

# Research Data Management

## Legal and Practical Aspects

Center for Advanced Internet Studies (CAIS), Bochum

June 30<sup>th</sup>, 2022



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

# Who are we?

## Dr. Anja Perry

- GESIS – Leibniz Institute for the Social Sciences
- Active in research data management since 2016
- Consulting and workshops on RDM, RDM competence development in KonsortSWD, development of domain data protocols
- Focus on quantitative data





## Qualiservice

Forschungsdatenzentrum für qualitative  
sozialwissenschaftliche Forschungsdaten

# Who are we?

## Dr. Jan-Ocko Heuer

- Research Data Center Qualiservice
- Sociologist; PhD (Dr. rer. pol.) in 2014
- Active in research data management since 2018
- Advice, workshops, development of workflows; coordinating development of RDM portfolio for qualitative social research within KonsortSWD
- Focus on qualitative data

# And who are you?

## Round of short introductions:

- What's your name?
- What types of data do you use?
- What do you expect from this workshop?

# Workshop rules

- Feel free to ask questions at any time
- We want to make this workshop as interactive as possible
- Feel free to share your own examples and experiences
- Slides and other materials will be available after the workshop

# Agenda

**1. Welcome and introduction**

**2. Introduction to RDM**

Coffee Break

**3. Data protection**

**4. Informed consent**

Lunch Break

**5. Anonymization**

**6. Documentation**

**7. Secure Storage**

Coffee Break

**8. Data Sharing**

# 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)**

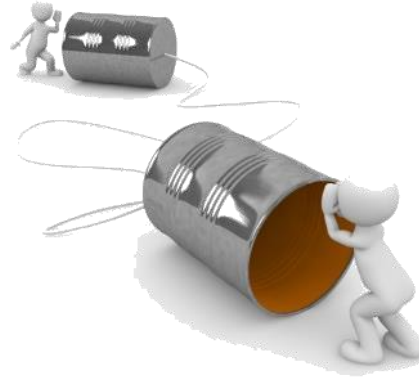


Image: pixabay (CC-0)



# Benefits for teams, researchers, & scientific community

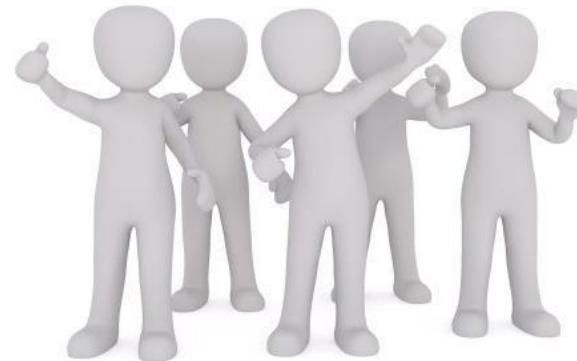
- Saves resources
- Correct results
- Better workflow



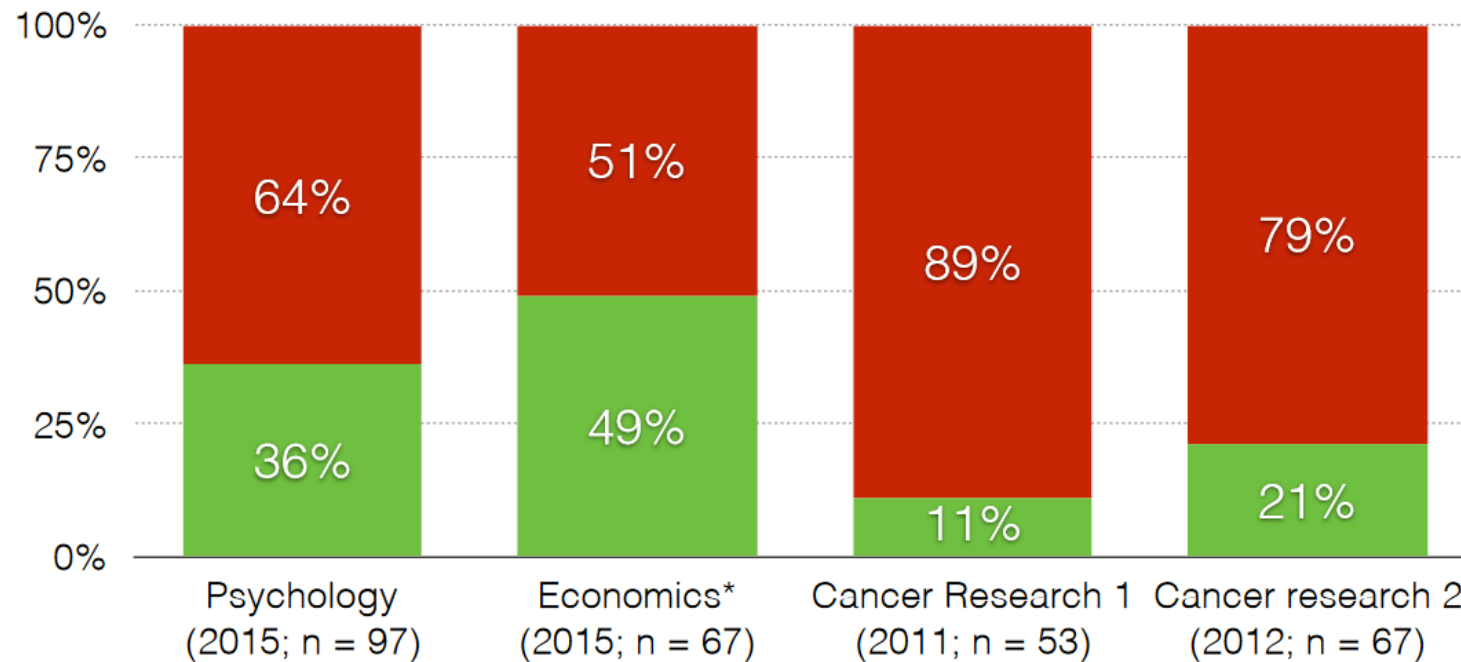
- Proof of transparent & valid conduct
- Facilitates re-use of own data



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



# Replication of published research findings



\*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)

# 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

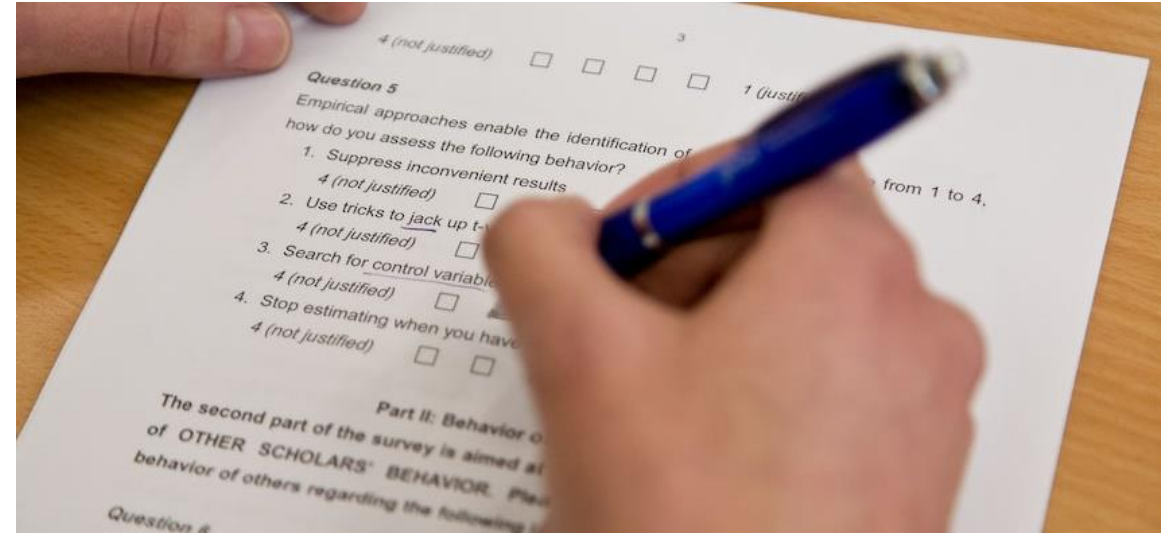
using the data in new  
contexts; answering new  
research questions

# Research Data Management Plan (DMP)

- Central to RDM
- Thought through from the beginning
- Defines responsibilities
- Roadmap for the project



# Quantitative data – Examples



```

Group
[id = BreakOffRouting; response condition = All | Hidden | layout = list]

[BreakOffRouting]

routing
IF ("DISP_CI ≠ 1") THEN Goto DC_CI_manual
ELSE
IF ("DISP_CI = 1" AND ("DISP_BQ ≠ 1")) THEN Goto DC_BO_manual
ELSE
IF (((("DISP_CORE ≠ 1" AND ("DISP_CORE ≠ 3" AND ("DISP_CORE ≠ 4")) AND ("COMPUTEREXPERIENCE = 1")) AND ("CBA_Core_Stage1_Score = NULL" OR ("CBA_Core_Stage2_Score = NULL")))) THEN Goto DC_CORE_manual
ELSE
IF (((("DISP_CORE = 1" AND ("CBA_Core_Stage1_Score > 3")) AND ("CBA_Core_Stage2_Score > 2")) AND ("DISP_CBA ≠ 1")) THEN Goto DC_CBA_manual
ELSE
IF (((("COMPUTEREXPERIENCE = 2" AND ("DISP_PP = 1")) OR (((("COMPUTEREXPERIENCE = 1" AND (((("DISP_CORE = 3" OR ("DISP_CORE = 4")) OR ("CBA_Core_Stage1_Score ≤ 3")) AND ("DISP_PP = 1")))) THEN Goto DC_PP_manual
ELSE
IF ((((((("COMPUTEREXPERIENCE = 2" AND ("DISP_PP = 1")) AND ("DISP_PRC = 1")) OR (((("COMPUTEREXPERIENCE = 1" AND ("DISP_CORE = 3" OR ("DISP_CORE = 4")) AND ("DISP_PP = 1")) AND ("DISP_PRC = 1")) OR (((("COMPUTEREXPERIENCE = 1" AND ("DISP_CORE = 1")) AND ("DISP_PRC = 1")) AND ("CBA_Core_Stage1_Score > 3")) AND ("DISP_PRC = 1")) OR (((("COMPUTEREXPERIENCE = 1" AND ("CBA_Core_Stage1_Score > 3")) AND ("CBA_Core_Stage2_Score ≤ 2")) AND ("DISP_PRC = 1")) OR (((("COMPUTEREXPERIENCE = 1" AND ("CBA_Start = 2")) AND ("DISP_PP = 1")) AND ("DISP_PRC ≠ 1")))) THEN Goto DC_PRC_manual
ELSE Goto BreakOffError
    
```

# Quantitative data – Examples (II)

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
1	1.040e+14	4	26	104012	99	1	1973	2	1642	1040	1040	1	6	1	5	5
2	1.040e+14	1	19	104008	99	2	1943	1	99	99	99	1	1	6	4	5
3	1.040e+14	4	98	104003	99	1	1990	1	99	99	99	1	6	3	99	11
4	1.040e+14	2	98	104010	99	1	1983	2	1756	1040	1040	1	6	3	99	11
5	1.040e+14	1	18	104007	99	2	1927	1	99	99	99	1	1	6	4	1
6	1.040e+14	2	19	104007	99	1	1983	1	99	99	99	1	6	2	5	3
7	1.040e+14	4	15	104005	99	2	1970	1	99	99	99	1	1	2	4	6
8	1.040e+14	4	19	104006	99	1	1942	1	99	99	99	2	1	6	5	88
9	1.040e+14	4	15	104005	99	1	1965	2	1040	1276	1040	7	4	1	5	5
10	1.040e+14	7	15	104004	99	2	1955	1	99	99	99	1	1	2	5	5
11	1.040e+14	7	15	104003	99	2	7777	1	99	99	99	1	1	6	77	77
12	1.040e+14	4	18	104005	99	2	1938	1	99	99	99	1	3	6	5	5
13	1.040e+14	7	17	104005	99	1	1945	1	99	99	99	1	1	6	5	4
14	1.040e+14	4	18	104005	99	2	1949	2	1040	1380	1040	1	4	6	5	5
15	1.040e+14	2	15	104003	99	2	7777	1	99	99	99	1	1	2	4	5
16	1.040e+14	4	32	104007	99	1	1974	1	99	99	99	1	6	2	5	5
17	1.040e+14	4	98	104006	99	1	7777	1	99	99	99	1	6	3	99	11
18	1.040e+14	4	25	104010	99	2	1968	1	99	99	99	1	1	2	5	1
19	1.040e+14	4	40	104007	99	1	1967	1	99	99	99	1	1	2	6	2
20	1.040e+14	1	23	104004	99	1	1932	1	99	99	99	1	1	6	4	3
21	1.040e+14	2	18	104003	99	2	1965	1	99	99	99	1	6	2	5	5
22	1.040e+14	1	27	104011	99	1	1956	1	99	99	99	2	1	1	5	2
23	1.040e+14	4	18	104004	99	2	1923	1	99	99	99	1	3	6	3	3
24	1.040e+14	1	19	104006	99	2	1952	1	99	99	99	1	4	2	5	3
25	1.040e+14	3	16	104004	99	2	1947	2	1276	1040	1040	1	1	6	3	3

# Qualitative data – Examples

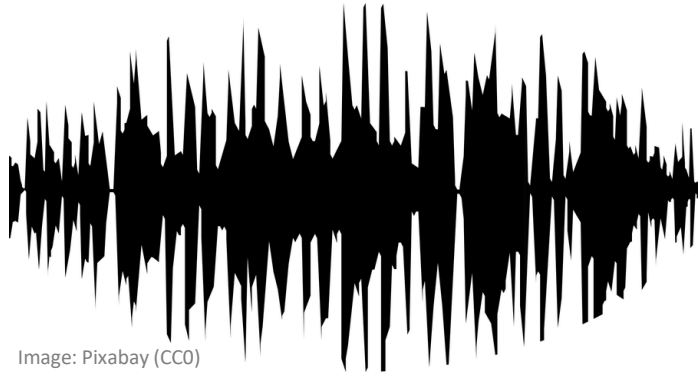


Image: Pixabay (CC0)



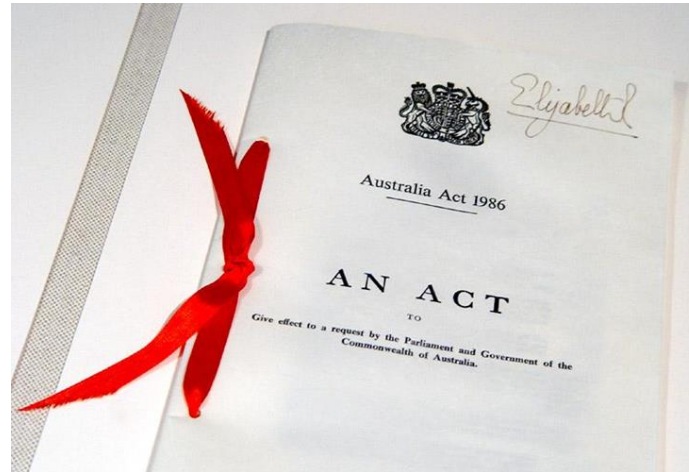
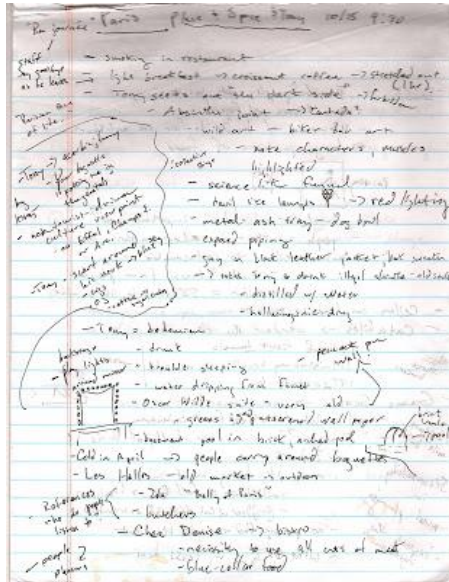
Images above & right: Pinxio (CC0)

Data	Codes
<p>(Q1) Moderator: What do you think about the apparel industry in Sri Lanka? Who are the customers, competitors, suppliers and influential parties and <i>what</i> is their influence on the business and management controls of the business?</p> <p>Finance Director: Global competition and GSP plus [Generalized System of Preferences] and the Trans-Pacific Trade Partnership have a direct impact on our operations, and on management controls. More than 60% of Sri Lankan apparel exports are to the United States and other Western countries. Our operations and controls have to focus on their demands in order to survive in the market. In the meantime, competition in the industry is increasing from emerging countries like Laos, Cambodia and Vietnam. So, our controls have always to focus on reducing costs and producing products at competitive prices. The economic environment of our country too is not supportive of the apparel industry because other industries such as tourism and hospitality are growing. Thus finding labour for the apparel industry will be difficult in about another five years because they expect</p>	<p>Global statutory rules</p> <p>Customer regulatory demands</p> <p>Influence of changing market conditions</p> <p>Facing changes in the local industry</p> <p>Influence of community</p>

Audio or video recordings and transcripts (of interviews, focus groups, group discussions etc.)



# Qualitative data – Examples (II)



Observation protocols

Field notes

Research diaries

‘Naturally occurring’ data  
(e.g. legal acts, court files,  
public records, letters etc.)

Photos/images/pictures

Audiovisual data

# Social science data – Characteristics

**Social science data usually contain information about human beings, making them subject to specific ethical and legal considerations**

- Personal data (name, date of birth, address, tax ID etc.), including ‘special categories’ with particular protection
- Confidential information (trade/business secrets, military information etc.)



Image: Pixabay (CC0)

## Social science data – Characteristics (II)

**The (formal) differences between quantitative and qualitative data partly result in specific characteristics of these two kinds of data**

- Quantitative data are usually (from the outset) more standardized and structured than qualitative data
- Qualitative data may contain information that are not subject/focus of the research – put differently: they are very information-rich
- Quantitative data can be more easily transferred from an empirical into a numerical relative (i.e., with less loss of information)

# Quantitative vs. qualitative social research

	<b>Quantitative</b> research	<b>Qualitative</b> research
Example	Quantitative survey	Case study
Research orientation	Natural sciences; positivism	Humanities; interpretative
Aim of research	Causal explanation, 'laws'	Understanding (of meaning)
Research process	Highly standardized	Open, flexible
Research logic	Test of theory / hypotheses	Exploration, theory generation
Scope of results	Prevalence, correlations, statistical representativeness	Typologies, exemplary cases, theoretical generalization

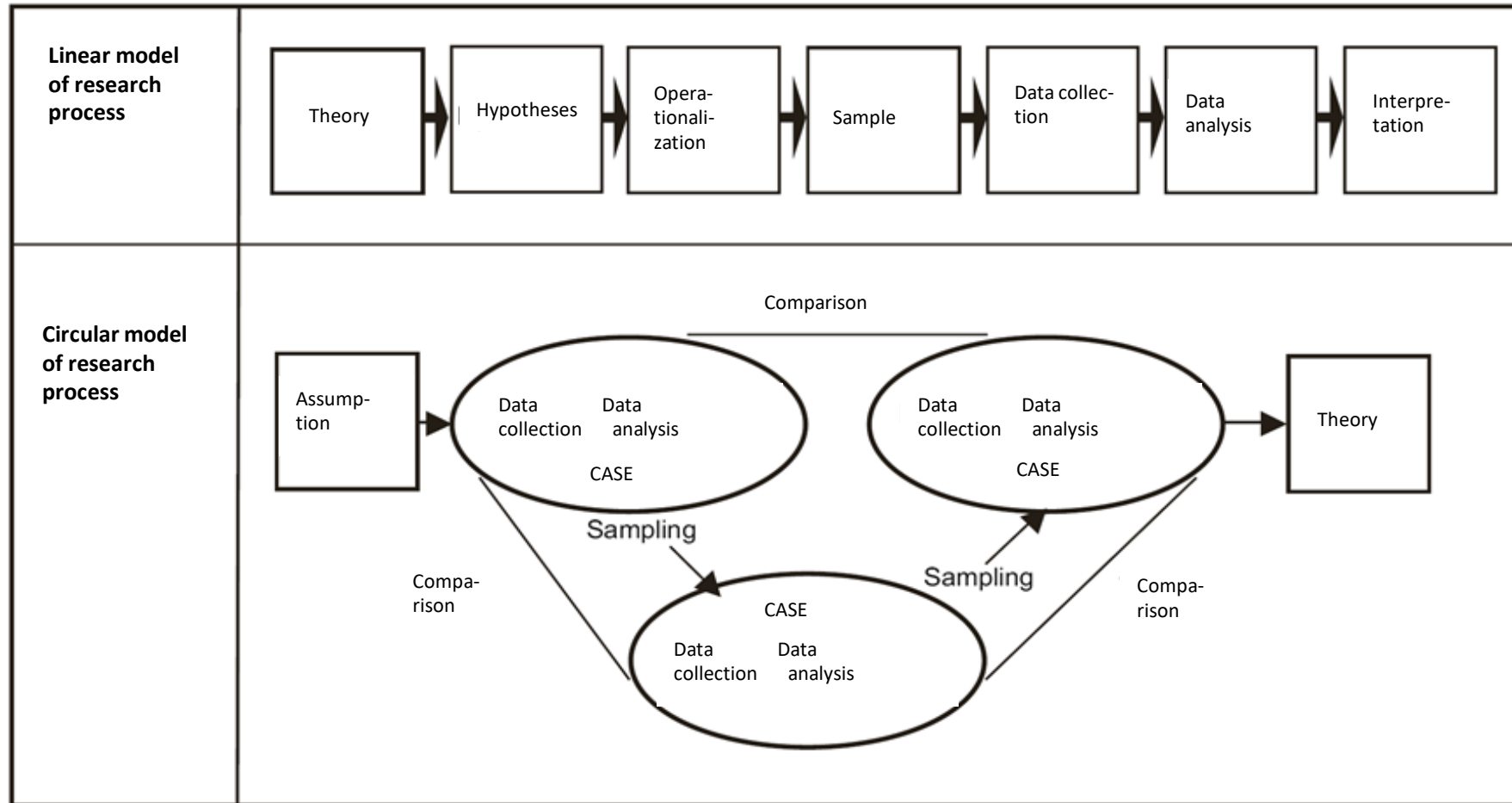


# Quantitative vs. qualitative social research (II)

	<b>Quantitative</b> research	<b>Qualitative</b> research
From data to theory	Reduction, abstraction	Explication & contextualization
Methodological orientation	Ahistorical, closed, particularist, random sample	Historical, open, flexible, holistic, theoretical sampling
Data collection	Researchers rarely collect data themselves	Researchers often collect data themselves ('in the field')
Typical methods of data collection	Survey/questionnaire, experiment, controlled observation	In-depth interview, focus group, participant observation
Typical methods of analysis	Descriptive & inferential statistics	Content analysis, narration analysis, discourse analysis



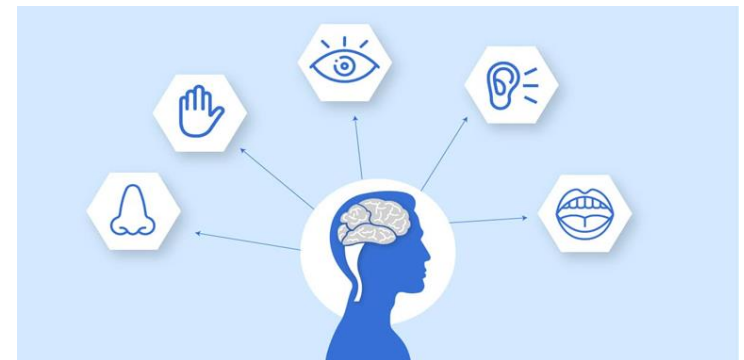
# Quantitative vs. qualitative research process



Source: Flick 2009 (own translation)

# RDM challenges – same but different?!

- Social science data – both quantitative and qualitative – have many RDM challenges in common (e.g. ethical and legal aspects)
- Yet, there are differences from a practical perspective (i.e., in developing solutions), partly, but not exclusively, resulting from differences between quantitative versus qualitative research and data (e.g. due to the degree of standardization)



Images: Pixabay (CC0)

# Questions?



# Data protection

What do you associate with the term  
*data protection*?



# Research ethics in empirical research

- **Research ethics:** norms and guidelines for responsible conduct of research (e.g. honesty, objectivity, integrity, social responsibility etc.)
- Empirical research **on/with human beings** → specific issues regarding research practices and the researcher-researched relationship
- **Principles** e.g.: minimize harms and risks; maximize benefits; respect human dignity, privacy, and autonomy



© E&HR

# Empirical research ethics in practice

In recent years, **increasing awareness** of research ethical issues in the social, behavioral, educational & economic sciences

- Many resources/collections (e.g. [VerbundFDB \[German\]](#), [RatSWD](#)),
- **Principles and recommendations** by research infrastructures →→
- **Statement by ethics committee** ([list of ethics committees](#)) as part of project application



# Research ethics and data protection

- **Research ethics** involves various norms and guidelines, and is thus not limited to legal norms and regulations
- **Data protection** usually refers statutory requirements
- Yet, data protection law can be seen as reflecting both ethical principles and legal rights (**privacy & autonomy**)
- Many data protection concepts (e.g. **Informed Consent**) also arise from both ethical and legal considerations



Image: Pixabay (CC0)



Image: Pixabay (CC0)

# Data protection: important concepts

## Personal data

Art. 4(1) GDPR

- = any information relating to an identified or identifiable natural person
- Natural (NOT judicial) persons
  - Living (NOT deceased) human beings
  - Not only identified: 'identifiable'
  - Not only direct: 'indirect' identifiers (e.g. 'one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity')

## Processing

Art. 4(2) GDPR

- = any operation or set of operations which is performed on personal data
- catch-all phrase for all operations on personal data (collection, storing, transfer etc.)

# Data protection: important concepts (II)

**Special categories of personal data** → special protection

Art. 9 GDPR

- 'racial' or ethnic origin
- political opinions
- religious or philosophical beliefs
- trade union membership
- genetic data
- biometric data for the purpose of uniquely identifying a natural person
- health
- sex life or sexual orientation

# Data protection: important concepts (III)

## **(Data) Controller**

Art. 4(7) GDPR

= the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data

→ entity responsible for data processing; part of information obligations

## **(Data) Processor**

Art. 4(8) GDPR

= a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller

→ any other entity involved in data processing (e.g. translation service)



# Legislative framework of data protection

- EU law: **General Data Protection Regulation (GDPR)** (German: *Datenschutz-Grundverordnung, DSGVO*) → in force since 25 May 2018
- National law: ‘Federal Data Protection Act’ (*Bundesdatenschutzgesetz, BDSG*) → supplements and specifies the GDPR in some respects
- Subnational law: e.g. *Datenschutzgesetz Nordrhein-Westfalen (DSG NRW)* → (additionally) applies to all public bodies in NRW
- Domain-specific regulations → apply depending on research topic/area

# Main legislative source: GDPR

- Directly binding and applicable in EU countries [& EEA: EU + ISL, LIE, NOR]
- Provides flexibility for certain aspects of the regulation to be adjusted by individual Member States (and subnational authorities)
- Applies if data controller OR data processor OR data subject (= individual) is based in the EU/EEA [research outside EU/EEA: other laws apply too!]
- Also addresses export of personal data to third countries and international organizations

# GDPR: objectives

Art. 1 GDPR + Recitals 1-7

## **Protection of natural persons in relation to processing of personal data**

- Data protection as a fundamental right (provided by Article 8(1) of EU Charter)
- Not an absolute right; must be considered in relation to its function in society and be balanced against other fundamental rights

## **Harmonisation of law for free movement of personal data within the EU**

- Exchanges of personal data have increased and are often necessary
- Harmonization ensures a high level of data protection

→ Natural persons should have **control** of their own personal data

→ Legal & practical **certainty** for natural persons, businesses, public authorities

# GDPR: principles for data processing

Art. 5 GDPR

- Lawfulness, fairness and transparency
- Purpose limitation – BUT: ‘research exemption’ (cf. Art. 89 GDPR)
- Data minimization
- Accuracy
- Storage limitation – BUT: ‘research exemption’ (cf. Art. 89 GDPR)
- Integrity and confidentiality



# GDPR: lawfulness of data processing

Art. 6 GDPR

Data processing is only lawful if and to the extent that at least one of the following applies:

- a) If the data subject has given **consent** to the processing of his or her personal data
- b) To fulfill **contractual obligations** with a data subject, or for tasks at the request of a data subject who is in the process of entering into a contract
- c) To comply with a data controller's **legal obligations**
- d) To protect the **vital interests** of a data subject or another individual
- e) To perform a task in the **public interest** or in official authority
- f) For the **legitimate interests** of a data controller or a third party, unless these interests are overridden by interests of the data subject or her or his rights according to the Charter of Fundamental Rights (especially in the case of children)

# GDPR: lawfulness of data processing

Art. 6 GDPR

Data processing is only lawful if and to the extent that at least one of the following applies:

- a) **If the data subject has given consent to the processing of his or her personal data**
- b) To fulfill contract with the data subject or subject who is acting on behalf of the data subject
- c) To comply with a legal obligation
- d) To protect the vital interests of the data subject or another natural person
- e) To perform a task in the public interest or in official authority
- f) For the legitimate interests of a data controller or a third party, unless these interests are overridden by interests of the data subject or her or his rights according to the Charter of Fundamental Rights (especially in the case of children)

**The main basis for data processing in social research is Informed Consent**

# GDPR: information obligations

Art. 13ff. GDPR

- Identity and contact details of controller (and of data protection officer)
- Purpose of processing and legal bases for processing
- Recipients or categories of recipients
- Period of data storage (or criteria used to determine that period)
- Rights of data subject: access, rectification, erasure, restriction of processing, object to processing
- Right to lodge a complaint with a supervisory authority



Image: Pixabay (CC0)

# Questions?



# Informed Consent

Dear Sir or Madam,

We invite you to participate in our survey.

Your participation in the survey is **voluntary**. Refusing to participate or revoking consent will not result in any disadvantages for you. We operate in accordance with the provisions of the General Data Protection Regulation (GDPR).

**You can assert the following rights at any time:**

- You can have access to the information about your personal data stored by us at any time.
- You can request a correction of this data as well as its deletion at any time.
- You can request a restriction of your data's processing or object to its further processing at any time.

Your data will remain in the project and will not be passed on.

For clarification of queries or if you wish to assert your rights, please contact: [shk.gesis@gesis.org](mailto:shk.gesis@gesis.org).

If you believe that we have not complied with the data protection regulations when processing your data, you can contact the GESIS data protection officers at: [dataprotection@gesis.de](mailto:dataprotection@gesis.de).

By participating in the survey, you agree to these terms.

# Pitfalls in Informed Consent

- Information about the project is missing
- “Your data will remain in the project and will not be passed on.” or  
“Your data will only be used within the research project”
- “By participating in the survey, you agree to these terms.”
- Provide stable contact details!

# Further pitfalls

- No separation between address and survey data: "Your data will be anonymized."
- Contradictory statements: "Your address data will be deleted." but "We will contact you again."
- Ensure appropriate language and scope
- Data sharing and subsequent use of data should already be mentioned in the consent form

# What is Informed Consent

- Obtaining consent to participate and protecting participants from harm
- Consent must be given voluntarily
- Weak consent leads to bad data
  - Creates mistrust
  - Respondents try to protect themselves and their personal data
  - Leads to non-response to questions (sensitive data) or even abandonment of the survey

**Researchers must be able to "demonstrate that the data subject has consented to processing of his or her personal data" (Art. 7.1 GDPR)**

# Information about participants' rights

- **Participants have the right at any time**
  - To withdraw their consent, i.e., to cancel the survey (Art. 7.3 GDPR)
  - To have your data deleted (Art. 17 GDPR)
  - To obtain access to the data and have incorrect data corrected (Art. 16 GDPR)

# Information about the survey and data processing

- Data processor (name and contact information)
- Purpose of data collection
- Recipients or categories of recipients of the data
- Period for which the personal data will be stored

**→ Therefore consent knowing the circumstances is required!**

# Two parts

## Information sheet

- Project and goal
- Involved researchers and contact
- Consequences of participation
- Use of data in the project and beyond
- Protection of data
- Rights of the participants

→ **Remains with participants**

## Consent

- Participating
- Have understood information sheet
- Feel informed
- Consent voluntarily
- Clarification of specifics
- Audio or video recording
- Verbatim quotes are used
- Signature

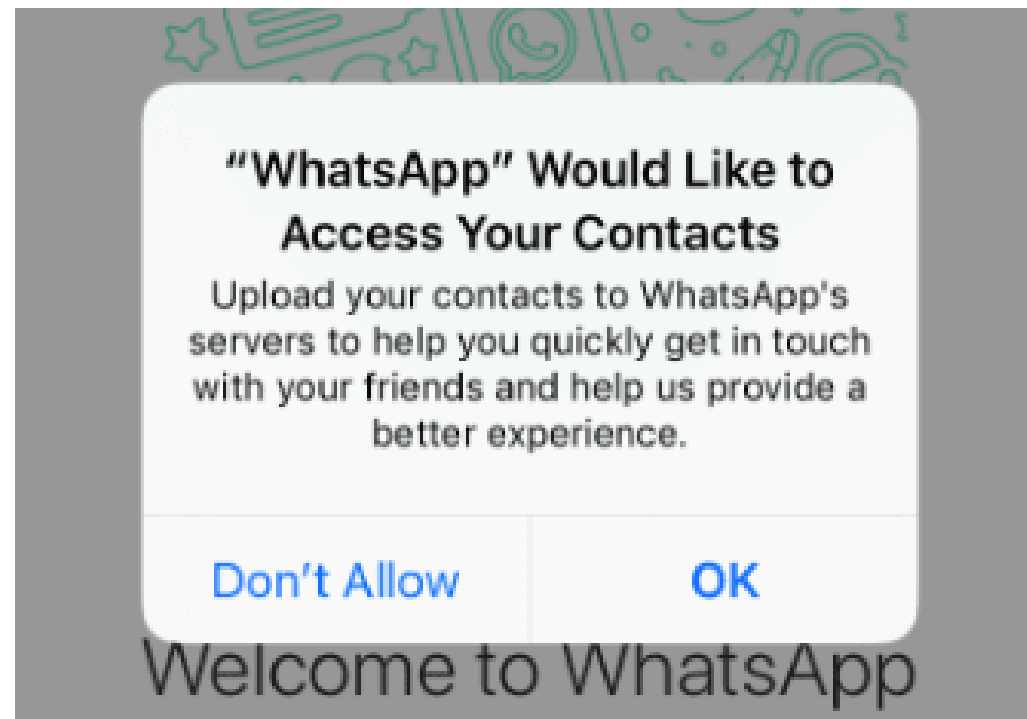
→ **Secure storage, separate from survey data!**



# Special cases

- Telephone and online surveys
- Illiterates
- Children
  - Too young to understand legal situation
  - Parents/guardians and child must always consent
  - Age limit is 16
  - Opening clause between 13 and 16 years (Art. 8 GDPR)
- Sense of duty towards the one who asks (e.g., employers, Federal Office for Migration and Refugees)

# What do you do here?



# How to seek Informed Consent

- How to obtain consent can depend on the nature of the research, the kind of data gathered, the data format, and how the data will be used
- There is no obligation to obtain consent in written form
- But: researchers have to prove that they have obtained consent
  - freely given, specific, informed and unambiguous, clear affirmative action
  - Many information duties; e.g. name/contact details of controller, rights etc.

**Thus, written consent is the ‘gold standard’ and should be obtained whenever possible**

# How to seek Informed Consent (II)

## Advantages and disadvantages of written vs. verbal consent

Written	<ul style="list-style-type: none"><li>- More solid legal ground, e.g. participant agrees to disclose personal information</li><li>- Often required by Ethics Committees</li><li>- Offers more protection for researcher</li><li>- Not possible in some cases: illiteracy, infirm participants, participants wary of formal documentation (illegal activities, refugees), informal research setting</li></ul>
Verbal	<ul style="list-style-type: none"><li>- Can be difficult to make all issues clear</li><li>- Possibly greater risks for researcher</li><li>- Best if recorded (audio or video)</li></ul>

Source: UK Data Service

# When to seek Informed Consent

Obtaining informed consent is not necessarily an on-off occurrence:

**One-off consent** = participant is asked to consent to taking part in the research project only once, typically at the beginning of the project

- *Advantages*: simple; least hassle to participants
- *Disadvantages*: research process and outputs are not always known in advance; participants will not know all the information they will contribute

**Process consent** = the participant's consent takes place throughout the research cycle, e.g. before data collection, before data analysis, and before data archiving

- *Advantages*: ensures active informed consent
- *Disadvantages*: may not get all consent needed before losing contact with participants; repetitive, can annoy participants

# Granular consent

Even if participants are asked to give consent only once, one might offer them the option to consent on a granular level (from CESSDA, modified):

The interviews will be archived at ..... and disseminated so other researchers can reuse this information for research and teaching purposes:	
I agree to the non-anonymized audio recording of my interview being archived and disseminated for reuse	yes/no
I agree to the anonymized transcript of my interview being archived and disseminated for reuse	yes/no

→ *What could be advantages and disadvantages of this?*

# Informed consent – resources

[CESSDA Data Management Expert Guide:](#)

[Informed Consent](#): information, recommendations, examples & exercises etc.

[VerbundFDB: Informierte Einwilligung \[German\]](#):

information, checklists, templates, webinars

[Qualiservice](#): information, templates (German & English), working papers, webinars

[UK Data Service](#): information, templates, examples

## Formulierungsbeispiele für „informierte Einwilligungen“<sup>1</sup>

fbinfo Nr. 4 // November 2018 // Version 2.1

Bitte zitieren als: Verbund Forschungsdaten Bildung (2018): Formulierungsbeispiele für „informierte Einwilligungen“. Version 2.1. fbinfo Nr. 4.

### Vorbemerkungen

Die folgenden Formulierungen dienen als Orientierungshilfe für die Erstellung einer Einwilligungserklärung für empirische Studien. Das schriftliche Einholen des Einverständnisses von Studienteilnehmer/innen geschieht in der Regel mit einem Text bestehend aus drei Elementen: (1) einem Informationsteil, in dem die Studie beschrieben und das Anliegen (Bitte um Einwilligung) vorgetragen wird, (2) den Hinweisen zum Datenschutz, in dem den Betroffenen erläutert wird, welche Arten von Daten erhoben und wie diese Daten verwendet werden sollen und welche Rechte die Betroffenen in diesem Kontext haben sowie (3) der Einverständniserklärung selbst, die von den Betroffenen zu unterschreiben ist.

Die folgenden Formulierungsbeispiele enthalten sowohl optionale als auch verpflichtende Textbausteine. Welche Angaben verpflichtend sind, bestimmt sich nach den jeweils gültigen datenschutzgesetzlichen Regelungen (DS-GVO, BDSG, LDSG oder datenschutzgesetzliche Vorschriften in Spezialgesetzen) und ist daher im jeweiligen Einzelfall zu prüfen. Die nachfolgenden Formulierungen sind Beispiele. Das heißt, sie können sowohl umformuliert als auch unterschiedlich angeordnet und miteinander kombiniert werden. Die Rechtsgültigkeit ist im Einzelfall zu prüfen, denn sie hängt von den konkreten Bedingungen

*Elisabeth Huber & Sabine Imeri*

**Informed consent in ethnographic research:  
A common practice facing new challenges**

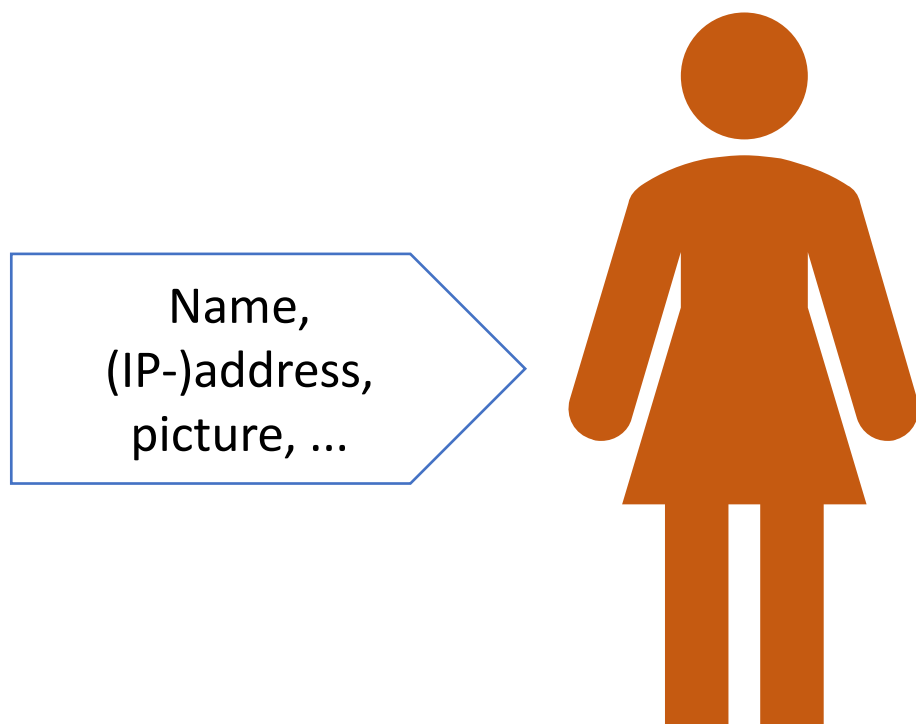
# Questions?



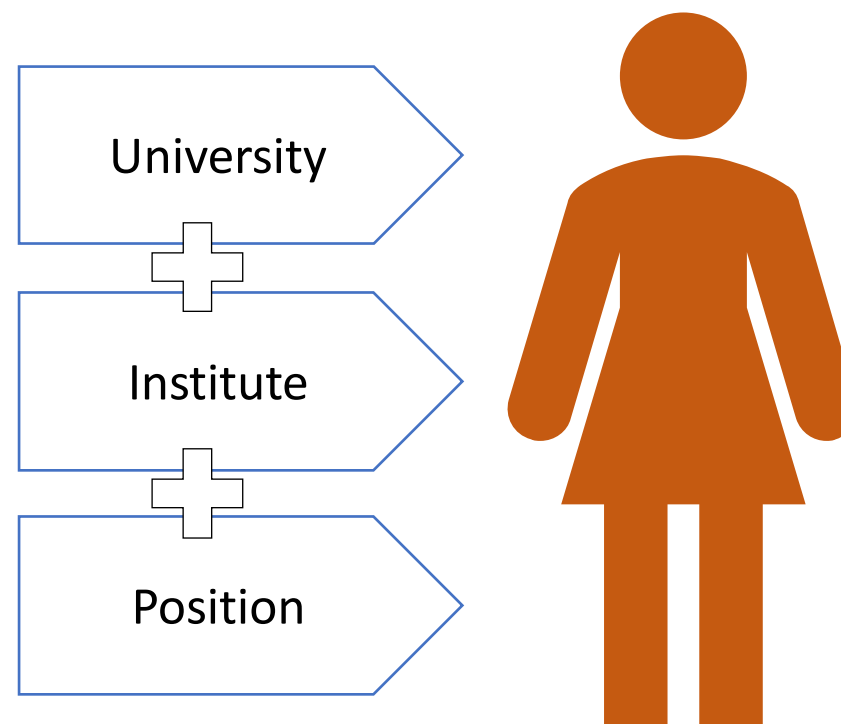
# Anonymization: quantitative data

# Direct and indirect identifiers

## Direct



## Indirect



## Anonymization strategies

Delete direct identifiers or keep them separated from survey data

Delete sensitive data when not needed anymore. Or don't ask them in the first place (Data Minimalism)

Check combinations of characteristics

Top and bottom coding (e.g., number of children, income)

Aggregate data (e.g., federal states, municipal size, occupation)

Document all changes!

# The population matters!

## Ausländer: Deutschland, Stichtag, Geschlecht/Altersjahre/ Familienstand, Ländergruppierungen/Staatsangehörigkeit

Ausländerstatistik Deutschland Ausländer (Anzahl)						
Stichtag Staatsangehörigkeit	Familienstand					
	ledig	verheiratet	verwitwet	geschieden	Lebenspartnerschaft	Lebenspartnerschaft aufgehoben
31.12.2020						
Insgesamt	4 961 395	4 767 230	216 105	612 670	10 365	1 645
davon:						
Ägypten	18 345	14 295	210	1 390	20	5
Äquatorialguinea	85	30	5	5	..	..
Äthiopien	12 495	5 320	215	760	40	..
Afghanistan	166 260	74 165	4 405	2 630	155	10
Albanien	32 870	31 310	360	1 740	30	5
Algerien	7 665	7 835	220	1 715	15	..
Andorra	15	5	..	5	..	..
Angola	3 390	1 290	80	335	5	..
Antigua und Barbuda	10	15	..	..	..	..
Argentinien	2 825	2 980	45	335	40	5

Source: Ausländerstatistik, GENESIS-Datenbank, German Federal Statistical Office

# Excursus: k-anonymity and l-diversity

## Definition

- No fewer than a certain number ( $k$ ) of individuals, with same indirect identifiers ( $k$ -anonymity)
- This group cannot have the same characteristics ( $l$ -diversity)

## Advantage

- Clear and transparent rule, Criteria to determine whether data is anonymized

## Disadvantage

- Information loss

**Used for full censuses and for very visible individuals (e.g., politicians, researchers, figures in the public eye)**

# Exercise: Which information is problematic?

Nr.	Type of information	Text	Continue with
001	Question	Date of birth	002
		__/__/----	
002	Question	Gender	003
		1: Male	
		2: Female	
		9997: Refused	
003	Question	Marital status	004
		1: Married	
		2: Registered civil partnership	
		3: Divorced / registered civil partnership annulled	
		4: Widowed / registered partner deceased	
		5: Single (never married or in registered civil partnership)	
		9998: Don't know	
		9999: Not specified	
004	Question	Are there children under 16 living in your household, and if so, how many?	005
		1: Yes, namely ___ child(ren) under 16 years of age	
		2: No	
		9999: Not specified	
005	Question	In which country were you born?	006
		1: Germany	
		2: Different country, namely:	
		9998: Don't know	
		9999: Not specified	
...	...	...	...

# Anonymising surveyed data

Critical variable	Risk	Implementation
<b>Date of birth</b>	Can make identification easier	Group “year” or “month” and “year”
<b>Marital status</b>	“registered partnership” identifies homosexuality	Aggregate to “married”
<b>Children in household</b>	High number of children can make a person unique	Top coding, e.g., “6 children or more”
<b>Country of origin</b>	Rare country of origin can make a person unique	Group countries

# Questions?



# Anonymization: qualitative data

# Exercise: anonymization of qualitative data

Source for exercise:

<https://dmeg.CESSDA.eu/Data-Management-Expert-Guide/5.-Protect/Anonymisation> (“A practice in anonymising qualitative data”)

Task:

- Read through the interview transcript (including study background)
- Mark the passages that need anonymization/pseudonymization
- Think about how you would anonymize/pseudonymize the data

# Good practices for anonymizing text data

- Do not collect (disclosive) data you don't need
- Plan anonymization/pseudonymization early
- Use pseudonyms/replacements that are consistent
- Use 'search and replace' techniques carefully
- Show where/how you have changed the original text
- Document your steps; create a log of all replacements
- Store unedited/sensitive data separate from anonymized data

# Flexible anonymization of text data

Qualiservice has developed the concept of ‘flexible anonymization’

**Aim:** enabling (better) data re-use while ensuring data protection

- Sensitive personal data are replaced by information (in varying degrees of abstraction) relevant for social science research
- Partial standardization possible through use of hierarchical lists
- Replacements can be ‘opened up’ on different abstraction levels
- Thus, for certain research questions, more specific information can be provided in ‘less’ anonymous data sets (onsite)



# Flexible anonymization – examples

Category	Subcategories	Example	Replacement			
			Original	Level 1	Level 2	Level 3
<b>Locations</b>	Country, region, state, city, street/address etc.	City of Berlin	Location	Location 1, city in Germany	Location 1, big city in Germany	Location 1, big city in Germany with > 1,000,000 inhabitants
<b>Occupations</b>		geriatric nurse	occupation of interviewee	occupation of interviewee, personal care workers (ISCO-08, 53)	occupation of interviewee, personal care workers in health (ISCO-08, 532)	occupation of interviewee, home-based personal care workers (ISCO-08, 5322)
<b>Individual specifics</b>	diseases, awards, accident, ...	massive heart attack	disease	severe disease	cardiovascular disease	severe cardiovascular disease

# QualiAnon – tool for anonymizing text data

- Mark and replace sensitive information
- Semi-automatic; user in control
- Documentation of replacements
- Replacements on case- or study-level
- Integration of (hierarchical) classifications
- At export: select degrees of abstraction
- Separate storage of different file types
- Create, share & re-use lists/classifications



# QualiAnon – tool for anonymizing text data (II)

The screenshot shows the QualiAnon application interface. On the left, there is a 'Categories' panel with a 'Location' dropdown menu. Below it, there are input fields for 'Pseudonym', 'Description', 'Typ', 'Bundesland', and 'Gemeinde'. A 'Replacements' table is also visible, listing original terms and their corresponding pseudonyms.

Originals	Category	Pseudonym	Replacement	#
Katja Meier	Person	Person 1	Female	2
Zwickau	Location	Location 1	Größere Mittels...	4
SPIEGEL ONLINE	Person	Person 2		32

The main text area on the right shows a document with several paragraphs. The first paragraph (5) contains the text: "(5) [Person 1], Jahrgang 1979, stammt aus [Typ=Größere Mittelstadt]. Studium der Politikwissenschaft, Soziologie in Jena, Münster und Estland. Seit 2015 für die Grünen im sächsischen Landtag, bei der Wahl 2019 Spitzenkandidatin. Seit Dezember 2019 ist Meier Justizministerin in der Kenia-Koalition unter CDU-Ministerpräsident Michael Kretschmer." The second paragraph (6) discusses a song by Katja Meier. The third paragraph (7) mentions an interview with SPIEGEL. The fourth paragraph (8) is a question from SPIEGEL. The fifth paragraph (9) is an answer from Meier. The sixth paragraph (10) is another question from SPIEGEL. The seventh paragraph (11) is an answer from Meier. The eighth paragraph (12) is a question from SPIEGEL. The ninth paragraph (13) is an answer from Meier.

Tom Nicolai, Kati Mozygamba, Susanne Kretzer, Betina Hollstein (2021): QualiAnon – Qualiservice tool for anonymizing text data. Qualiservice, University of Bremen. Available at: <https://github.com/pangaea-data-publisher/qualianon>

Register via e-mail: [qualianon@uni-bremen.de](mailto:qualianon@uni-bremen.de)

# Questions?



# Documentation: quantitative data

# What's this data about?

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
1	1.040e+14	1	26	104012	99	1	1973	2	1642	1040	1040	1	6	1	5	5
2	1.040e+14	1	19	104008	99	2	1943	1	99	99	99	1				5
3	1.040e+14	1	98	104003	99	1	1990	1	99	99	99	1				11
4	1.040e+14	1	98	104010	99	1	1983	2	1756	1040	1040	1				11
5	1.040e+14	1	18	104007	99	2	1977	1	99	99	99	1				1
6	1.040e+14	2	19	104007	99					99	99	1	6		5	3
7	1.040e+14	4	15	104005	99					99	99	1	1	2	4	6
8	1.040e+14	4	10	104005	99					99	99	2	1	6	5	88
9	1.040e+14				99	1	1977	2	1040	1276	1040	7	4	1	5	5
10	1.040e+14				99	2	1955	1	99	99	99	1	1	2	5	5
11	1.040e+14				99	2	7777	1	99	99	99	1	1	6	77	77
12	1.040e+14	4	17	104005	99	2	1938	1	99	99	99	1	3	6	5	5
13	1.040e+14	7	17	104005	99	1	1945	1	99	99	99	1	1	6	5	4
14	1.040e+14	4	18	104005	99	2	1949	2	1040	1380	1040	1		6	5	5
15	1.040e+14	2	15	104003	99	2	7777	1	99	99				2	4	5
16	1.040e+14	4	32	104007	99	1	1974	1	99	99				2	5	5
17	1.040e+14	4	98	104006	99	1	7777	1	99	99				3	99	11
18	1.040e+14	4	25	104010	99	2	1968	1	99	99				1	2	5
19	1.040e+14	4	40	104004	99		1967	1	99	99	99	1	1	2	6	2
20	1.040e+14	1	23	104004	99		1932	1	99	99	99	1	1	6	4	3
21	1.040e+14	2	18	104004	99		1965	1	99	99	99	1	6	2	5	5
22	1.040e+14	1	27	104004	99		1956	1	99	99	99	2	1	1	5	2
23	1.040e+14	4	18	104004	99	2	1923	1	99	99	99	1	3	6	3	3
24	1.040e+14	1	19	104006	99	2	1952	1	99	99	99	1	4	2	5	3
25	1.040e+14	3	16	104004	99	2	1947	2	1276	1040	1040	1	1	6	3	3

Who was asked?

How was the survey conducted?

What do the codes mean?

When was asked?

Who asked?

???

# Fukushima disaster: What happened at the nuclear plant?

🕒 10 March

🔗 Fukushima nuclear disaster

Are you for or against nuclear power in Germany?



Image: <https://www.bbc.com/news/world-asia-56252695>

# Findability

## Employed in Times of Corona (May 2020)

[Presse- und Informationsamt der Bundesregierung](#)

GESIS Data Archive, Cologne. ZA7673 Data file Version 1.0.0, <https://doi.org/10.4232/1.13635>

**Abstract:** The Corona crisis (COVID-19) affects a large proportion of companies and freelancers in Germany.

Against this background, the study examines the personal situation and working conditions of employees in Germany in times of corona. The analysis mainly refers to the situation in May 2020 and can only make limited statements about the further situation of the employed persons in the course of the corona pandemic.


**Principal Investigator/ Authoring Entity, Institution:** Presse- und Informationsamt der Bundesregierung -

**Publisher:** GESIS Data Archive

**Study number:** ZA7673

**DOI:** 10.4232/1.13635


**Publication year:** 2020

**Current Version:**  1.0.0, 2020-12-23, <https://doi.org/10.4232/1.13635>


**Date(s) of Data Collection:** 04.05.2020 - 18.05.2020

**Geographic coverage:** Deutschland (DE)


**Universe:** German-speaking population aged 16 and over, working full-time or part-time at the time of the survey

**Analysis Unit:**  Individual

**Sampling Procedure:**  • Non-probability: Quota;

**Mode of Data Collection:**  • Self-administered questionnaire: Web-based (CAWI);

**Temporal Research Design:**  Cross-section

**Kind of Data:**  Numeric

**Data Collector:** Kantar Public Division, Berlin

### Downloads

[Datasets](#)

[Questionnaire](#)

[Other documents](#)

### Actions

[Bookmark](#)

[Cite](#)

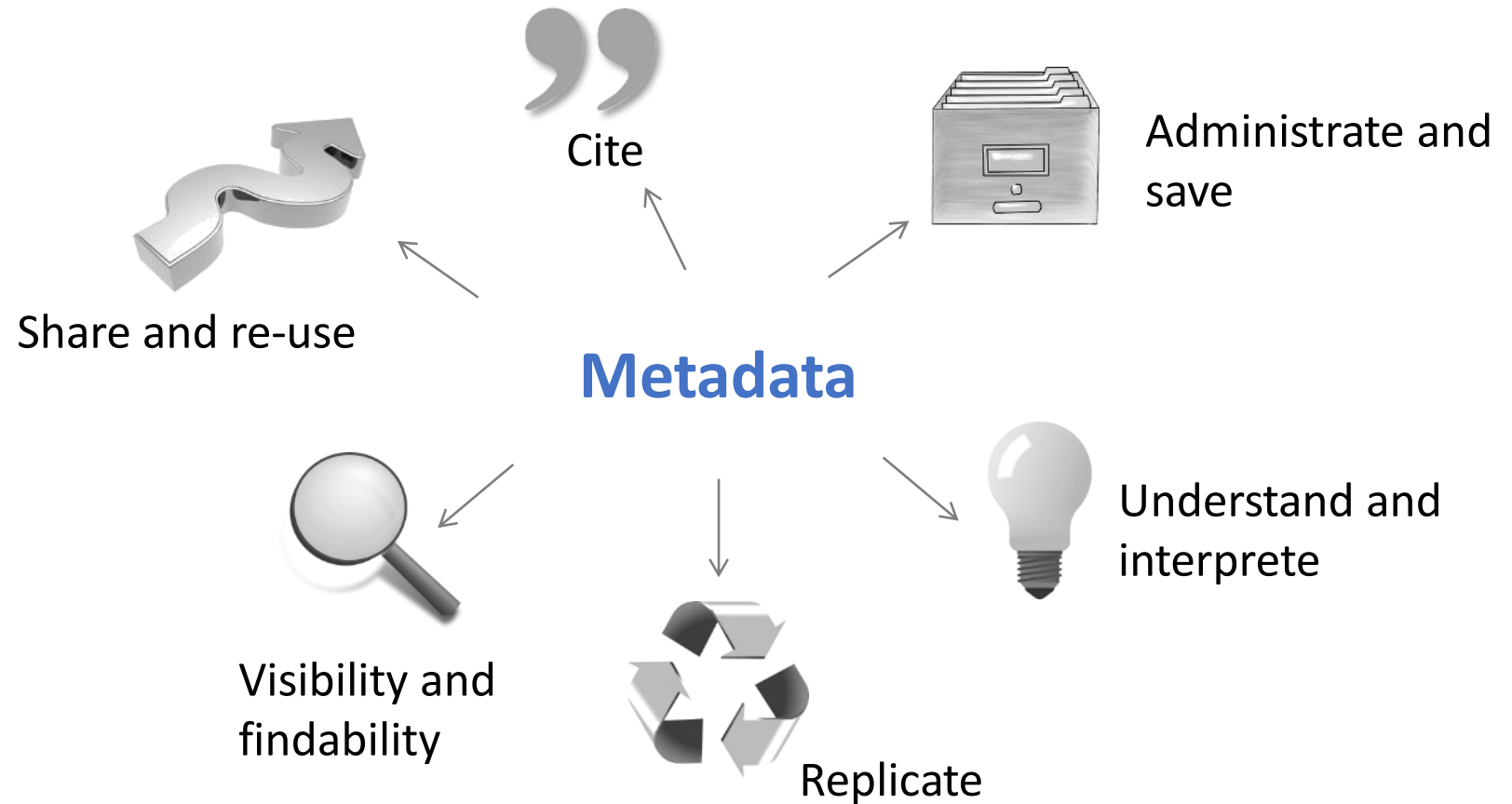
# What do we expect for data publication?

1. Questionnaire
2. Codebook
3. Method report
4. Study description

And everything else that helps to understand the data.

# Metadata

- **Metadata** = Data about data
- Functions:



# FAIR data

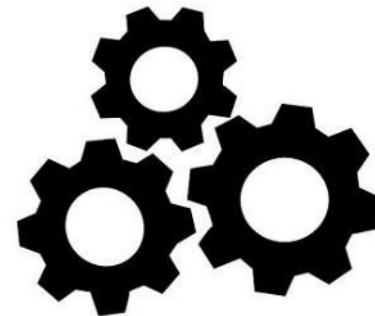
**F**  
Findable



**A**  
Accessible



**I**  
Interoperable

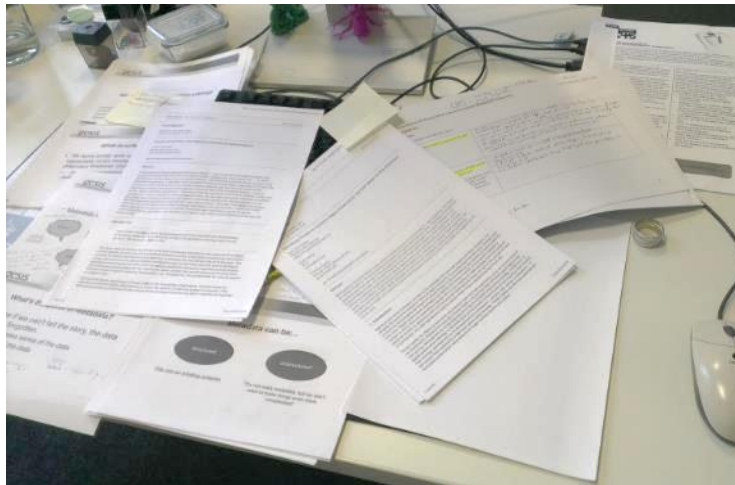


**R**  
Reusable



# Metadata can be...

## unstructured



- Codebooks
- Method report
- Protocols

## structured



- ISCO, ISCED, ISEI (classifications)
- DDI (metadata)



# Metadata standards

- Different standards
  - [Dublin Core](#)
  - [Data Documentation Initiative](#) (DDI)
  - [Statistical Data and Metadata eXchange](#) (SDMX)
- Standardized guidelines on documentation
- Enables exchange and reuse of metadata
- Only indirectly relevant for research projects
- Directly relevant for large survey programs, repositories and data archives

# Questions?

# Documentation: qualitative data

# Qualitative data: metadata standards

- **Qualitative data also need metadata**
- The same **international metadata standards** are used as for quantitative data (e.g. Dublin Core, DDI, DataCite)
- But: metadata standards & controlled vocabularies **sometimes do not ‘fit’** to qualitative data (because they usually were created for quantitative data)

## A Qualitative Data Model for DDI

*Larry Hoyle and the DDI Qualitative Data Working Group: Louise Corti (Chair), Arofan Gregory (Convener), Agustina Martinez (Co-Architect), Joachim Wackerow (Technical Implementation Committee representative), Eirik Alvar, Noemi Betancort Cabrera, Damien Gallagher, Tobias Gebel, Jani Hautamaki, Arja Kuula, Steve McEachern, Cornelia Zuell*

### Abstract

The Qualitative Data Model Working Group was established in January 2010 with the charge “To develop a robust XML-based schema for qualitative data exchange (compliant with DDI) and encourage tools development based upon these needs.” This report describes the preliminary model developed by that group via online meetings, and working meetings in Gothenburg (2011) and Bergen (2012).

This model, described in UML, was developed to cover three main scenarios:

1. Qualitative data collections needing metadata at the object level only
2. Qualitative data collections where segments of objects need to be delineated and described and where segments of different physical representations of the same logical objects possibly need to be linked
3. Qualitative data collections as in the second case where related quantitative data have been generated through techniques such as text mining

### Introduction

The working group began in April of 2010 by gathering a set of use cases and summarizing them in terms of types of qualitative data and types of metadata elements potentially needed. Analysis of the use cases led to enumerating the three main scenarios described above.

At the Gothenburg meeting the group developed a first draft of a model expressed in UML. This model was refined further in the Bergen meeting and is presented here.

### Examples of Qualitative Data

The term “qualitative data” applies to a wide range of collections of digital, and, potentially real, objects. Qualitative data analysis tools can work with text, audio and video recordings, and still images. In developing the metadata model we considered cases where identification of segments of original real objects might need to be linked to some sort of digital representation of the objects. An example might be a collection of preserved archeological artifacts of which multiple digital images were taken. We did not, however, describe any methods by which segments of arbitrary physical objects might be identified. We will need to explore whether others have addressed this issue. One place to start might be techniques used by game hardware to recognize facial features, hands, feet, etc., from video input.

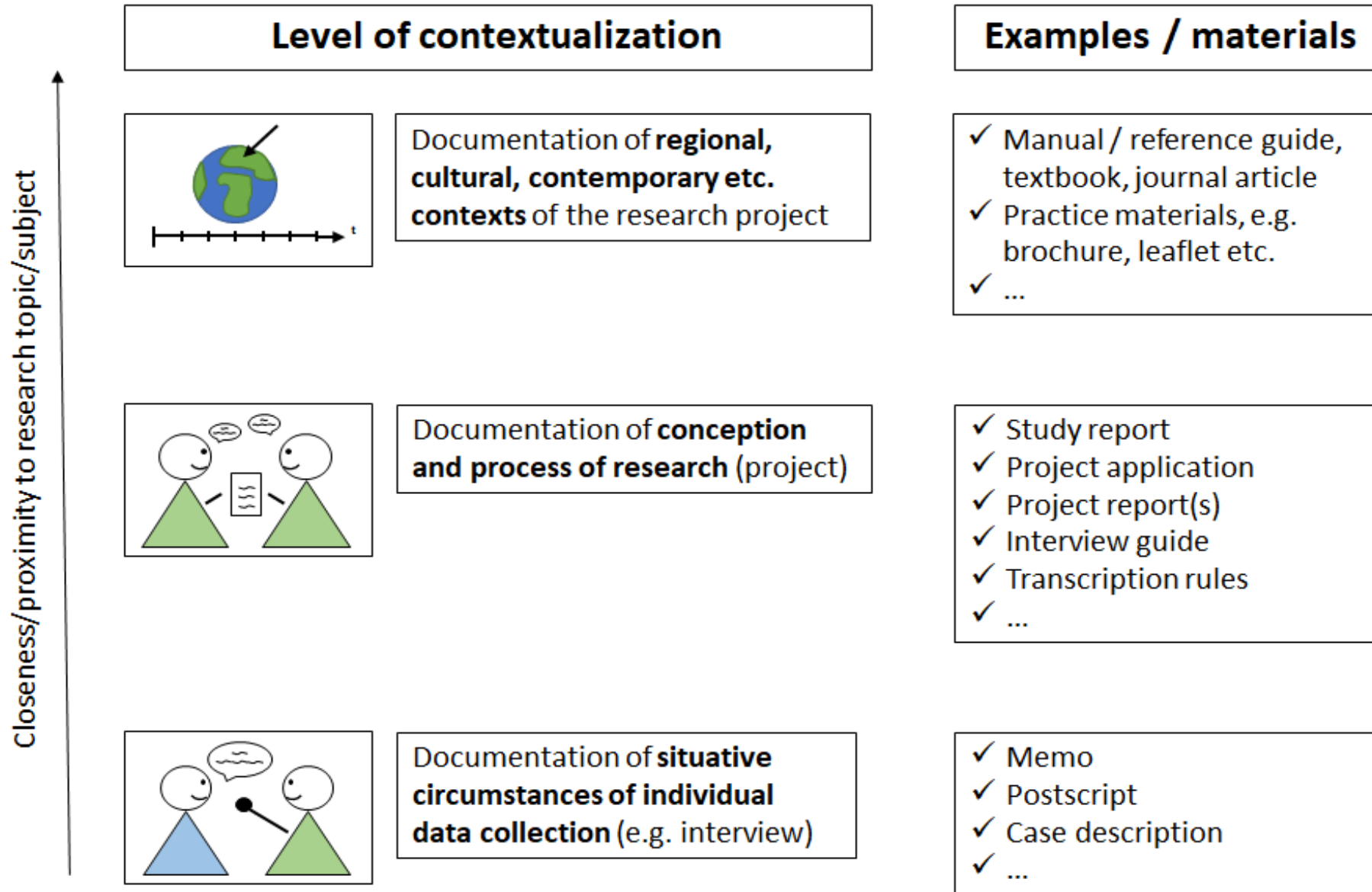
### Use Cases

#### Object Level Only

Both the UK Data Archive and the Finnish Social Science Data Archive offered use cases in which a vector of arbitrary attributes might be used to select a subset of qualitative data objects from an archive. Examples might be selecting all of the audio and transcripts of interviews where an interviewee was a female over 40 years of age, or all of the video clips where the season was winter.

# Documentation: the importance of context(s)

- Qualitative research aims at understanding (= making sense of) actions and statements – which requires context knowledge
- There is not just one context of research data, but various (e.g. institutional, conceptual, methodological, situative etc.)
- These contexts are accessible to the researchers who produce (“collect”) the data – but not so to other researchers
- Documentation thus aims at transferring context knowledge to those not directly involved in data production



Different levels of context & contextualization

# Recommendations for documenting contexts

- Keep (context) materials that help to understand your research & data
- Document your research process as detailed and systematic as possible
- One possibility: writing a ‘Study Report’ about the research & data
- While writing, try to put yourself into the shoes of others and consider which information they would need to (re-)use your research data
- When you archive/share your data, this Study Report can be published
- Most important advice: Document each step while you are doing it!

# In sum: documenting qualitative research

- **Metadata** (standards; for search-/findability etc.)
- **Context materials** (everything that helps to understand your data)
- **Study report** (the core of your documentation, possibly for publication)



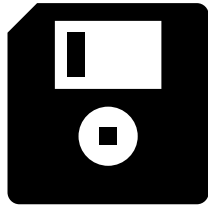


# Secure storage

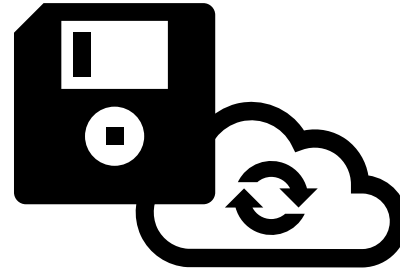
# Your backup?



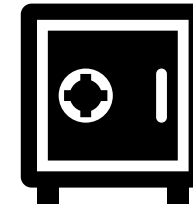
# A good backup strategy



**At short intervals,  
in the same place**



**In (less) short intervals,  
in another place**



**At longer intervals,  
in a very safe place**

# Organizational and technical measures

## Organizational measures

- Separate personal data from research data
- Anonymize/pseudonymize the data
- Store personal data only in countries with similar protection

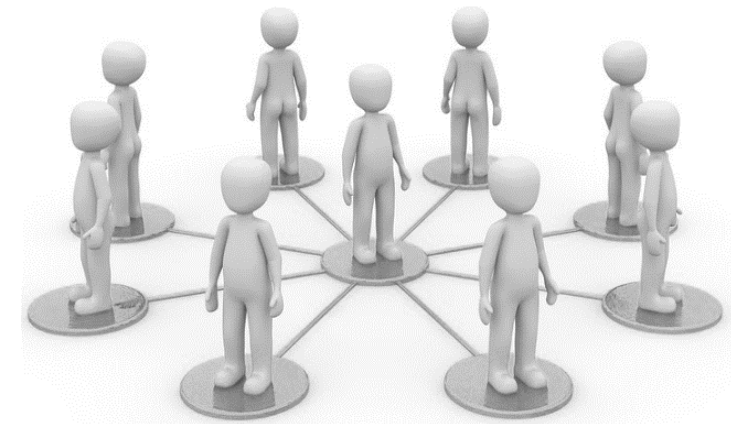
## Technical measures

- Control access to data
- Protect data with a secure password
- Encrypt data
- Destroy data that is no longer needed

How does your backup work?

# Online collaboration using clouds

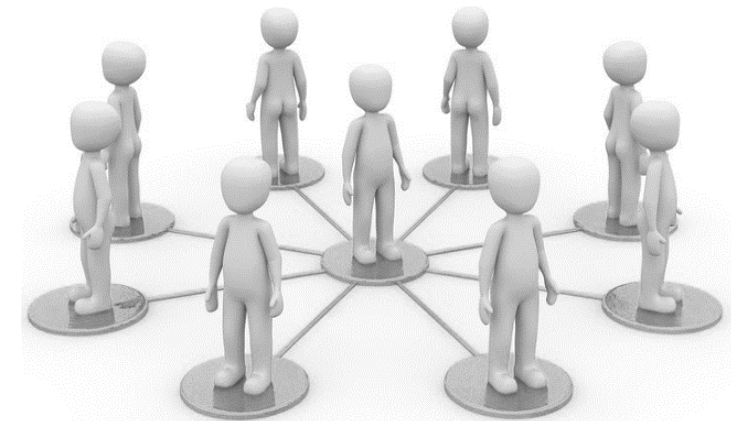
- Critical look at conditions
- Personal data must be stored in the EU or in countries with a similar level of protection
- Email is not a secure way to send sensitive information
- Consider early on how to share data within the project



Picture: Pixabay (CC0)

# Online collaboration using clouds

- Critical look at conditions
- Personal data must be stored in the EU or in countries with a similar level of protection
- **Email is not a secure way to send sensitive information**
- Consider early on how to share data within the project



Picture: Pixabay (CC0)

# Questions?



# Data sharing

# Case study

A group of researchers conducted a survey. A member of the project team leaves the project due to a dispute. However, she still wants to publish a paper with the results of her research.

Is she allowed to copy the data from the university's server to a private hard drive and take it with her? Is she allowed to publish the data?

Can the project lead deny her access to the data after she leaves the project?

**What do you think?**

# What is copyright\*?

- **Protects scientific work, if**
  - Personal, intellectual creation
  - Individuality expressed
- **Does not protect**
  - (research) ideas
  - Facts, e.g., single entries in a database

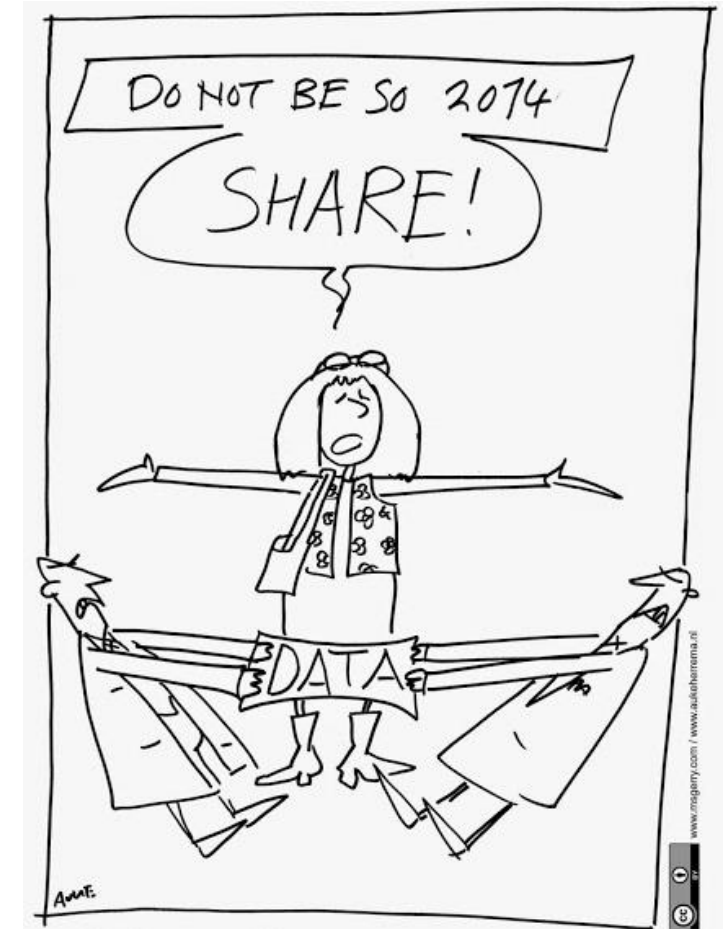
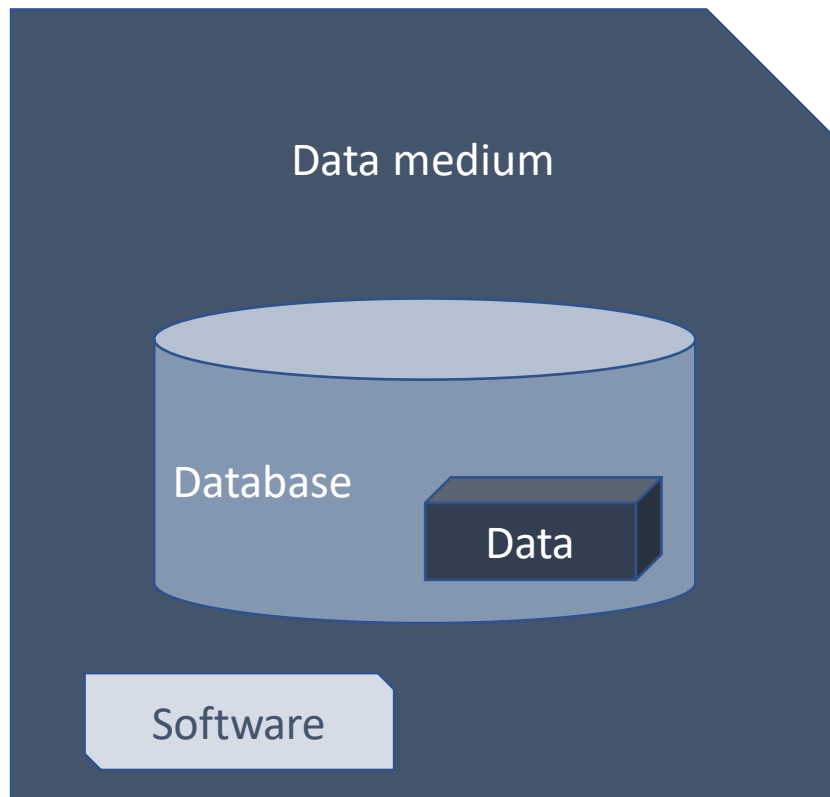


Bild: Herrema (2014) (CC-BY)

\* The following slides reflect the German copyright law

# What about research data?



- **Data**
  - If applicable, protection of works (Werksschutz)
  - But: Not all research data are protected by copyright!
  - If necessary, observe data protection!
- **Database**
  - Rights to the database work (Datenbankwerk, original conception)
  - Database right (Datenbankherstellerrecht, crucial investment)
- **Medium**
  - Property
- **Software**
  - Protection of works (Werksschutz)

# Protection of research data

## Likely protected

Qualitative research data

Data in social science and  
humanities research

## Likely not protected

Quantitative research data

Data in STEM

- No definite answer
- Think of other rules for creative work! E.g., photographs

# Protection of research data

## Protection possible

- Articles, papers, monographs – if not in the public domain yet
- Graphs, maps
- Tables with individual graphic design
- Photographs, videos, infrared and x-rays
- Databases
- Research software

## Protection usually NOT possible

- Facts
- Results of software simulation
- Metadata

Check each **individual case!**

# Who owns the work?

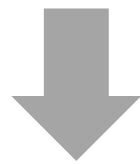
## Researchers

Protected data they created

Databases

Developed software

Private data medium



## University / research institution

Database right

Computers and servers

## Third parties

Rights to re-used data

If applicable, co-authorship through participation (e.g., interviews)

Rights to third-party database works

Rights to re-used software

Data protection does not grant property rights

## Funders

Database right

# Back to the case study

A group of researchers conducted a survey. A member of the project team leaves the project due to a dispute. However, she still wants to publish a paper with the results of her research.

Is she allowed to copy the data from the university's server to a private hard drive and take it with her? Is she allowed to publish the data?

Can the project lead deny her access to the data after she leaves the project?

**What do you think?**



# Back to the case study

Research results	Protection	Owner of rights
<b>Raw data:</b> Replies from respondents	Not protected by copyright(?)	
<b>Database</b> in which these data are entered	<ul style="list-style-type: none"> <li>Database work? Only when creatively/originally conceptualized</li> <li>Database right</li> </ul>	<ul style="list-style-type: none"> <li>Project lead / researcher (creator)</li> <li>University (investor)</li> </ul>
<b>Data medium:</b> Server on which the data is stored.	Ownership of property	University

**Taking away / publication of the data:** Researcher may violate (co-)copyrights of other project participants.

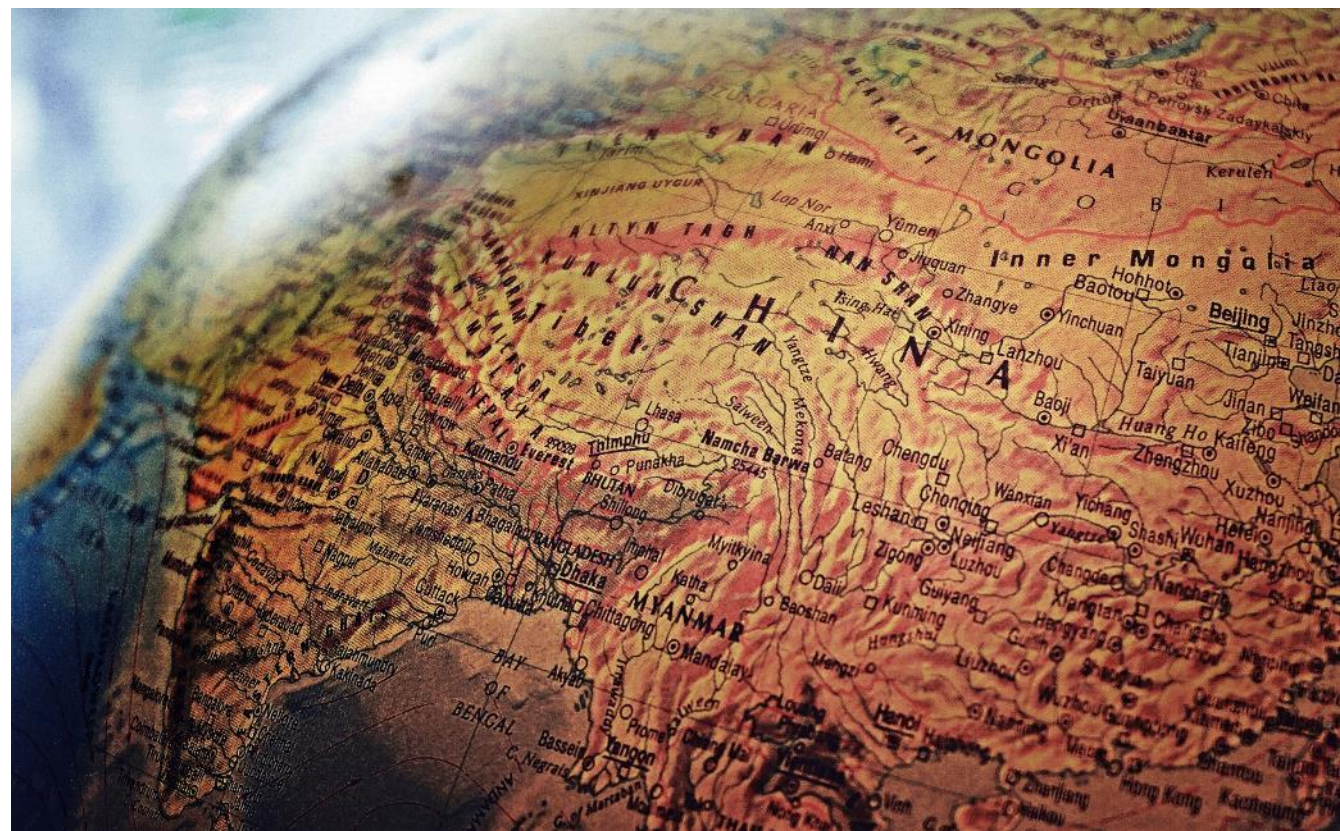
**Denial of access:** Project lead may deny the researcher's right to access data / databases assigned to her by copyright.



- Seek legal advice at an early stage!
- Make contractual agreements at an early stage!

# International copyright

- Not universal!
- Territorial rights:
  - Where they are applied
  - Not where work comes from
- **EU directive: 2001/29/EG**  
(copyright directive) and further directives

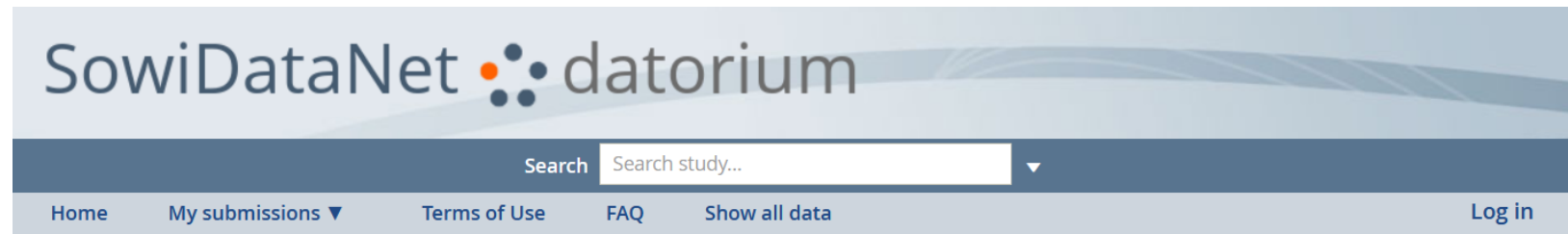


# Questions?

# Licenses

What licenses do you know?

# What is a license?



## Terms of Use

### 3. Legal Framework

On submission, depositors grant GESIS a **non-exclusive right of use** of the digital datasets and documents deposited in SowiDataNet | datorium, in particular

1. the right to systematically archive these datasets and documents and to modify them for the purpose of long-term digital storage and dissemination. When so doing, GESIS may use all appropriate technical means, formats and methods.
2. the right to make datasets and documents publicly accessible on a permanent basis – also for download.

As this does not affect the depositor's copyrights, the data and documents may, for example, also be submitted to other institutions for the purpose of archiving or publication.

On submission to SowiDataNet | datorium, data and additional materials must be free from third-party rights, and the relevant data protection provisions must be respected. In particular, datasets that contain individual-level data (for example survey data) must be anonymised in a way that makes the identification of individuals impossible (de facto anonymity).

GESIS reserves the right to decline to accept data and other materials or to remove already accepted data and materials from the repository should doubts arise as to the legal permissibility of archiving them.

GESIS continues to develop SowiDataNet | datorium. We retain the right to modify the service both technically and organisationally; to cease operations at any time; or to replace SowiDataNet | datorium with another service. In the event of the cessation of operations at

# What is a license?

- Permission to use a right
- E.g., permission to save, adapt, and share

## Commercial licenses

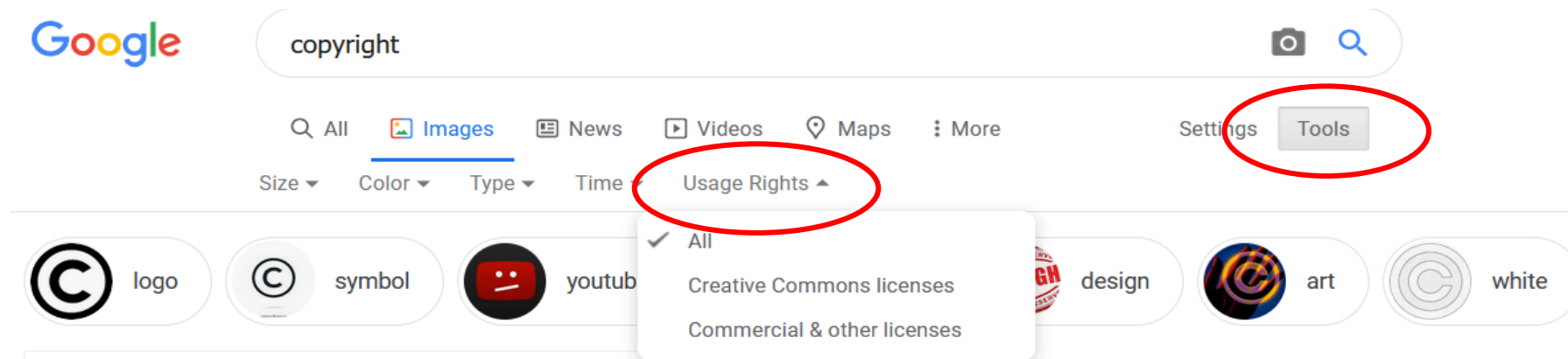
- Allow usage for a fee
- Commercial interests, e.g., publishers
- Examples: Data from a publisher for a fee

## Free licenses

- Allow free usage
- Maximize reuse and visibility
- Examples: CC licenses, ODC licenses

# “If I find it on the internet, I can use it”

- **Wrong!**
- Material that is not licensed is not freely available!
- Ask the author about reuse
- Search for licensed material:





# Exclusive vs. non-exclusive rights

## Non-exclusive rights

- Permits usage in a certain way
- Does not rule out usage by others
- Example: many picture databases, CC licenses

## Exclusive rights

- Usage by excluding everyone else
- One-off fee
- Example: Exclusive rights for a publisher

# “Good” licenses

- Standardized deed
- Do not grant exclusive rights
- Internationally used and compatible with many national laws
- Machine readable



Recommendation



<https://creativecommons.org/>

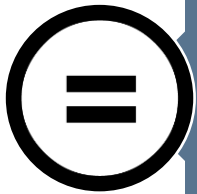
# CC-Licenses



**Attribution (BY):** distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator



**Non-commercial (NC):** distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only

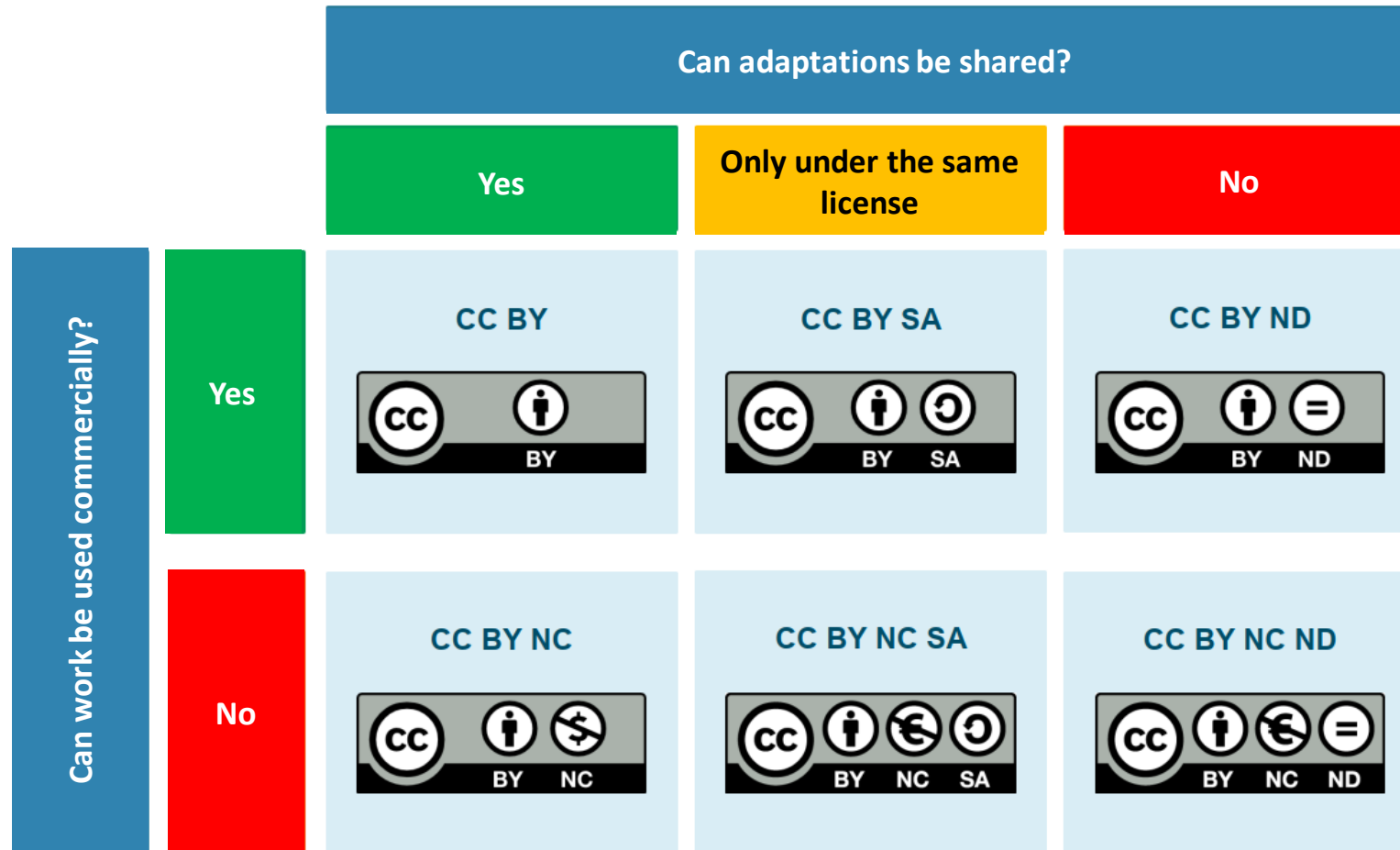


**No Derivative Works (ND):** copy and distribute the material in any medium or format in unadapted form only



**Share-alike (SA):** distribute, remix, adapt, and build upon the material in any medium or format, so long as the modified material is licensed under identical terms

# Choosing the appropriate license



# Reuse of CC licensed material



- Titel:** "Furggelen afterglow"
- Author:** "[Lukas Schlagenhauf](#)" with link to profile page
- Source:** "[Furggelen afterglow](#)" with link to original
- Licence:** "[CC BY-ND 2.0](#)" with link to deed

"[Furggelen afterglow](#)" by [Lukas Schlagenhauf](#) is licensed under [CC BY-ND 2.0](#).

# Public Domain

- Waiver of all copyrights is not possible in Germany
- What's possible:
  - Unconditional licensing or waiver of rights enforcement ("no rights reserved")
  - CC0
- Recommended for data that is not copyrighted
  - Indication that data is in the public domain
  - There are no rights of use to be granted (compared to CC-BY)
  - Public domain mark (no license)
- Good scientific practice requires data to be cited (attribution)



# Open Data Commons

## **Public Domain Dedication and License (PDDL)**

Waives the enforcement  
of the copyright  
(unconditional licensing)

Analog to CC0 / Public Domain

## **Attribution License (ODC-BY)**

Allows sharing and editing  
databases for any purpose

Attribution is a condition

Analog to CC-BY

## **Open Database License (ODC-ODbL)**

Allows sharing and editing  
databases for any purpose

Attribution is required as  
well as provision under the  
same license

Analog to CC-BY-SA

# Licenses and how open they are

Datenlizenz Deutschland

DL-DE->BY-2.0

DL-DE->Zero-2.0

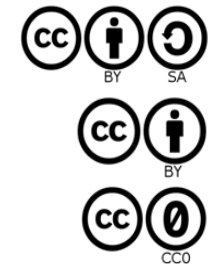
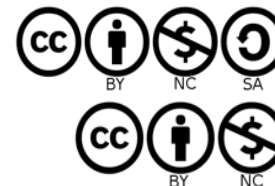
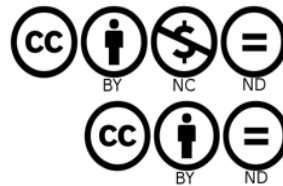
Open Data Commons

ODC-ODbL

ODC-BY

PDDL

CC licenses





# Questions?

# Archiving & sharing data

# Why sharing? – Benefits for ...

## Research

- Transparency and comprehensibility
- For quantitative data: replication
- Secondary analyses (e.g. using other methods, asking other research questions)
- Longitudinal analyses; time comparisons
- Inspiration for new research
- Cross-disciplinary research
- Research about research
- Extended use of data (saving costs/efforts)
- Connecting data to publications

## Teaching & young researchers

- Research methods/methodology in practice
- Learning about research data management
- Learning about legal/ethical aspects of data reuse
- Learning about specifics of secondary analyses
- Using existing data for BA/MA/PhD theses

## Scientific community

- Contribution to discourses on standards and best practices of data management, data protection and research ethics

# Why sharing? – Benefits for your own project

- Increased awareness of RDM practices and standards
- Documentation → knowledge transfer within research team; reflection
- ‘Data/Study Report’: publication; allows detailed methods description
- Long-term data availability (via persistent identifier, e.g. DOI) → increased transparency/comprehensibility and better publishing options in journals
- Continuing use of your data → more citations and academic reputation
- Last not least: funding organizations might require data archiving/sharing

# How to find a(ny) repository



<https://www.re3data.org/>

# Selecting a good repository/archive/RDC

Recommendations for selecting a data repository: [based on [openaire.eu](https://openaire.eu)]

1. Use a disciplinary repository if there is one
2. Alternatively, use the institutional repository [*long-term availability?*]
3. Use a catch-all repository, e.g. Zenodo [*data protection, metadata?*]

→ In contrast to simple repositories, archives or Research Data Centers offer advice/support (e.g. with metadata, sensitive data, usage options), high security & protection, and (possibly) data curation for better quality

# Research Data Centers (accredited by RatSWD)



# Creating good conditions for data sharing

To create good conditions for archiving & sharing your research data:

- If you haven't done so: think deliberately about research data & RDM ✓
- Contact archives/RDCs as early as possible (before project application)  
→ support with RDM planning and RDM cost calculations
- Apply for financial means to cover expenses for data sharing (e.g. fees of RDCs; additional resources within the project)

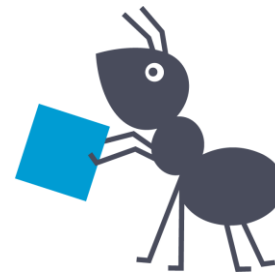


# Building a National Research Data Infrastructure



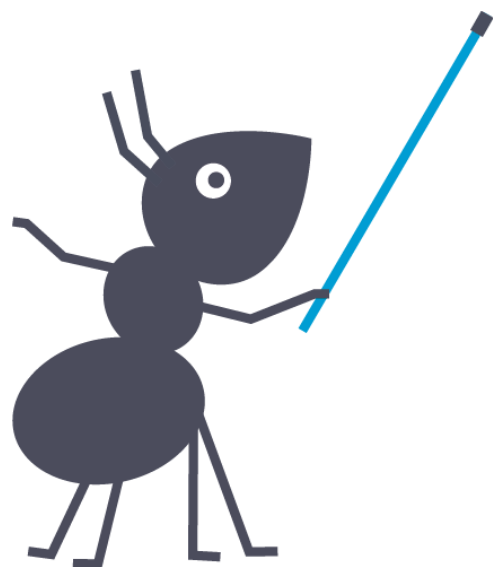
“The National Research Data Infrastructure (NFDI) has the objective to systematically index, edit, interconnect and make available the valuable stock of data from science and research.”

## KonsortSWD



“KonsortSWD provides services around research data in the social, educational, behavioural, and economic sciences.”

# KonsortSWD: services



Automatic coding of <b>open questions</b>	Actively involve <b>professional associations</b>	<b>Q&amp;A Forum</b> for RDC users	<b>RatSWD</b> (German Data Forum)	<b>Open Metadata &amp; Data format</b>	Expand <b>network of partners</b>
Improve <b>findability</b> (Google et al.)	<b>Quality assurance for RDC</b>	<b>RDM Grants</b>	<b>Linking Text</b> and other data types	<b>PID Service</b> below study level	Unified <b>points of access</b> to RDC
<b>Training</b> for RDM Personell	federated archiving <b>qualitative data</b>	<b>Harmonisation</b> of core variables	Support for <b>RDC</b> (CoreTrust Seal, ...)	<b>Interfaces</b> for Data exchange	Support for <b>qualitative RDM</b>

# Questions?

**Thank you!**

## **Kontakt**

**Dr. Anja Perry**

GESIS – Leibniz Institute for the Social Sciences

[anja.perry@gesis.org](mailto:anja.perry@gesis.org)

+49 221 47694-464

**Dr. Jan-Ocko Heuer**

Research Data Center Qualiservice, University  
of Bremen

[jheuer@uni-bremen.de](mailto:jheuer@uni-bremen.de)

+49 421 218-57053

# References

- Brettschneider, P., Biernacka, K., Böker, E., Danker, S. A., Jacob, J., Perry, A., Wiljes, C. & Wuttke, U. (2021). Urheberrecht und Lizenzierung bei Forschungsdaten. Zenodo. <https://doi.org/10.5281/zenodo.5243232>
- CESSDA Training Team (2017 - 2022). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. Available at: <https://dmeg.cessda.eu/>
- Deutsche Gesellschaft für Soziologie (DGS) und Berufsverband Deutscher Soziologinnen und Soziologen (BDS) (2017). Ethik-Kodex. Verfügbar unter: <https://soziologie.de/dgs/ethik/ethik-kodex>
- European Parliament and European Council (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- European Parliament and European Council (2001). Directive (EU) 2001/29/EG (Copyright Directive). Available at : <https://eur-lex.europa.eu/eli/dir/2001/29/oj>
- Flick, U. I. (2009). Qualitative Sozialforschung: Eine Einführung. Hamburg: Rowohlt-Taschenbuch-Verlag.
- Heuer, J.-O.; Kretzer, S.; Mozygemba, K.; Huber, E. & Hollstein, B. (2020). Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung – eine Handreichung für Forschende zur Erstellung eines Studienreports. Qualiservice Working Papers 1-2020, Bremen. <http://dx.doi.org/10.26092/elib/166>
- Huber, E. & Imeri, S. (2021). Informed consent in ethnographic research: A common practice facing new challenges (preprint), Qualiservice Working Papers 4-2021, Bremen. <http://dx.doi.org/10.26092/elib/1070>
- Kretzer, S. (2013). Arbeitspapier zur Konzeptentwicklung der Anonymisierungs-/Pseudonymisierung in Qualiservice. Available at: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47605-2>
- Mozygemba et a. (forthcoming). Anonymisierung und Pseudonymisierung qualitativer textbasierter Forschungsdaten – eine Handreichung. Qualiservice Working Papers, Bremen

# References (cont.)

- Nicolai, T; Mozygemba, K., Kretzer, S., Hollstein, B. (2021): QualiAnon – Qualiservice tool for anonymizing text data (version 1.1.0). Qualiservice. University of Bremen. Software verfügbar unter: <https://github.com/pangaea-data-publisher/qualianon>
- OpenAIRE (2022). Guides for researchers: How to find a trustworthy repository for your data. Available at: <https://www.openaire.eu/find-trustworthy-data-repository>
- Qualiservice (2022): Helpdesk – Data protection. Available at: <https://www.qualiservice.org/en/the-helpdesk/data-protection.html>
- RatSWD (2017). Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften. RatSWD Output 9 (5). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.1>
- RatSWD (2020): Datenerhebung mit neuer Informationstechnologie. Empfehlungen zu Datenqualität und -management, Forschungsethik und Datenschutz. RatSWD Output 6(6). <https://doi.org/10.17620/02671.47>
- RatSWD (2021): Übersicht sozial- und wirtschaftswissenschaftlicher Ethikkommissionen. Stand: 07. Dezember 2021. Available at: <https://www.konsortswd.de/ratswd/themen/forschungsethik/ethikkommissionen/>
- RatSWD (2022): Best-Practice Forschungsethik: Best-Practice-Sammlung zur Forschungsethik im Forschungsprozess. Available at: <https://www.konsortswd.de/ratswd/best-practice-forschungsethik/>
- UK Data Service (2022): Consent for sharing. Available at: <https://ukdataservice.ac.uk/learning-hub/research-data-management/ethical-issues/consent-for-data-sharing/>
- UK Data Service (2022): Anonymising qualitative data. Available at: <https://ukdataservice.ac.uk/learning-hub/research-data-management/anonymisation/anonymising-quantitative-data/>
- Verbund Forschungsdaten Bildung (2022). Ethische Aspekte. Available at: <https://www.forschungsdaten-bildung.de/ethische-aspekte>
- Verbund Forschungsdaten Bildung (2022). Informierte Einwilligung. Available at: <https://www.forschungsdaten-bildung.de/einwilligung>

# Image sources

- BBC (2021). Fukushima disaster: What happened at the nuclear plant? Available at: <https://www.bbc.com/news/world-asia-56252695>
- Brenger, B., Baurmann, B., López, A., Rehwald, S., & Wilms, K. (2017). "Wo sind deine Forschungsdaten in 10-Jahren?!" Awareness für Forschungsdatenspeicherung. Zenodo. <https://doi.org/10.5281/zenodo.1000538> (CC-BY)
- Hazelwood, S. (2017). "Brown World Map Illustration". Pexels. <https://www.pexels.com/de-de/foto/brown-world-map-illustration-1098526/>. (Pexel License)
- Herrema, A. (2014). FOSTER Cartoon: Don't be so 2014 – Share. FOSTER-Project (EU-Funding 612425). Available at: <https://www.fosteropenscience.eu/foster-taxonomy/research-data-management>. (CC-BY).
- A. Herrema & H. Bouwteam (2014). FOSTER Cartoon: Data? – Over there. FOSTER-Project (EU-Funding 612425). Available at: <https://www.fosteropenscience.eu/foster-taxonomy/research-data-management>. (CC-BY).
- [SangyaPundir](#) (2016). FAIR guiding principles for data resources. Available at: [https://commons.wikimedia.org/wiki/File:FAIR\\_data\\_principles.jpg](https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg). (CC-BY-SA).
- German Federal Statistical Office (2022). Ausländerstatistik. Available in GENESIS-Online: <https://www-genesis.destatis.de/genesis/online?sequenz=statistikTabellen&selectionname=12521#abreadcrumb>

# Data upload at GESIS

SowiDataNet  datorium

<https://data.gesis.org/sharing>

## Register

Registration at GESIS is free of charge, open to all and gives you access to various GESIS services. You can use your GESIS Account to log in to the following GESIS services:

- [GESIS search](#) (Download of research data)
- [GESIS Notebooks](#) (Virtual research environment)
- [GESIS-Training](#) (Scientific training events)
- [SowiDataNet|datorium](#) (Research data repository for the social and economic sciences)

\*Mandatory fields

**First name\***

**Last name\***

**Country\***

**Discipline\***



# Only 4 mandatory fields

Basic Information

Title \*

Other title Other title type +

Primary researcher \*

Name Institution +

Publisher \* +

Publication year \* 2022

Availability \*

# But it makes sense...

-
Content

**Subject area**

Please fill in
+

**Topic classification**

▼

**Abstract**

Please fill in

**Internal comment on content**

Please fill in

... to add more metadata...

Methodology

Geographical location - State, Country (according to ISO-3166)

Geographical location - Region

Temporal research design

Universe

Selection method

Data collection mode

Survey period date format

# ... to understand the data better.

## + Additional Notes

**New feature:** It is now possible to assign the availability "Free access (without registration)" to supporting materials such as codebooks, questionnaires, or syntax files, even if the availability of the data itself is more restricted.  
To activate this option, simply check the box "Set a different availability for this file" after uploading a resource in the area below.

### Files (max. 1 GB)

If you wish to upload files bigger than 1 GB, please contact [curator2@gesis.org](mailto:curator2@gesis.org).

Drop files here to upload.

Click to Upload

Save

Review and Submit

# Further data services



We make your data **FAIR**. The data is **described in a standardized way, secured for the long term**, provided with a permanent identifier (**DOI**), and can be **easily found and reused** through browser-optimized catalogs.



## Archiving BASIS

Free storage (bitstream preservation) and publication of your data via our repository SowiDataNet/datorium

## How-to Guide Data Sharing

Instructions for data depositors - procedure for securing, documenting and publishing your data

## Archiving PLUS

Chargeable storage (long-term digital preservation), description and publication of your data

## Archiving PREMIUM

Chargeable storage (long-term digital preservation), quality assurance, description and publication of your data