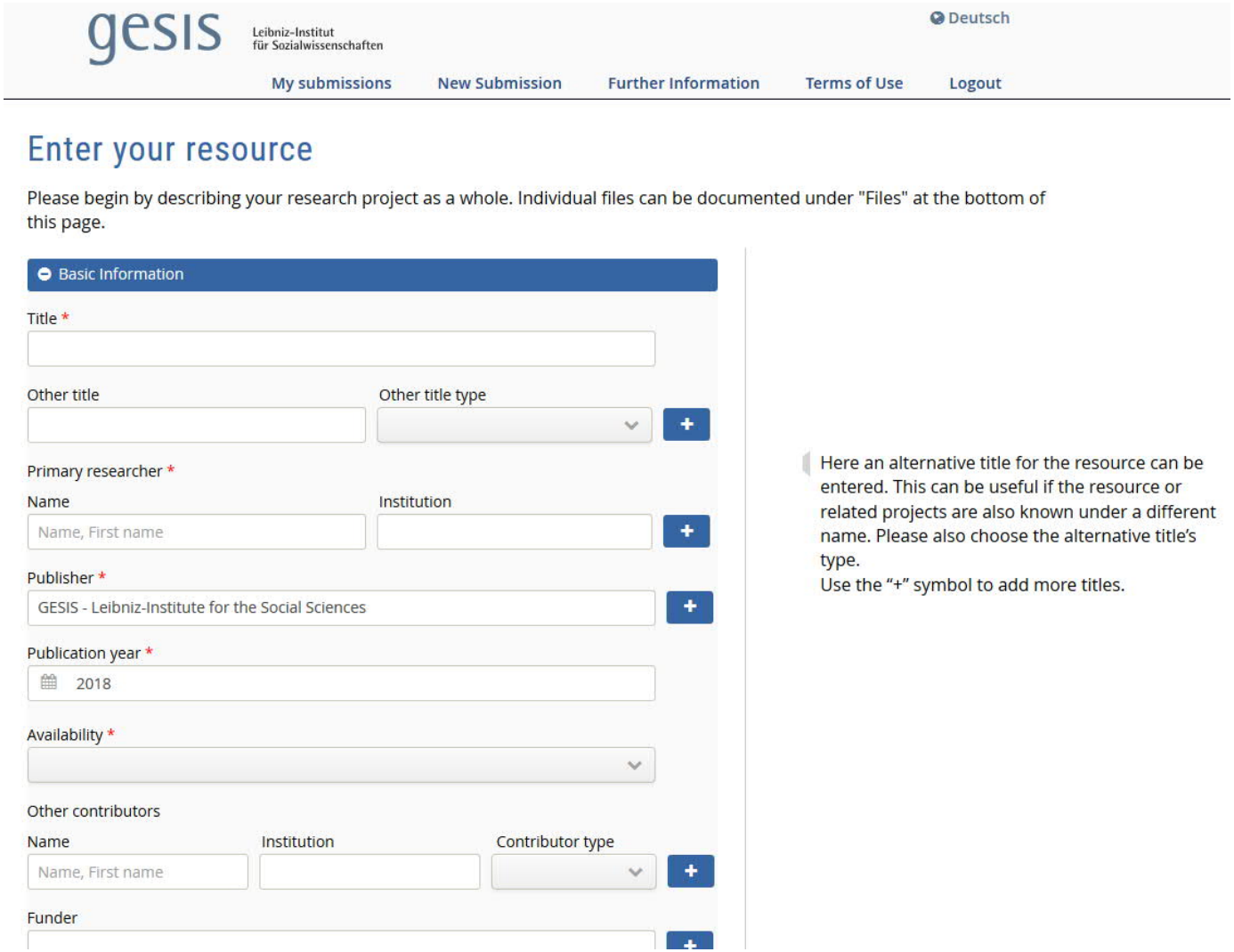
Review process

- Technical review:
  - check whether all delivered material is complete, correct and in a suitable technical condition (e.g. readable, virus free, etc)
- Content review:
  - concerning plausibility, consistency, data weighting and data privacy

gesis Leibniz Institute for the Social Sciences

cessda eric

Work in progress



gesis Leibniz Institute for the Social Sciences

cessda eric

Benefits of datorium

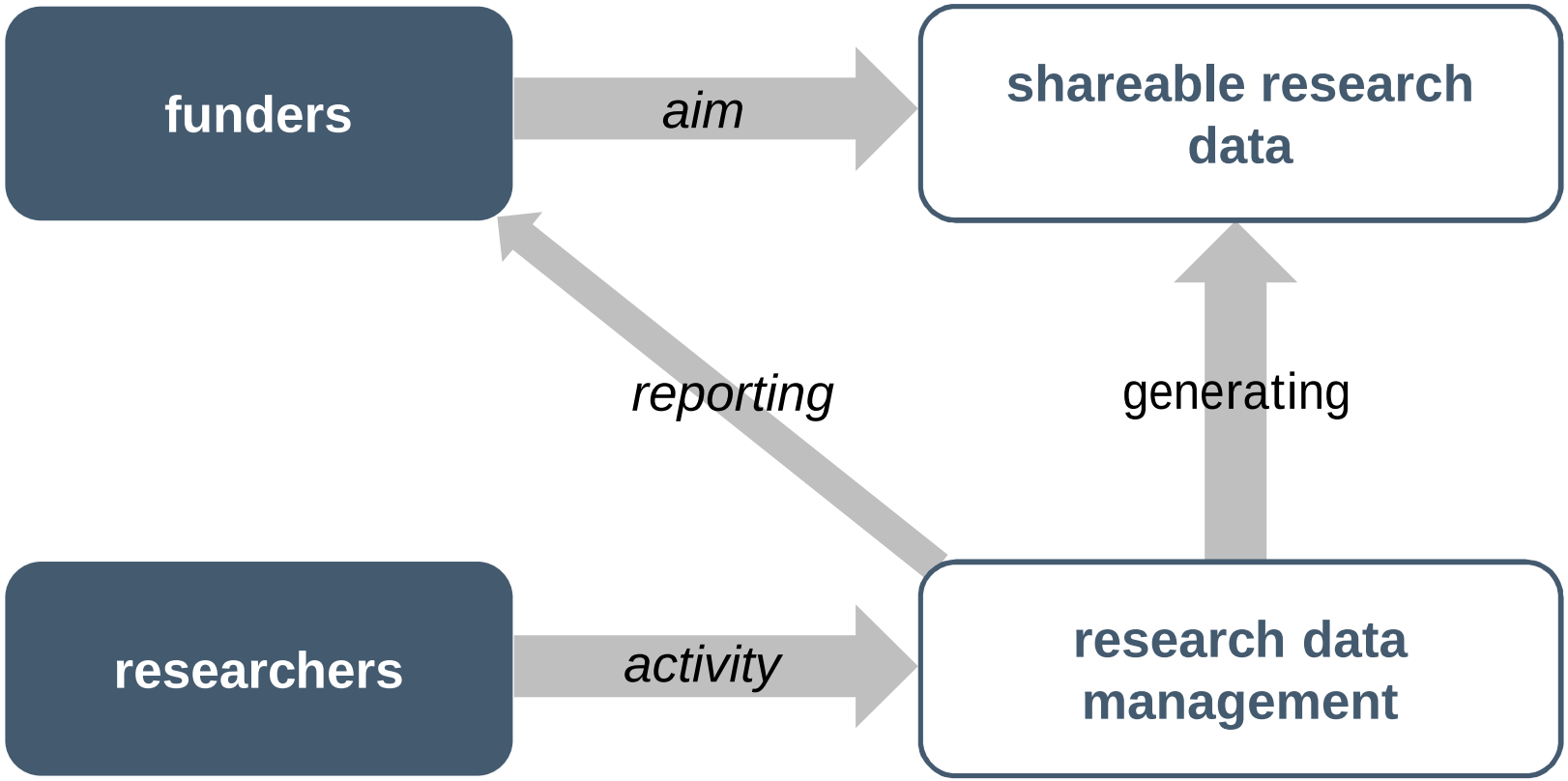
- Easy upload of research data **free of charge** in a secure environment
- All datasets receive a persistent identifier (DOI)
- Data accessible via
  - datorium webpage
  - DBK
  - da|ra
  - DataCite
- Well-balanced mixture of mandatory and optional metadata fields in its submission form

7. WRAP-UP

gesis Leibniz Institute for the Social Sciences

cessda eric

Open Science and Funder Requirements



!!! Keep in mind to plan ahead !!!

gesis Leibniz Institute for the Social Sciences

cessda eric

The Data Spectrum

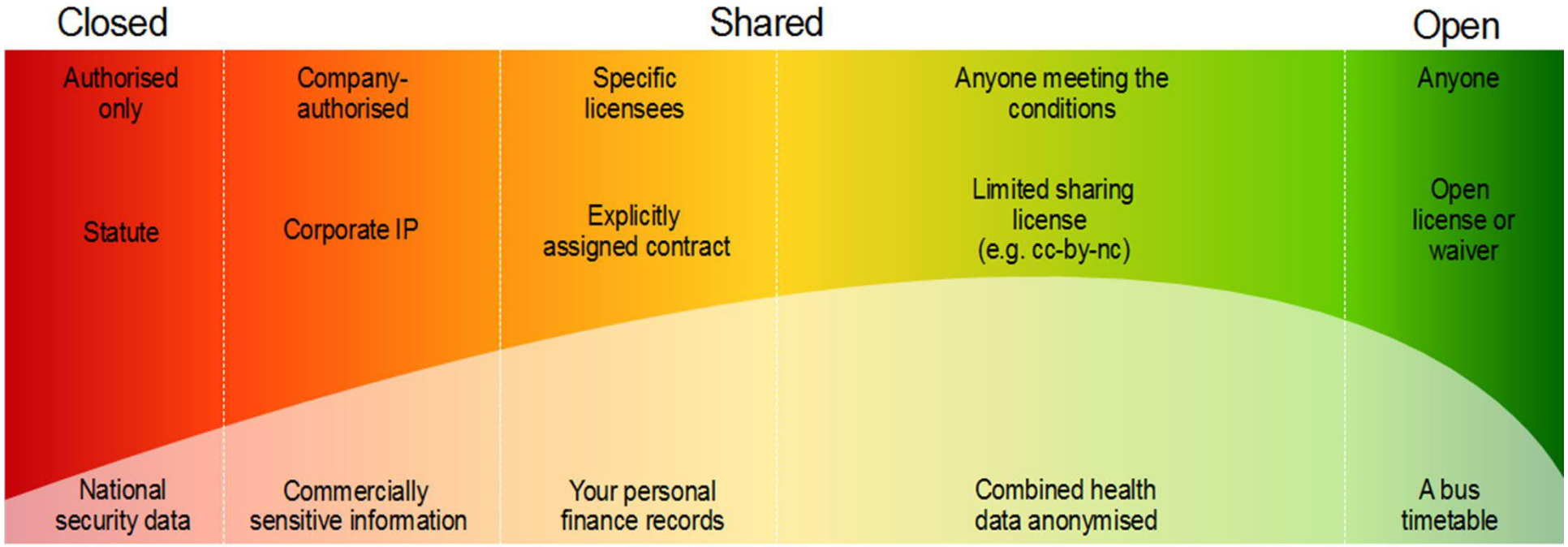


Image: “The Data Spectrum”. Adapted from http://theodi.org/data-spectrum (Open Data Institute, cc-by)

USE FOSTER TO:

Access Free Courses

Get Badges

Earn Specialisation

Attend live events

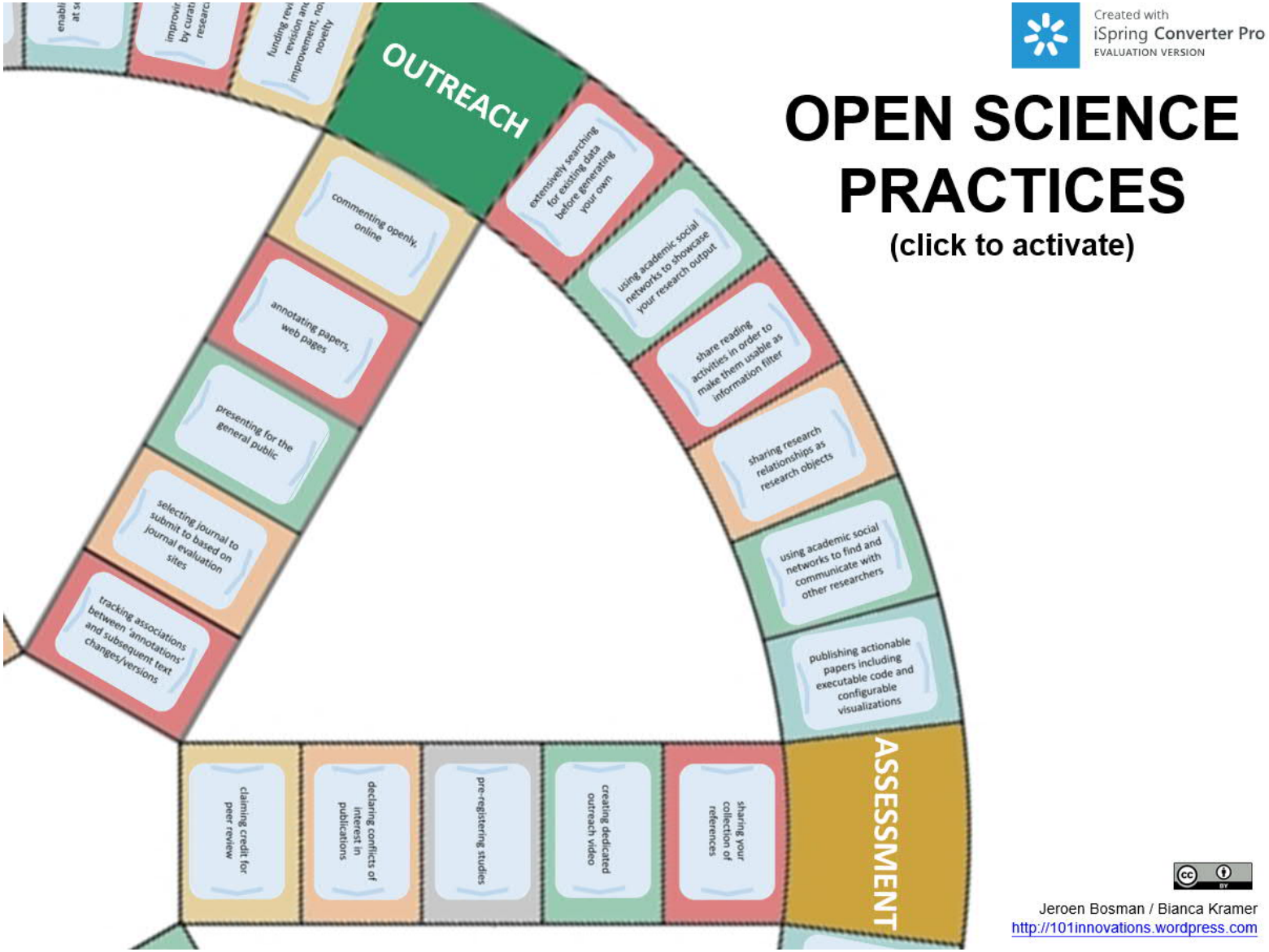
Participate in the community

https://www.fosteropenscience.eu/

But also…

… be an open researcher yourself!

- Be FAIR
- Promote Open Science
- As a reviewer, ask for data citation or for the authors to share their data or code
- Include Open Science statement in your papers



https://bmkramer.databox.me/Public/Wheel_of_Open_Science/

Thank you for your attention!

anja.perry@gesis.org
kristi.winters@gesis.org

This work is licensed under Creative Commons Namensnennung 4.0 International Lizenz.