GDPR and it's implementation in data sharing in social sciences

Legal and ethical issues in data management and open science

***Anne-Mette Somby**, CESSDA ERIC*

CESSDA Training Round table,11th of December 2019, Belgrade

cessda.eu @CESSDA_Data

My background

- Political and social sciences.
- At NSD- Norwegian Centre for Research data
- My position now: a senior research advisor at hvl.no/en

IASSIST sponsors this event

- IASSIST is an international organization of professionals working in and with information technology and data services to support research and teaching in the social sciences.
- Its 300 members are from a variety of workplaces, including data archives, statistical agencies, research centres, libraries, academic departments, government departments, and non-profit organizations.

From: https://iassistdata.org/

Content of my talk is based on the workshop we did with Alexandra at ESRA 2019 conference.

NSD

FORS
explore.understand.share.

Anne-mette.somby@hvl.no

Alexandra.stam@fors.unil.ch

Open science concepts

- Open data
- Open publications
- Transparency – societal benefit- economic interest
- And more…

The general Data Protection regulation

- The GDPR applies from the 25 May 2018.
- The GDPR applies to any data controller or data processor in the EU who collects personal data about a data subject of any country, anywhere in the world.
- A data controller or data processor that is based outside the EU but collects personal data on EU citizens will also be covered by the GDPR.
- This means that a researcher (data controller) based within the EU who collects personal data about a participant, from any other country within the EU, or the world, needs to comply with the GDPR.
- Also means a researcher (data controller) outside the EU who collects personal data about a participant in the EU will be covered when this relates to offering goods/services or the monitoring of their behavior within the EU.

GDPR art. 5 – rules and exemptions for research

Personal data must be:

a) Processed lawfully, fairly and in a transparent manner

b) Collected for specific purposes and not processed further for incompatible purposes (purpose limitation) – exemption for research/archiving: further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes

c) AdeAccurate and where necessary up to date

d) quate, relevant and limited to what is necessary – (Data minimisation)

e) Kept in identifiable form no longer than necessary - exemption for research/archiving personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1)

f) Processed with appropriate security – integrity and confidentiality

Experiences from Norway, Germany and UK

- UK: A change in legal basis from consent to public task. Revisiting what anonymization means in research.
- Germany: Uncertainty in the interpretation of anonymity, more involvement of data protection officers or legal advisors at universities. More explicit consent covering data archiving.
- Norway: NSD still provides assessment for most research. GDPR has resulted in a growth in the numbers of approx. 60 percent. Universities focus on training for their staff. Consent as the legal basis for most research.

Open data

- New requirements – sometimes in contradiction
- Increasingly complex research landscape
- Special care needs to be given to sensitive and personal data
- We are here to bridge the gap between law and research practice

Special categories of personal data

- Racial or ethnic origin
- Religious opinion, political and philosophical beliefs
- Health
- Trade union membership
- Sexual orientation or sex life

- Demands higher protection and documentation of societal benefit
- Data Protection Impact Assessment

Legal basis for research

All processing of personal data requires legal basis.
The most common for <u>research</u> are:

- (Article 6): a) *consent* , e) necessary for the performance of a task carried out in the *public interest*
- Special categories of data (Article 9)
- Prohibited unless: a) *explicit consent* e) personal data are manifestly made public by the data subject j) *necessary for archiving, scientific or statistical purposes*

Informed consent for research

- Freely given: must be a genuine choice, be able to refuse/withdraw without consequences, not be in a dependent relationship
- Specific - clear information on extent and consequences
- Informed: Content and form requirements, should be easily understood, easily accessible, clear and simple language, especially when the information is given to children
- Active: “opt in” - silence, pre-ticked boxes, and inactivity are not valid
- It should be as easy to withdraw consent as to give it
- Controller shall be able to demonstrate that consent is given (Article 7)

Documenting consent

- Under the GDPR, consent needs to be documented, which means (in the context of research) it will be important for researchers to maintain documented and accurate records of the consent obtained from their participants.

- This could, for example, be written consent or audio recorded oral consent.

How to fulfill the data subjects right to information ?

- Use a template if you find one that suits your project.
- Information should be given individually.
- If not possible you can give it to a larger population on your website.
- If the population is too large you can argue an that it’s impossible to fulfill this right.

Data Protection Impact Assessment (DPIA) when:

- Article 35 defines when DPIA is required, what it should contain and who will implement it.
- New obligation with the GDPR.
- Ensure safeguards for the registered.
- When likely to result in a high risk to the rights and freedoms of natural persons.
- The controller shall seek the advice of the data protection officer when carrying out a DPIA.

Strategy for Sharing Data

- Adding the discussion of data sharing and archiving permits the participant to make an informed decision. This empowers them and puts them in charge of choosing whether they wish for their contribution to the research project – and their data – to be available for use in future research projects.

- The best way to achieve informed consent for data sharing is to **identify** and **explain** the possible **future uses** of their data and offer the participant the option to consent on **a granular** level.

Strategy for Sharing Data example

- For example, in a qualitative study, this may involve allowing the participant to consent to data sharing of the anonymized transcripts, the non-anonymized audio recordings and the photographs.

- Discuss ethical and juridical implications from collection to dissemination.

Collaboration across institutions and countries

Collaboration with countries outside EEA

- Data controllers and data processors
- Requirements for information security when sharing personal data
- EU data protection rules apply to the European Economic Area (EEA), which includes all EU countries and non-EU countries Iceland, Liechtenstein and Norway.

Basic concepts

- Data controller:

The institution that decides the use of the data

- Data processor:

The institution that process data on behalf of the controller.

You will need be a written data processor agreement between the two institutions.

Consent as legal basis?

- Consent is the most common legal basis for research in some countries (In Norway this is the main rule). This is different from Finland, Sweden and Denmark where the most common legal basis for processing personal data is public interest (art. 6.1 e and art. 9. 2 j)

- In Norway public interest is most commonly used as legal basis for:
  - Data collected on the internet
  - Public registries
  - Data collected in anthropology studies/fieldwork

Other legal basis

- *Public interest* as legal basis is common in many EU countries. This means that you have to ensure an “ethical consent” when needed and possible.

- You will also have to ensure the *rights* of the data subjects (transparency (art. 12), **information (art. 13**), **access (art. 15),** rectification (art. 16), **erasure (art. 17),** restriction of processing (art. 18), notification (art. 19), data portability (art. 20)) *when possible and relevant*.

Collaboration inside /outside EU/EEA

- You can use the same DPIA in different EU/EEA countries, but there may be different practices.
- If you don’t’ share personal data you don’t need to agree on a common DPIA.
- Meet and discuss legal basis and how to gain consent.
- Use a DMP.
- Check national legislation (e.g age for consent will be different).

EU data protection rules apply to the European Economic Area (EEA), which includes all EU countries and non-EU countries Iceland, Liechtenstein and Norway.

Collaboration with countries not in the EEA

- Standard EU agreement on transfer of personal data outside the EEA.
- Agreement between institutions on securing personal data
- Meet and discuss…
- Check national legislation
- DPIA is needed for EU/EEA institutions

Requirements for information security when sharing personal data

- Your institution will need to provide guidelines and infrastructure for handling personal data.
- The institution can also give advise on how to share personal data in a secure way.
- It is (probably) not accepted to share personal data on open clouds, e-mails and google doc.
- (These guidelines will have to meet the demands from national legislators).

Best Practice for Legal Compliance 1

- Investigate early which laws apply to your project.
- Do not collect personal and sensitive data if not relevant to your project.
- Seek advice from your research office
- Meet your partners from other institutions
- If you need personal and sensitive data:
  - Inform and get consent from the participants
  - You may need an ethical and legal assessment

Best Practice for Legal Compliance 2

- You may need to perform a DPIA.
- Use a DMP.
- Obtain informed consent, also for data sharing and preservation or curation.
- Protect identities e.g. anonymization and not collecting personal data if not necessary.
- Regulate access where needed (all or part of data) e.g. by group, use or time period.
- Securely store and protect personal and sensitive data.

Data Management and Open Science

How can we make data FAIR?

1. New requirements

- From funders:
  - data management plans (DMPs)
  - data sharing (in FAIR repositories)
- From journals:
  - deposit of data used in publications
  - sufficient documentation

Hvorfor datahåndteringsplan?

Findable

Accessible

Interoperable

Reusable

NSD
NORSK SENTER FOR FORSKNINGSDATA

2. Researcher perspective

- digitalisation of data: more and more data are produced;
- new research fields, including new types of data (Big Data);
- facilitated access to data by the community;
- new analytical/data extraction tools
- ‘contradictory’ forces: protection and openness

Different levels of resistance

- rethinking relation to data
- rethinking practices in relation to new requirements, including data sharing
- looking at ethical and data protection issues

across disciplines
across methodologies
across cultures

3. Data Management

Data management can be defined as a "process by which the data are acquired, validated, stored, protected, and processed, and by which its accessibility, reliability, and timeliness is ensured to satisfy the needs of the data users".*

*Source: http://www.businessdictionary.com/definition/data-management.html

Data Management Plan

1. Building a data management strategy

- **DMP**

2. Instrument design and data collection

- **Informed consent**

- **Documentation**
- **Data security**
- **File and folder organisation**

3. Data preparation

- **Transcription**
- **Anonymisation**
- **Quality checks**

4. Data analysis and publications

5. Data preservation

- **Long term policy**
- **Formats**

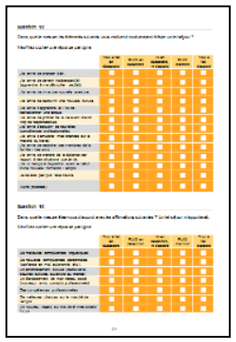
6. Data sharing

- **Data deposit**
- **Data promotion**

What data to share?

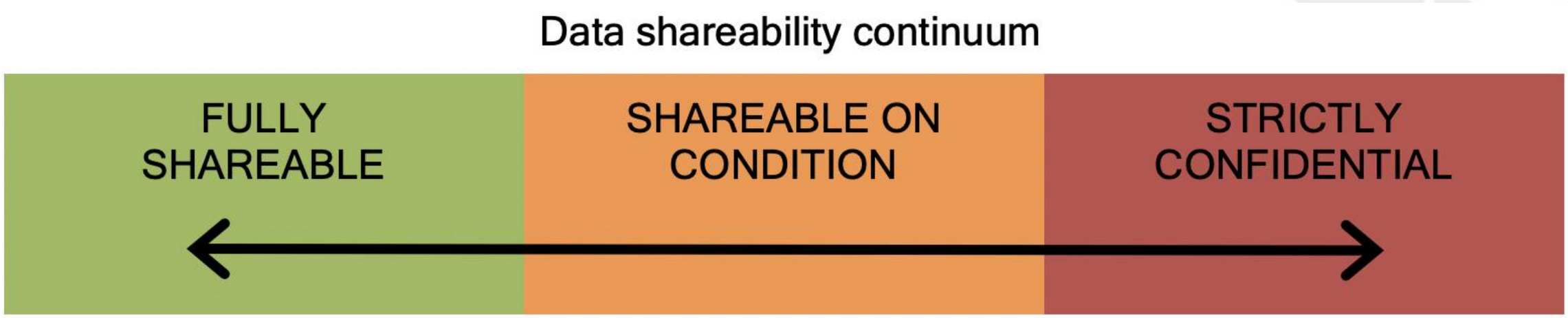
- Research data and related documentation

Research data can be defined as «recorded factual material commonly *retained by and accepted in the scientific community as necessary to document and validate research findings*"[1]

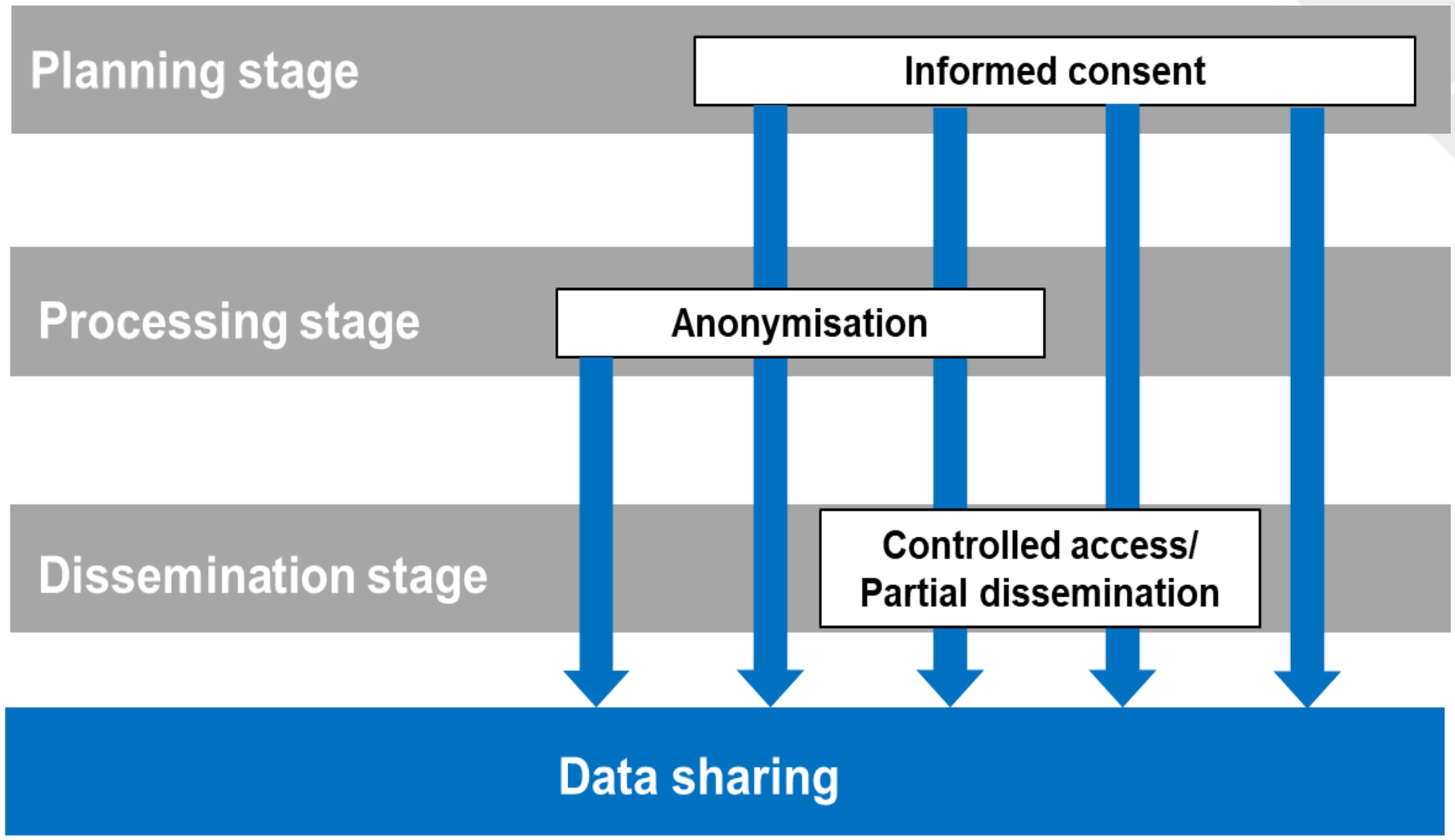




The collection of personal and sensitive data is not a reason not to share.

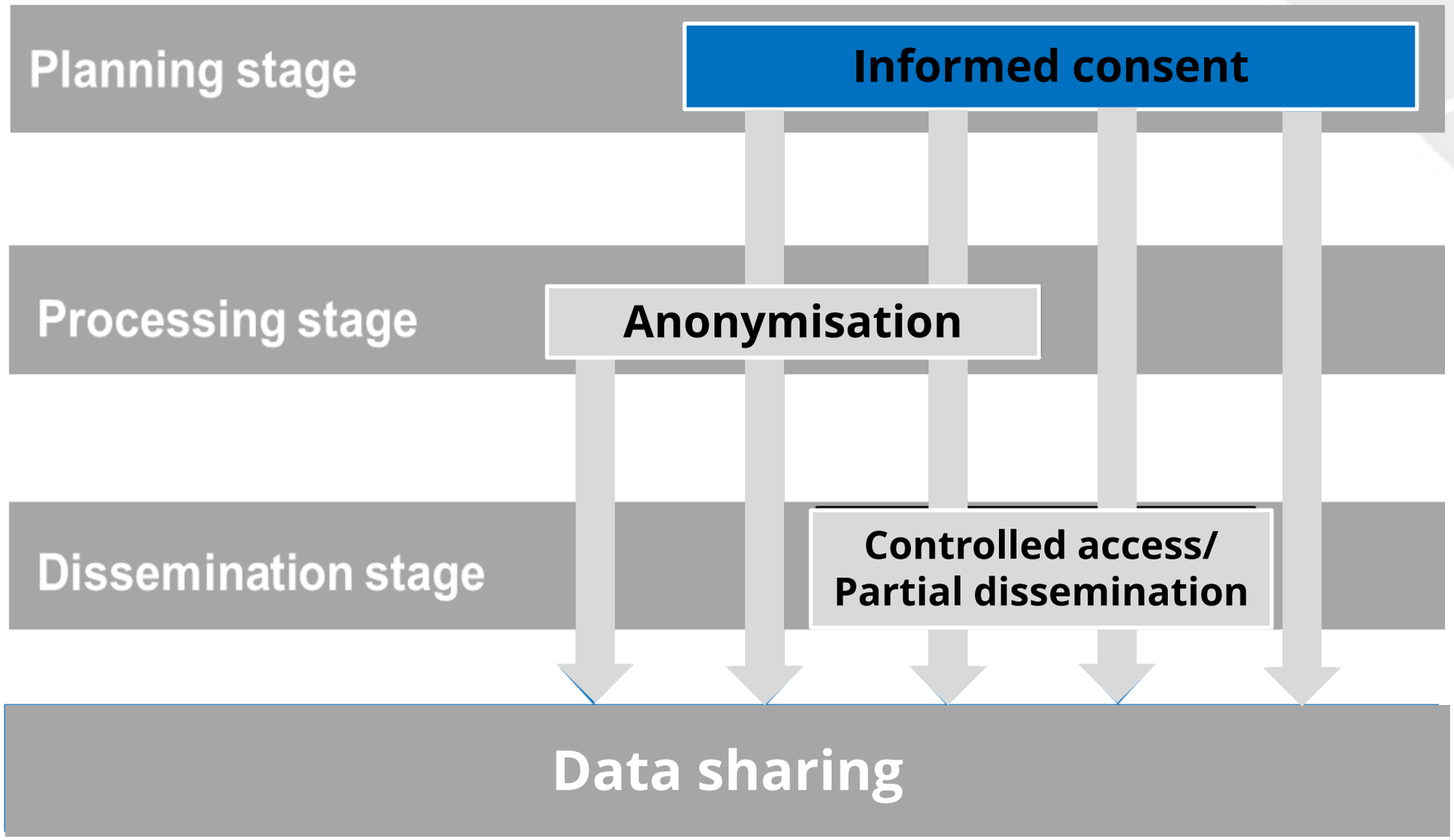
[1] Definition adapted from https://www2.le.ac.uk/services/research-data/rdm/what-is-rdm/research-data


Data shareability continuum
FULLY SHAREABLE
SHAREABLE ON CONDITION
STRICTLY CONFIDENTIAL

Key practices to protect respondents



Informed consent

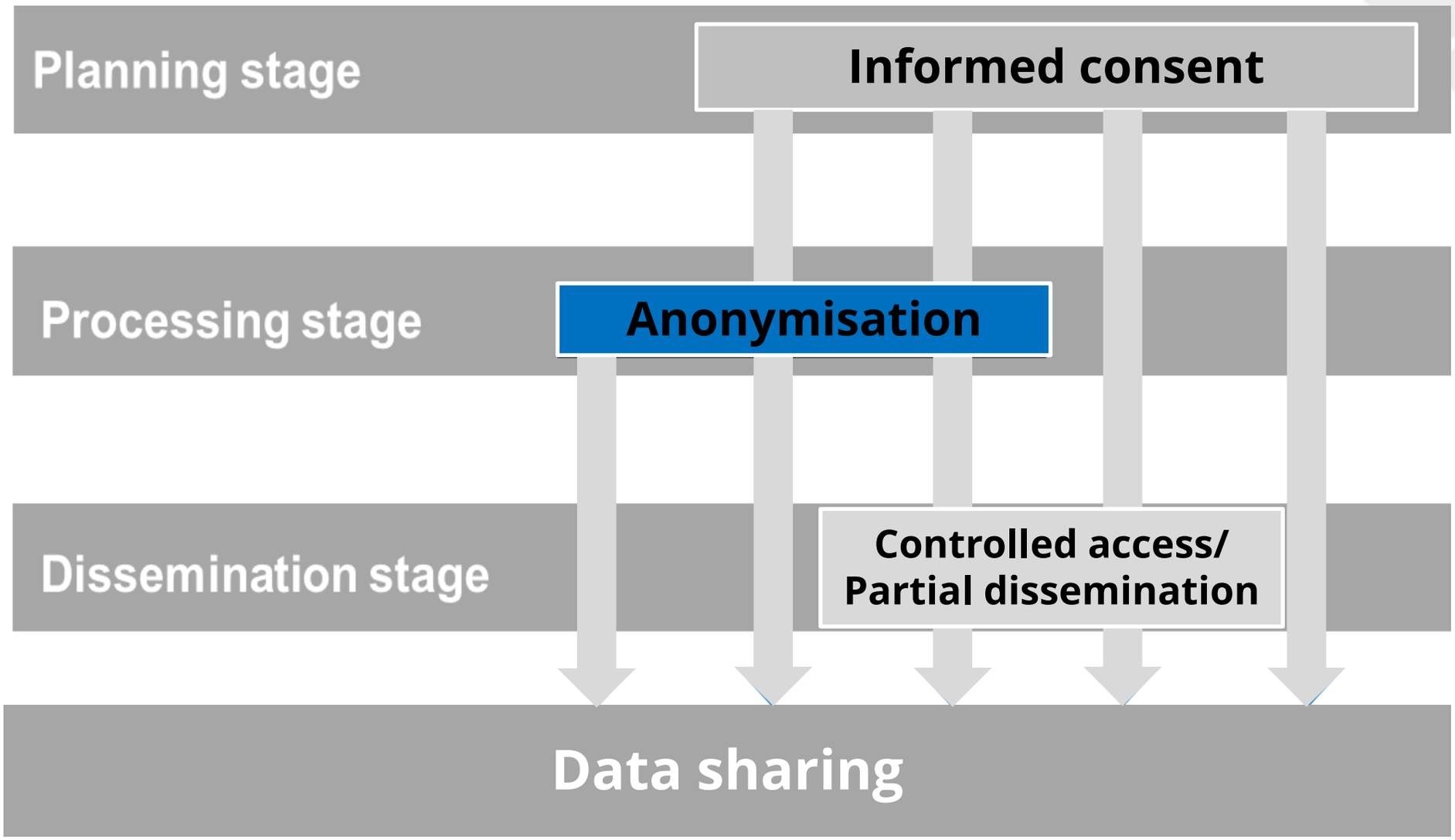


Some ethical reflections

Anyone wishing to analyse or archive and share non-anonymised personal/sensitive data must ensure that individuals have freely and properly given their consent. This raises a number of ethical considerations with respect to:

- The original purpose of gaining consent
- The “truly informed” and “ free” nature of consent
- Exceptions to consent

Anonymisation



How can identity be disclosed?

A person's identity can be disclosed through identifying information: a value may, possibly in combination with other values, lead to (re)identification.

Identifying information may consist of:

- Direct identifiers (e.g., name, address, telephone number, voice, picture, bank account number, social security number, …);
- Indirect identifiers – possible disclosure in combination with other information (e.g., occupation, geography, unique or exceptional values or characteristics, …)

Some key steps for quantitative anonymization:

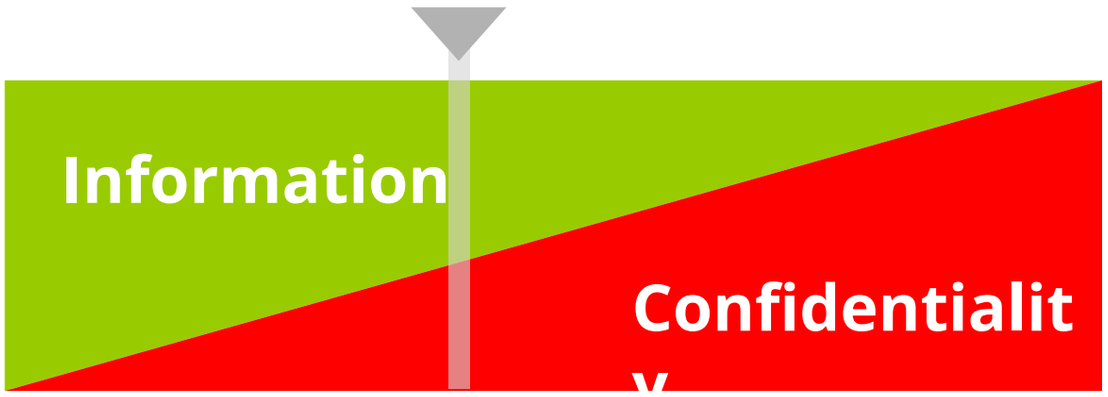
- Remove direct identifiers (e.g., names, address, institution, photo).
- If necessary, limit the number of identifying variables.
- Reduce the precision / detail of a variable through aggregation and/or rounding (e.g., birth date, educational categories, replace a value with a less precise value).
- Restrict upper and/or lower ranges of a variable to hide outliers (e.g., income, age) (winsorization).
- Combine variables (e.g., create a non-disclosive rural/urban variable from place variables).
- Alter values (substitute values, suppress values).
- Generalize meaning of detailed text variables (e.g., occupational expertise)
- Watch out for open text (replace names with pseudonyms).

Some extra steps for qualitative anonymization:

- Do not collect disclosive data unless necessary
- Plan or apply editing at the time of transcription
- Where possible replace rather than remove
- Avoid blanking out: use pseudonyms or replacements
- Be consistent throughout the project
- Identify replacements (e.g. with brackets)
- Keep an anonymization log of all replacements, aggregations or removal made

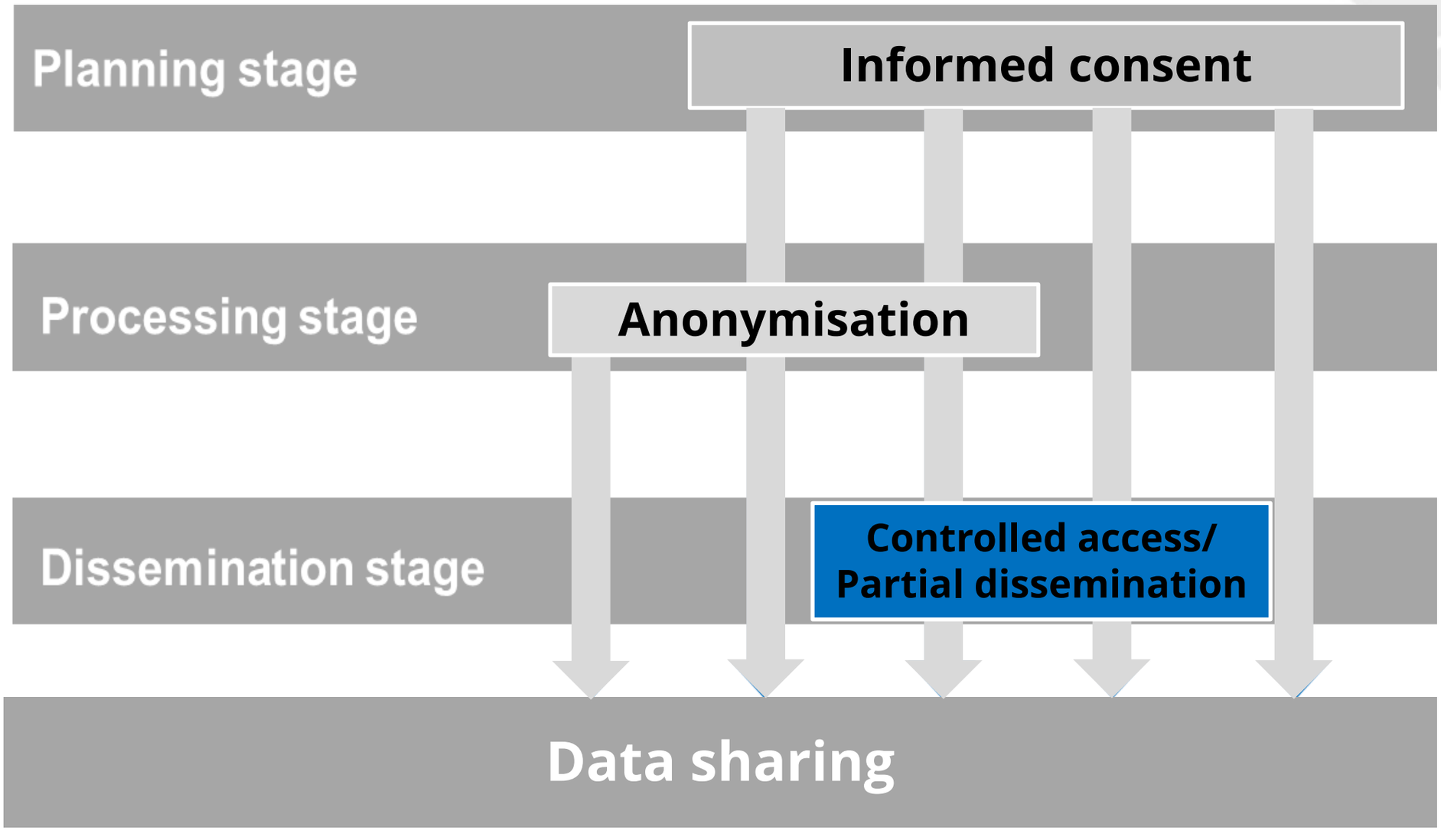
How much anonymization is enough?

Risk/utility balance



Risk assessment

Access control



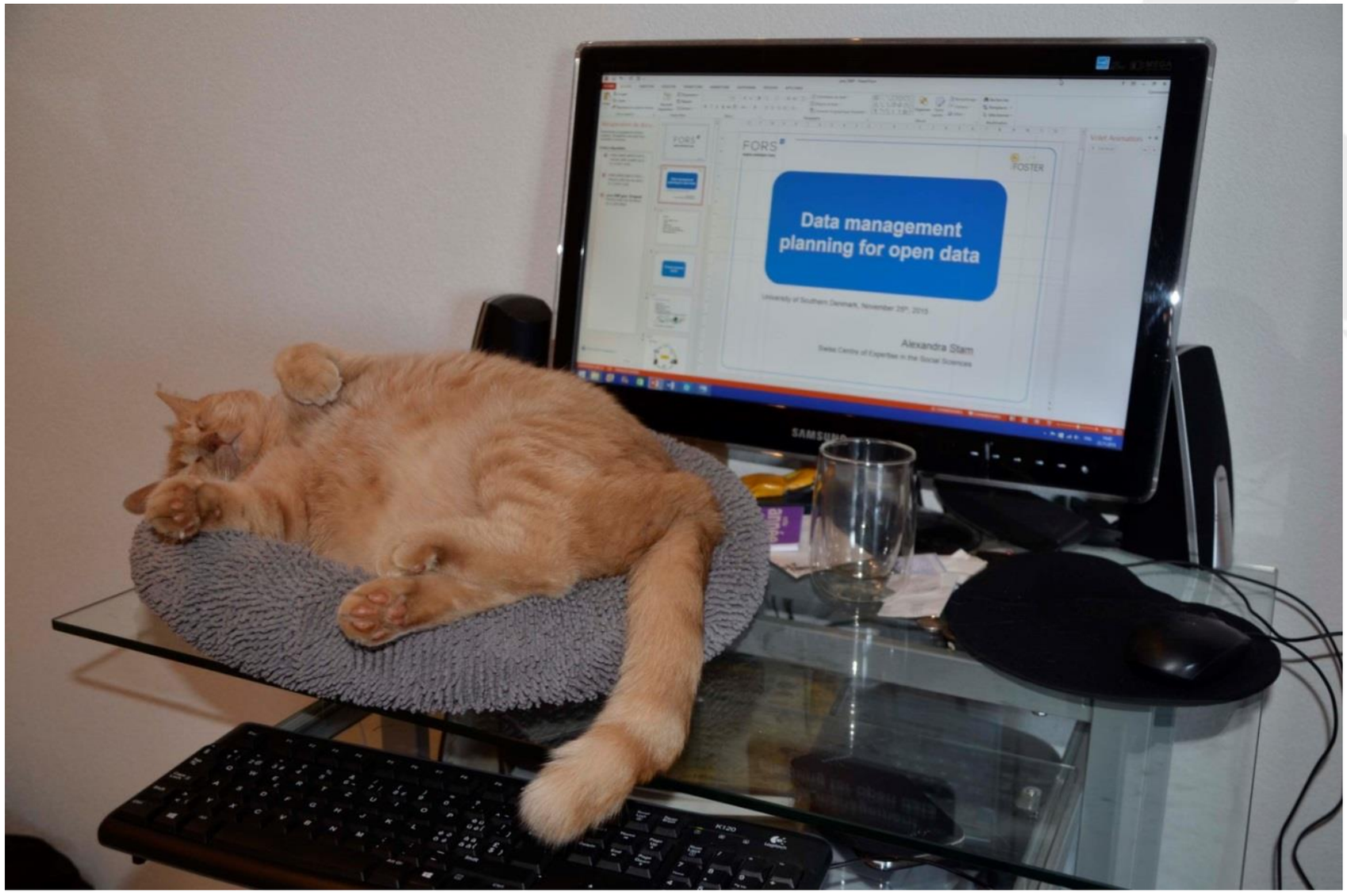
Data access can be managed by

- Delaying access to data
- Limiting access to data
- Controlling access to data

Note that most repositories have end-user contracts, hence the importance of favoring established repositories over more informal ways of sharing your data.

Conclusion

- The Open research framework is an invitation to make data as open as possible depending on their nature;
- Ethics is a matter of practice. Doing research in an ethical way implies a number of trade-offs and compromises;
- Data management planning allows to identify and plan for key actions needed for making data sharable throughout the research lifecycle;
- Consider informed consent, anonymization and access control jointly;
- Do not hesitate to get in touch with repositories at an early stage.



Thanks for your attention

anne-mette.somby@hvl.no
alexandra.stam@fors.unil.ch

From CESSDA Training 2019 team!

amso@hvl.no

cessda.eu @CESSDA_Data

cessda