Collecting data: Procedures and formalities from inception to archival – A practical introduction for social scientists

March 20, 2025

Slideshow: https://github.com/kmannemar/data-collection

Anders Søndergaard Røn, asr@ps.au.dk
Data manager

Kim Mannemar Sønderskov, ks@ps.au.dk
Professor, Research data management advisor

DEPARTMENT OF POLITICAL SCIENCE
AARHUS UNIVERSITY

Collecting data: Procedures and formalities from inception to archival – A practical introduction for social scientists

Motivation

- Increase in volume and modes of data collection
- Increase in rules and norms regarding data collection, esp. regarding individuals

$\rightarrow$ Massive increase in complexity for empirical scientists (studying individuals)

Goal

- Practical overview of the rules and procedures that face social scientists when collecting data – from inception to archival (More than just *GDPR*)

Collecting data: Procedures and formalities from inception to archival – A practical introduction for social scientists

Motivation

- Increase in volume and modes of data collection
- Increase in rules and norms regarding data collection, esp. regarding individuals

$\rightarrow$ Massive increase in complexity for empirical scientists (studying individuals)

Goal

- Practical overview of the rules and procedures that face social scientists when collecting data – from inception to archival (More than just *GDPR*)

Disclaimers!

- The rules they are a'changing
- Complicated are them rules (and I am no expert)
- Bored? Interrupt!
- (Most) Rules make sense

Agenda

Goal: Practical overview of the rules and procedures that face social scientists when collecting data – from inception to archival

Program:
1. Introduction
2. Before collection
3. During collection
4. After collection

1+2 takes the longest...
8.30–12.00

Introduction

1. Introduction
   - Sources of rules and procedures
   - Topics and important concepts
2. Before collection
3. During collection
4. After collection

Sources of rules and procedures

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Funders
- Data providers
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Sources of rules and procedures

Missing sources?

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Funders
- Data providers
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

- Scope of collection
- Attributability to individuals/Personal data
- Research Subjects
- Location of research subjects
- Direct engagement with human subjects?
- Collector and Purpose
- Format: Quant, Qual etc.

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and important concepts

Scope of collection: One study vs. entire research project

- A *study* is one data collection process (e.g., a field experiment, survey, a panel survey, and a series of face-to-face interviews), where the collected data may be used in more than one research *project* (and more than one publication). A project and a publication may use/report more than one study.
- Focus here: Study
- ...which more or less correspond to practice in some disciplines, AU's "Fortegnelse" and AU's IRB.

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

Attributability to individuals ("pensonhenførbarhed")?
→ Personal data/"Personhenførbare data"

- Crucial distinction for GDPR, storage, ethics, sharing

To signal distinction here: `Only Personal Data`

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

Attributability to individuals ("pensonhenførbarhed")?
→ Personal data/"Personhenførbare data"

- Crucial distinction for GDPR, storage, ethics, sharing

To signal distinction here: `Only Personal Data`

1. *Non-anonymous data*: Data with direct identifier (e.g. name, CPR, address) → Yes, Personal data
2. *Pseudononymized data*: No identifier in data, but data can be linked with data with identifier → Yes, Personal data
3. Anonymous data: Non-attributable to individuals, even in combination with other data → No, Non-personal data

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

Attributability to individuals ("pensonhenførbarhed")?
→ Personal data/"Personhenførbare data"

- Crucial distinction for GDPR, storage, ethics, sharing
  To signal distinction here: `Only Personal Data`

1. *Non-anonymous data*: Data with direct identifier (e.g. name, CPR, address) → Yes, Personal data
2. *Pseudononymized data*: No identifier in data, but data can be linked with data with identifier → Yes, Personal data
   - "can" → in theory
   - "linked" → also without explicit key, e.g. categorical answers in a survey
3. Anonymous data: Non-attributable to individuals, even in combination with other data → No, Non-personal data

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

Attributability to individuals ("pensonhenførbarhed")?
→ Personal data/"Personhenførbare data"

- Crucial distinction for GDPR, storage, ethics, sharing
  To signal distinction here: `Only Personal Data`

1. *Non-anonymous data*: Data with direct identifier (e.g. name, CPR, address) → Yes, Personal data
2. *Pseudononymized data*: No identifier in data, but data can be linked with data with identifier → Yes, Personal data
   - "can" → in theory
   - "linked" → also without explicit key, e.g. categorical answers in a survey
3. Anonymous data: Non-attributable to individuals, even in combination with other data → No, Non-personal data

- Note: The status of a study may change over time, typically from "Yes" to "No"

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

Research subjects

- Individuals, Companies, Organizations, beyond
- *Individuals*, *Companies* and sometimes *Organizations* are special categories
- Individuals: Privacy and ethics
- Companies/organizations: Privacy
- Focus here: Collection of data about individuals
- However, knowledge about companies, organizations etc. often require data about *individuals*

Location (of research subjects)

- GDPR: Applies everywhere if collected/controlled by anyone residing in EU/EEA
- Ethical requirements may vary by country, more on ethics later

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

Mode of collection: Direct engagement with human subjects?

- Direct engagement: Making subjects do something they would not otherwise have done in that instance (e.g., answer questions) or exposing them to information and stimuli (e.g., in a field experiment), they would not otherwise have been exposed to in that instance.
- We consider both modes here
- Implications for ethics and GDPR, later

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

Collector and Purpose

- Collector: Scientist, administrator, student
- Purpose: Science, teaching, administration, fun

- Collector and purpose here: By scientist for science
- Not here, but relevant to many: Data collection by students, but see AU-LINK and PS-LINK
  - Main take away: Stricter informal norms for scientists for science, but more liberal rules wrt. personal data (Scientific research purpose/ "Forskningshjemlen")

Topics and Important concepts

Collector and Purpose

- Collector: Scientist, administrator, student
- Purpose: Science, teaching, administration, fun

- Collector and purpose here: By scientist for science
- Not here, but relevant to many: Data collection by students, but see AU-LINK and PS-LINK
  - Main take away: Stricter informal norms for scientists for science, but more liberal rules wrt. personal data (Scientific research purpose/ "Forskningshjemlen")

Format: Quantitative, qualitative, other

- Less important distinction here, although it often has implications for anonymizability and hence attributability to individuals

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Data providers
- Funders
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Topics and Important concepts

- Scope of collection
- Attributability to individuals
- Research Subjects
- Location of research subjects
- Direct engagement with human subjects?
- Collector and Purpose
- Format: Quant, Qual etc.

Sources of rules and procedures

- Legislation
  - GDPR/Databeskyttelsesloven [Data protection act]: Data about and attributable to individuals
  - Arkivloven [Archives Act]: Data collected by Danish authorities
- Aarhus University
  - Information Security Policy
  - AU's Data Protection Officer (DPO)
  - AU's Technology Transfer Office (TTO)
- Collaborators' institutions
- Funders
- Data providers
- Publishers
- Open Science movement/FAIR (Findable, Accessible, Interoperable and Reusable)

Before collection

1. Introduction
2. Before collection
   - First steps: Idea, design, money, collaboration
     - Talk to Anders!
   - Processing/controlling personal data - GDPR issues
   - Storage
   - IRB/Ethical approval
   - Pre-Analysis Plan
3. During collection
4. After collection

First steps: Idea, design, money, collaboration

First steps: Idea, design, money, collaboration

First steps...

- Get the idea
- Design the overall study incl. data collections
  - Make informed choices about the data collection
  - Power calculation?
  - Pilot study? That's a study/data collection in itself...
- Estimate costs
  - Get 2-3 quotes if a contractor is needed
  - if $> 1.6M \rightarrow$ Public procurement, contact udbud@au.dk
  - See more here: AU-LINK
- Get funding if needed
  - May require research ethical considerations /ethical approval of the study
- Make agreement with collaborators regarding co-authorship, also about subsequent projects using the data
  - Funding acquisition/designing a study qualifies for being invited as coauthor, but acquisition is probably not enough in itself. (Co-authorship: Vancouver guidelines)

Processing/controlling personal data? A GDPR thing

Next step: Determine if you and/or others in the project will be **processing** (i.e. collecting/manipulating/storing/analyzing) personal data `Only Personal Data`

Personal data: Non-anonymous and Pseudonymised data (see above)

If "Yes": → GDPR → There must be at least one **data controller**

- Data controller: Determines the purposes for which and the means by which personal data is to be **processed**
- Should you (i.e. AU) be data (co-)controller?
- PI, co-PI, other? (Note: You can e.g. co-design collection, analyze etc. without being controller)
- If you are not PI and not processing personal data → Let someone else be data controller (remember to acknowledge them!)
- More than one data controller? Practical if two or more institutions are handling personal data → Joint data controller agreement and collaboration agreement Contact TTO/Anders.
- Rarely: You could also be a data processor (→ Data processor agreement (data controller's responsibility))
- If AU is data controller → Data controller obligations (below)

Cases – discuss 2 and 2

Who, if any, are the data controller(s) and data processor(s) in one/more of these situations

1. The study you are working on right now
2. YouGov have collected monthly public opinion data. The data has been anonymized. You purchase the data from YouGov
3. You conduct semi-structured interviews with civil servants. They are identifiable from the material. You send the audio-files to MyGoodTape who make transcriptions of the interviews. MyGoodTape deletes the data after processing.
4. You collect data from X about politicians and store their names and other information
5. You pay YouGov to collect data for a survey experiment designed by you; respondents are sampled from their respondent pool. YouGov removes identifiable information before sending you the data.

Processing/controlling personal data II - Outside EU/EEA

`Only Personal Data`

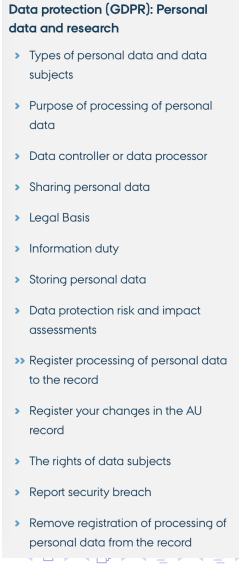
- GDPR applies to processing of personal data inside EU/EEA regardless of the location of the subjects
- What if all data controller(s) and the subjects are located outside EU/EEA?
- Note: If data is collected outside EU/EEA under your instruction: AU is data-controller

Data controller obligations

`Only Personal Data`

Primary obligations: (See also AU-LINK)

- Ensure a legal basis for processing personal data (next slide)
- Make a joint data controller agreement, if more than one controller (following slides)
- Make data processor agreements if processor(s) is used (following slides)
- Think ahead regarding sharing and storing of personal data (following slides)
- Perform risk assessment (following slides)
- Register the study at AU's record of data processing activities (following slides)



Ensure a legal basis for processing personal data

`Only Personal Data`

Formally a choice between: *Scientific research purpose* ("Forskningshjemlen") and *Valid consent to data processing*

See AU-LINK – also for templates regarding information and consent forms

Scientific research purposes

- Purpose must be research
- Duty to inform participants
- ...which can be dropped if data is collected without *direct engagement* (register data, social media data, etc.) and, presumably field experiments etc.
- Note: Data collected on this legal basis can only be used for research
  - or obtain consent from all subjects; or anonymize it

Valid consent to data processing

- Obtain consent
- Specific requirements for the consent

Make a joint data controller agreement

`Only Personal Data`

More than one data controller? Practical if two or more institutions are possessing personal data
$\rightarrow$ Joint data processor agreement and collaboration agreement
Contact Anders/TTO

Make data processor agreements if data processors are used

`Only Personal Data`

- Data processor: processes (collects/stores/manipulates/analyses etc.) personal data under the instruction of the data controller
- E.g. QuestionPro, YouGov*, Statistics Denmark, Nvivo, MyGoodTape, ChatGPT**
- If your personal/pseudononymized data leaves AU: You are using a data processor
- You need a data processor agreement! Templates exists at TTO
- Rules about auditing exists...
- Special rules regarding processors outside EU/EEA $\rightarrow$ Contact TTO
- *YouGov etc. If data is collected under your instruction, but from their pool of respondents: Joint data processor agreement (templates exists)
- **ChatGPT: Off-line LLM exists. Anders? CoPilot?

Think ahead regarding sharing of personal data

`Only Personal Data`

Sharing of *personal data* with collaborators, data processors, or yourself?

- Will you use data processors for personal data?
- Will you share personal data with collaborators outside AU?
- Will you share personal data with collaborators/data processors outside EU/EEA?

Think ahead regarding sharing of personal data

`Only Personal Data`

Sharing of *personal data* with collaborators, data processors, or yourself?

- Will you use data processors for personal data?
- Will you share personal data with collaborators outside AU?
- Will you share personal data with collaborators/data processors outside EU/EEA?
- Should you be able to access personal data if/when you leave AU???
- This should be clear from the information material/consent form
  - With yourself (inside EU/EEA) = collaborators: Mention in consent form

Plan also ahead if you want to share anonymized data as this can be used for other purposes → Should be clear from the info material/consent form (ethics vs. GDPR)

Cases – discuss 2 and 2

In light of the preceding slides...New complications/questions etc.

1. Your own study
2. **Another case**
3. YouGov have collected monthly public opinion data. The data has been anonymized. You purchase the data from YouGov
4. You conduct semi-structured interviews with civil servants. They are identifiable from the material. You send the audio-files to MyGoodTape who make transcriptions of the interviews. MyGoodTape deletes the data after processing.
5. You collect data from X about politicians and store their names and other information
6. You pay YouGov to collect data for a survey experiment designed by you; respondents are sampled from their respondent pool. YouGov removes identifiable information before sending you the data.

Perform risk assessment (and potentially impact assessment)

`Only Personal Data,but`

Template at AU-LINK

Perform risk assessment (and potentially impact assessment)

What will be the consequences of unintentional access to personal data for the people whose data is being processed in the project? (*Please tick relevant box*)

- ☐ No or insignificant consequences (1)
- ☐ Inconvenient consequences (2)
- ☐ Critical consequences (3)
- ☐ Unacceptable consequences (4)

State reason: Click here to write

Likelihood times consequences (e.g. unlikely x inconvenient consequences = *1x2 =2) (tick the table)*

Consequences X Likelihood	No or insignificant consequences	Inconvenient consequences	Critical consequences	Unacceptable consequences
Unlikely	1☐	2☐	3☐	4☐
Not very likely	2☐	4☐	6☐	8☐
Likely	3☐	6☐	9☐	12☐
Expected imminent	4☐	8☐	12☐	16☐

What has been done to limit the likelihood and the consequences?
Describe: Click here to write

Register the data collection to AU's record of data processing activities

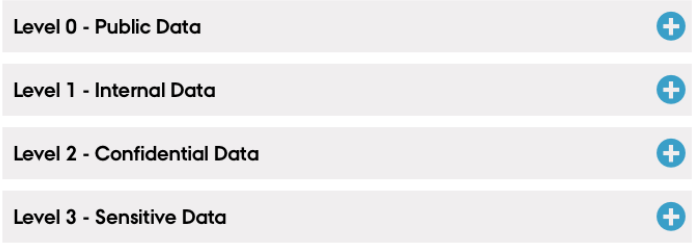
`Only Personal Data`

AU is obligated to maintain a record of all ONGOING personal data processing activities performed by the university in the role of data controller or data processor.
Information: Project information, types of data, data sharing
Remember to update registration if information changes – including if the data is no longer personal or you are no longer employed at AU.
Here: AU-LINK

The following slides are potentially relevant for all types of data

Plan storage ← Data type

4 types of data:



AU-LINK

- Public: Shared research data
- Internal: Unshared, non-personal research data, General Personal data (name, age, education, salary, nationality)
- Confidential: Personal research data (cpr, marital status, grades)
- Sensitive: Sensitive personal data (Race, political attitudes, genetic, health criminal record etc.)

Note: Arkivloven/Archives Act

- Archives Act concerns **all** data collected by Danish authorities including universities
- Researchers are obliged to ask The National Archive (*Rigsarkivet*) if they find the data worth preserving
- You cannot ask beforehand
- If yes: You are obliged to document it and send it to them with, preferably with as much info as possible $\rightarrow$ Not anonymized?
- Recent news: If data about "ordinary" citizens $\rightarrow$ ok to anonymize
- Personal data will be available after 50/70 years
- Standard phrasing in info material: "Efter denne dato vil dine personoplysninger blive anonymiseret, overført til Rigsarkivet eller slettet."

Plan storage ← Data type

AU-LINK

	PUBLIC	INTERNAL	CONFIDENTIAL	SENSITIVE
Brightspace + Panopto	Yes	Yes	No	No
Workzone	Yes	Yes	Yes	Yes
mitHR (HR)	Yes	Yes	Yes	Yes
U-drive (personal drive)	Yes	Yes	No*	No*
O-drive (shared drive) with limited log safety (standard)	Yes	Yes	Yes	No
O-drive (shared drive) with extended log safety (contact IT-support)	Yes	Yes	Yes	No*
STADS	Yes	Yes	Yes	Yes
TYPO3	Yes	No	No	No
OneDrive	Yes	Yes	No*	No*
Sharepoint	Yes	Yes	No*	No*
Outlook	Yes	Yes	No	No
Teams + Zoom	Yes	Yes	No*	No*
Other online Microsoft services	Yes	Yes	No*	No*
Survey-Xact	Yes	Yes	Yes	Yes
REDCap	Yes	Yes	Yes	Yes
Free Cloud services, e.g. Dropbox or Google drive**	Yes	No	No	No

Yes: You are ALLOWED to save/share data here
No: You are NOT ALLOWED to save/share data here
No*: Saving requires that the personal data has been PSEUDONYMISED

Plan storage ← Data type

	PUBLIC	INTERNAL	CONFIDENTIAL	SENSITIVE
Brightspace + Panopto	Yes	Yes	No	No
Workzone	Yes	Yes	Yes	Yes
mitHR (HR)	Yes	Yes	Yes	Yes
U-drive (personal drive)	Yes	Yes	No*	No*
O-drive (shared drive) with limited log safety (standard)	Yes	Yes	Yes	No
O-drive (shared drive) with extended log safety (contact IT-support)	Yes	Yes	Yes	No*
STADS	Yes	Yes	Yes	Yes
TYPO3	Yes	No	No	No
OneDrive	Yes	Yes	No*	No*
Sharepoint	Yes	Yes	No*	No*
Outlook	Yes	Yes	No	No
Teams + Zoom	Yes	Yes	No*	No*
Other online Microsoft services	Yes	Yes	No*	No*
Survey-Xact	Yes	Yes	Yes	Yes
REDCap	Yes	Yes	Yes	Yes
Free Cloud services, e.g. Dropbox or Google drive**	Yes	No	No	No

Yes: You are ALLOWED to save/share data here
No: You are NOT ALLOWED to save/share data here
No*: Saving requires that the personal data has been PSEUDONYMISED

AU-LINK

SIF (Sensitive Information Facility)

SIF (Sensitive Information Facility)

- **Use:** AU-wide system for storing research data that contain personal data or for other reasons require extra protection, e.g. due to IP or strategic concerns. Access to SIF must only happen after completion of the locally managed user course. Contact your local coordinator via the SIF guidance portal.

UCloud: LINK
Contact Anders

IRB/Ethical approval

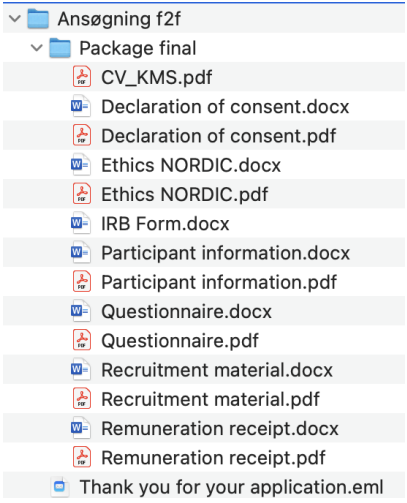
IRB/Ethical approval

- Drastic increase in demand for ethical approval of data collections from journals
- Rules are in flux
- Motivation: Protect research subjects from harm inflicted by the data collection including publication of results

When to apply

- Always:
  - When required by law (e.g. health related experiments involving medical treatment of human subjects)
  - When required by funder
- Probably always: When data collection involves direct interaction with human subjects, or when collecting, systematizing and publicizing data data about individuals even if the un-systematized data is public available
- Maybe not: All other situations
- See When to apply for a research ethical approval of research studies
- Apply: AU's Research Ethics Committee

IRB/Ethical approval – content of application



Self-declaration for AU's Research Ethics Committee

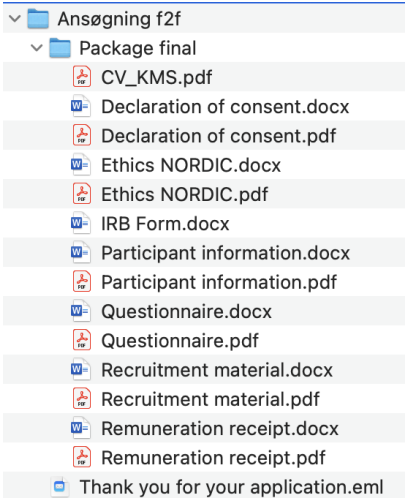
Study manager

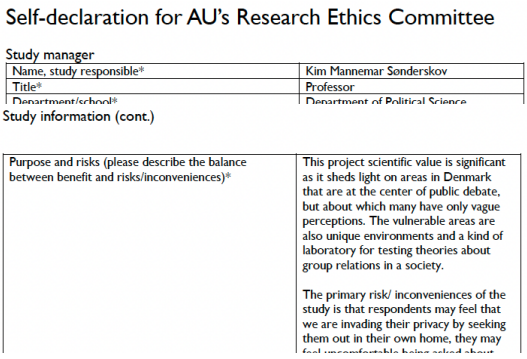
Name, study responsible*	Kim Mannemar Sønderskov
Title*	Professor
Department/school*	Department of Political Science

Study information (cont.)

Purpose and risks (please describe the balance between benefit and risks/inconveniences)*	This project scientific value is significant as it sheds light on areas in Denmark that are at the center of public debate, but about which many have only vague perceptions. The vulnerable areas are also unique environments and a kind of laboratory for testing theories about group relations in a society. The primary risk/ inconveniences of the study is that respondents may feel that we are invading their privacy by seeking them out in their own home, they may feel uncomfortable being asked about

IRB/Ethical approval – content of application





Self-declaration for AU's Research Ethics Committee

Study manager

Name, study responsible*	Kim Mannemar Sønderskov
Title*	Professor
Department/school*	Department of Political Science

Study information (cont.)

Purpose and risks (please describe the balance between benefit and risks/inconveniences)*	This project scientific value is significant as it sheds light on areas in Denmark that are at the center of public debate, but about which many have only vague perceptions. The vulnerable areas are also unique environments and a kind of laboratory for testing theories about group relations in a society. The primary risk/ inconveniences of the study is that respondents may feel that we are invading their privacy by seeking them out in their own home, they may feel uncomfortable being asked about

AU's Research Ethics Committee vs. other universities' IRBs

GDPR and Ethical Approval

- Somewhat unrelated...
- Not all *personal data studies* requires *ethical approval*, *ethical approval* can be relevant irt. *non-personal data*; GDPR-compliance does not imply sound ethics, Ethical approval does not imply GDPR compliance.
- Potential tension: Using research as legal basis without consent and esp. without information might be ethical dubious, but on the other hand: you need to argue for your case in those cases regardless.
- Ethical consent: Voluntary participation; GDPR consent: Acceptance of data processing
- Any thoughts/experiences?

Potential last step:
Pre-registration/Pre-Analysis Plan or Pre-registred Reports

Potential last step:
Pre-registration/Pre-Analysis Plan or Pre-registred Reports

- Pre-Analysis Plan: Registration of hypotheses, specifications, measures, sample modifications etc. prior to collecting (or seeing) the data
- Pre-Registred Report: Manuscript accepted for publication before collecting the data
- A response to the replication crisis
- Increasing popularity, not least in experimental studies...
- ...but also controversy over norms (transparency in deviations, failure to report all pre-pregistred hypotheses, details of the PAP etc); see more here (and sign a call for stricter norms) Open Letter on the Need for Preregistration Transparency in Peer Review
- Popular platforms and/or templates: OSF and aspredicted.org
- Thoughts or experiences (incl. observational studies or non-quantitative studies)?

During collection

1. Introduction
2. Before collection
3. During collection (short)
4. After collection

During data collection

- Store data adequately
- Update record in AU's record `Personal data only`
- Respond to messages from subjects (and treat that communication adequately)
- Other?

After collection

Storage

- Store data adequately
- Remove data from data processors, e.g. QuestionPro
- Or have data processors delete the data if possible (e.g. YouGov)

Anonymize or pseudonymize?

`Personal data only`

- Possible to anonymize data immediately?
- No, because
  - Not ever possible
  - Need personal data for analysis
  - Need identifier for follow-up collection
  - Classes with Archives act? See below
- Yes?
  - Do it! → Decomplicates storage, sharing, minimizes risks, etc.
  - Remember to remove entry in AU's record

Possible to pseudonymize data immediately?

- No, because
  - Not ever possible
- Otherwise do it → makes storing and sharing somewhat easier, and minimizes risks

Pseudonymization

`Personal data only`

- Remove direct identifiers (CPR, Name, address, e-mail etc.)
- Add key
- Store separately

Documentation and data management - hobbyhorses of mine (and others)

- Make replication data for projects available if/when possible
- Document you data asap, the entire
- Make documentation public (along with the entire data if/when possible)
- Take good care of your data: Store a master file that all analyses starts from

Post analyses

- Anonymization, deletion, archiving? $\rightarrow$ FAIR-principles: All materials used to produce research publications should be stored 5 years
- If Personal data was used in publication than is should be kept in that form for 5 years.
- Making replication data public, but not pseudonymized personal data
- Re-use? Depends on info material...

That's it

Comments? Questions?