EXERCISE BOOKLET

Research Data Management

Dr. Anja Perry, Oliver Watteler

September 16 - 21, 2019,

Muteesa I Royal University, Masaka, Uganda

Content

A note on the exercises

These exercises are based on real studies, data, reports, and requests. For this workshop we manipulated data and combined different reports as well as requests. Real names of projects and involved researchers were replaced or deleted.

gesis Leibniz Institute
for the Social Sciences

Exercise 1: Have a first look at Data Management Plans

> ✺   Work individually
> ☼   Time: about 15 minutes
> ☺   At the end, we will discuss your thoughts with the class.

Have a look at the Data Management Plan (DMP) template provided and 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 may have encountered in the past

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

gesis Leibniz Institute
for the Social Sciences

Exercise 2: Do your research I

> ♨ Work individually
> ☼ Time: Work on this task and the following "Do your research" tasks during the afternoons throughout the rest of the week
> ☺ Submit your work on Friday.

Find 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?

1) Consider the following bullet points to identify data needs to explore theories of your research question:
  • **Key concepts**
  • **Units of analysis** (e.g. individuals, households, regions or countries?)
  • **Geography** (e.g. which country or countries)
  • **Time** (What period do you want data to cover? E.g. recent, a specific period, a long period; Do you need data from the same sample at different time points?)
  • **What is the population?** (Everyone, young people, old people, voters)
  • **Study design and sample** (Do you need a representative sample, a random sample, a large sample?)

2) Search for suitable data and describe how you approached your search.
3) 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.

4) Start setting up a Data Management Plan (DMP) and fill in everything you can already answer. Feel free to use the template provided. You will return to your DMP throughout the week.

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

Exercise 3: Informed Consent

> 👥 Work in 5-6 groups
> ① Time: about 30 minutes
> ☺ At the end, one member of your group should briefly present the results of your discussion and your conclusions.

Have a look at the consent form on <u>page 5</u> of the exercise booklet. Discuss this consent form and particularly think about

- the information presented: can it be considered sufficient to inform (potential) participants about the project, their rights and the use of the data collected;
- the content of the consent form: are there any aspects you would not include in such a form, is there any phrasing that violates existing laws or that is contradictory;
- the employed language: is it appropriate, clear and understandable for a layperson;
- any other noticeable problems with this consent form.

Reference:
Netscher, S. & Recker, J. (2015). Exercise: Informed Consent. Cologne, Germany: GESIS – Leibniz Institute for the Social Science.

University of Sciences
Faculty for Research
Chair for Social Science Research (CSS)
Contact: shk.uos@css.uni

CONSENT FORM

Integral Gender Equality Research

The study **Integral Gender Equality Research** (IGER), implemented by the **Chair for Social Science Research** (CSS), **University of Sciences**, evaluates pupils' attitudes toward gender discrimination at vocational schools. It consists of two parts: a short quantitative questionnaire (about 60 min) and, for selected participants, a two hour group discussion.

	YES	NO
I'm adequately informed about the IGER project and …		
…participate in the quantitative part	☐	☐
…participate in the qualitative part	☐	☐
I'm adequately informed about my personal rights and understood that …		
…I can quit the quantitative study at any time	☐	☐
…I can quit the group discussion at any time without any explanation for my drop-out	☐	☐
…I can withdraw my data within a period of 24 hours	☐	☐
…my personal data will be kept confidentially and anonymized	☐	☐
…my personal data will only be used by researchers of the current project	☐	☐
I agree that		
…my wording (qualitative study) may be quoted in publications, reports, Webpages, and other research outputs, together with my real name, without any further approval from me	☐	☐
…my data will be archived at the end of the project and are thus re-useable for other researchers	☐	☐
…researchers may contact me by email (quantitative study) in the future	☐	☐

Signature of participant: _____

Date: _____

Thank you

The Project Manager

University of Sciences
Faculty for Research

our name says it all

Exercise 4: Do your research II

    💬   Work individually

    ☼   Time: Work on this task and the following "Do your research" tasks during the afternoons throughout the rest of the week

    ☺   Submit your work on Friday.

---

1) Adapt your Data Management Plan regarding the protection of your participants
   - go through the respective sections and fill in the information needed for your study
2) Develop a consent form, considering…
   - which information will be provided
   - how the handling of sensitive data and the question of data sharing will be addressed
   - when informed consent will be obtained

Exercise 5: Re-identification

> ♨ Work in 5-6 groups
> ☉ Time: about 30 minutes
> ☺ At the end, one member of your group should briefly present the results of your discussion and your conclusions.

Have a look at the email on <u>page 8</u>. In your group, try to re-identify Daniel, i.e. de-anonymize the data by extracting the corresponding ID-variables (D1005 or D1009) with the help of the given information.

If you are able to do so, discuss
- what this teaches us about the level of anonymization of the data;
- how we should respond to Elizabeth.

To deal with this exercise you should use the *CSES*-data stored in ILIAS.

> *Hint: The German electoral system is a two-vote system. While the second vote depends on a party list vote within each of the German states, the first vote is a candidate vote in small so-called single member districts. To identify Daniel, you should look for the primary electoral district where he may have cast his first vote ballot.*

From: Best, Colleague
Sent: Montag, 27. Juli 2015 11:23
To: Netscher, Sebastian
Subject: Playing CSES matchmaker?

Hi Sebastian,

I received this email from an old friend of mine. Knowing that you were involved in the CSES which contains the German post-election study for 2013, maybe you could help Elizabeth find her lost love?

However, even if we could identify Daniel in the dataset and get his contact information (based on his ID), I doubt that we are in an ethical position to give her the information! What do you think?

Cheers,
Colleague

From: elitru@junkmail.com
Sent: Montag, 27. Juli 2015 10:44
To: Best, Colleague
Subject: HELP PLEASE!!!

Hi Colleague,

I'm in desperate need of help. I lost touch with a person I met briefly back in 2014 but who affected me deeply.

We met in January 2014 on a skiing holiday in St Moritz, Switzerland. You know, I never believed in love at first sight but as soon as our eyes met it became evident he was my better half. I had the sense of being and wholeness like never before –being MYSELF, being ALIVE, being in LOVE!

We spent an entire day together and planned to meet again just before he was due to fly back to Koeln (Cologne) the following morning. When I woke up that day he had slipped a love message under my door. For the first time in years I felt loved. At that moment I suddenly panicked. The wounds of my past heartbreaks resurfaced and it was too much. I didn't go to meet him before his flight. All I could do was run away.
Soon afterwards I started suffering from severe insomnia and Daniel words began haunting me: 'This is the beginning of anything you want.' I MUST find Daniel, it's the only way I can reconcile with my past heartbreaks and hopefully move on.

I spent the last year looking for him unsuccessfully. My last hope lies with you. I remember that he once said he had been interviewed for the German post-election study in the fall of 2013. I know you are involved in political research and have been working for a Germany research institute for a couple of years now. Maybe you can help me to get in contact with Daniel by asking someone who is involved in the German post-election study?

Here is what I can tell you about him. (It is funny that we shared so many profound things but I can't recall his last name or his profession.) Daniel never got married but has a dog he truly adores. I recall him saying he felt so guilty undertaking the skiing trip without his dog.
I could write so much more but I don't want to take too much of your time. All I hope is that these few lines convince you of my good intentions. I regret not seeing him on that last morning and I have to do everything I can to find him. Please help me find my lost love.

Yours hopefully,
Elizabeth

Exercise 6: Do your research III

    👥   Work individually

    🕐   Time: Work on this task and the following "Do your research" tasks during the afternoons throughout the rest of the week

    ☺   Submit your work on Friday.

---

1) Adapt your Data Management Plan regarding
   a. Versioning
   b. Interoperability
   c. Anonymization
2) Work with the training dataset. Please complete the following tasks:
   a. Check the dataset for 'wild codes'. Are there any codes in the dataset that do not seem to have a substantial meaning?
   b. Check the dataset for 'missing values'. Are there any missings in the dataset? Beware of codes that look like missing values but belong to valid values.
   c. Variables Q55A through Q55F belong to three pairs of questions. The first question of each pair holds a filter. Check if the filters were used correctly. Please consult the data codebook for question wording and interviewer instructions.
   d. Variable Q29A is about the meaning of 'democracy'. It is the starting variable to a range of other variables concerning this issue. Please check Q29A and the following variable and see if you find anything that strikes you in the answering behavior of respondents. This task is rather philosophical than logical. There are no errors here but an important point for discussion.

gesis   Leibniz Institute
for the Social Sciences

Exercise 7: Data Documentation

> ✿ Work in 5-6 groups
> ① Time: about 30 minutes
> ☺ At the end, one member of your group should briefly present the results of your discussion and your conclusions.

Have a look at the following field report and discuss within your working group how useful you consider it for the purpose of re-using the data. Take into account

- the content of the document and the language used, i.e. is the report understandable and unambiguously written? Do you miss any important information or are there any statements you would not include in such a field report? Do you recognize any contradictions in the description, etc.;
- the structure of the document, i.e. how would you present the relevant information in such a field report?

FIELD REPROT InDiLeSo   14.09.2015

FIELD REPORT

RESEARCH PROJECT

Interdisciplinary Learning in Social Sciences

(InDiLeSo)

The current sample can be considered to represent the population of interest (students of social sciences at Swiss universities). In our research project we visited several colleges in Switzerland in the cantons of Basel, Lausanne and St. Gallen. Our highly qualified interviewers (most of them student research assistants) interviewed a total sample of 357 students, employing a standardized questionnaire. In the project, we randomly selected three universities and visited them on three Fridays in June and July 2015. There, we attended lectures based on availability and chose a number of students to briefly interview them (about 30 min) after the sessions.

Reference:
Netscher, S. & Recker, J. (2015). Exercise: Data Documentation. Cologne (Germany): GESIS – Leibniz Institute for the Social Science."

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

Exercise 8: Evaluate the data

> ♟ Work in 5-6 groups
> ☺ Time: about 10 minutes
> ☺ At the end, one member of your group should briefly present the results of your discussion and your conclusions.

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".

Reference:
Chylikova, J., J. Krejci (2018). The CESSDA Expert Tour Guide. Chapter 3. Process [presentation]. Bergen: CESSDA ERIC.

Exercise 9: Do your research IV

> ☁ Work individually
> ☼ Time: Work on this task and the following "Do your research" tasks during the afternoons throughout the rest of the week
> ☺ Submit your work on Friday.

1) Adapt your Data Management Plan regarding Data Documentation and Metadata.
2) Exercise: Data documentation

When as a researcher you want to reuse existing data, for example for new analysis or meta-analysis, it is important that you understand what exactly the data mean and how they were created. You also need to be able to assess whether the data are fit for your intended purpose. This requires the data to be well described and documented and well prepared.

Understanding how to best prepare and document your own research data when publishing them for future reuse, is easiest to appreciate from the perspective of the new user who is unfamiliar with the dataset.

For this exercise we will use the following datasets:

1. Malawi Household surveys for agricultural biodiversity assessment

2. Manufacturing growth and the lives of Bangladeshi women

3. All Ireland Traveller Health Study (AITHS)

Listed below are examples of key information that a researcher using data may need to know. Evaluate the example dataset descriptions you find on the websites to see whether you can find this information easily. Also indicate where you found the information, e.g. in the dataset descriptor, in supplementary documentation, in a related publication. Consider whether this could have been described or presented better and whether additional information would help to reuse data.

Did you find the information, and where?			
Key information needed for reuse of data (examples)	1 Malawi HH survey	2 Manufacturing Bangladesh	3 AITHS
Example: Number of respondents	340; found in the dataset description	1395; found in the ReadMe file (part of the data zip bundle)	8492; found in the AITHS Technical Report 1
Geographical area where the data were collected			
Is there sampling bias or is the sample random?			

Is there a control group?			
Were data collected directly in digital format or on paper and then submitted/transcribed into a database; if so was double entry or peer checking done to avoid errors?			
Which questions exactly were asked in the survey or interview (or which protocols used for measurements)?			
Can you find the hypothesis or aims of the research that generated this dataset?			
How was consent gathered?			
Can the data be used for commercial purposes?			
What access conditions apply to the data?			
Can you find a publication that describes the findings of this dataset?			
Does the file format and structure of the data facilitate easy reuse?			

Reference:
Veerle Van den Eynden, Scott Summers and Stuart MacDonald (2018) Exercise: Data documentation. UK Data Service, UK

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

gesis Leibniz Institute
for the Social Sciences

Exercise 10: Licenses and Data Re-Use

> 👥 Work in 5-6 groups
> ⓘ Time: about 30 minutes
> ☺ At the end, one member of your group should briefly present the results of your discussion and your conclusions.

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.

Discuss
- whether the different datasets can be re-used in your research project;
- whether you can cumulate all three individual datasets, merge the administrative data and publish this integrated dataset. If yes, under which license would you publish the integrated dataset?
- whether you are allowed to publish your research results with journal publisher.

The Licenses

(cc) BY	**Attribution (CC BY)** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
(cc) BY SA	**Attribution – share-alike (CC BY-SA)** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
(cc) BY NC	**Attribution – non-commercial (CC BY-NC)** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. You may not use the material for commercial purposes. A commercial use is one primarily intended for commercial advantage or monetary compensation.

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

Reference:
Netscher, S. & Recker, J. (2015). Exercise: Licenses and Data Re-Use. Cologne (Germany): GESIS – Leibniz Institute for the Social Science."

Exercise 11: Do your research V

> ♨ Work individually
> ☺ Time: Work on this task and the other "Do your research" tasks during the afternoons throughout the rest of the week
> ☺ Submit your work on Friday.

Adapt your Data Management Plan regarding...

1) Develop a back-up strategy for your data. Who is responsible for the back-up?
2) Describe the organization of files in your project.
3) How about data sharing? Which licenses do you assign? Does your choice regarding licenses comply with your informed consent? Where will you archive your data?