Best practices in digital data donation: A workflow for studies using data donation

WEB DATA OPP WORKSHOP
MARCH 18-19, 2024, BARCELONA

T.C. Carrière[1], Dr. L. Boeschoten[1], Dr. B. Struminskaya[1], Dr. H.L. Janssen[2], Dr. N.C. de Schipper[2], & Prof. Dr. T.B. Araujo[2]
[1]Utrecht University, [2]University of Amsterdam

This presentation is part of the Digital Data Donation Infrastructure (D3I) project (https://datadonation.eu), made possible by the Platform Digitale Infrastructuur SSH (https://www.pdi-ssh.nl)

Data donation

- Collection method for digital traces
  → Collecting own data + Active consent

- Relies on legislation for personal data
  → e.g., EU's GDPR

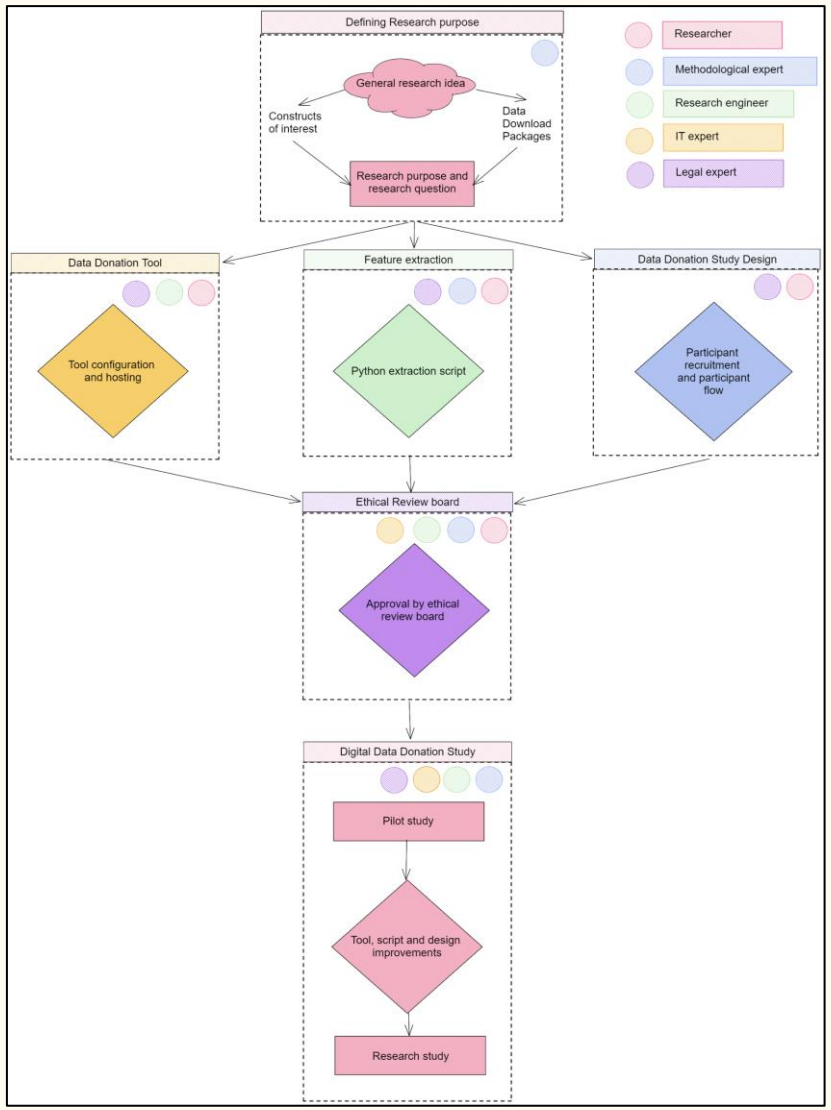
- Different approaches to privacy preservation
  → Kmetty (2022); Araujo (2022); Boeschoten (2022)

Port as Data Donation approach

Boeschoten and colleagues (2022): An approach for data donation, using their software Port.

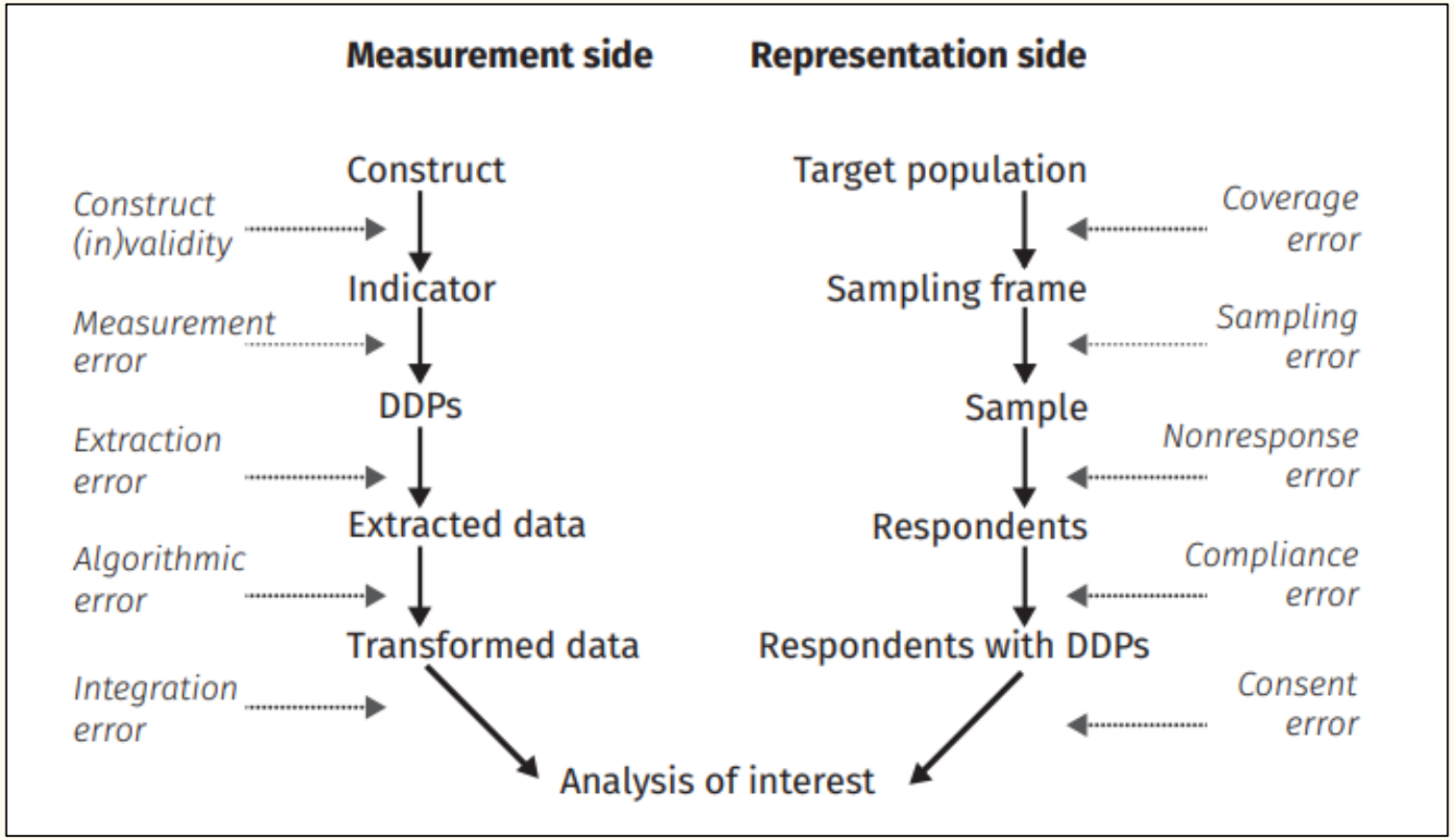
A workflow for data donation

- Based on Port, we created a workflow to guide in setting up a data donation study.
- Challenges & pitfalls identfied for each step & how to deal with these.
- Linked to TE framework for data donation (Boeschoten et al. 2022).



Workflow based on TE framework

Workflow accounts for sources of error in TE framework for data donation (Boetschoten et al., 2022)

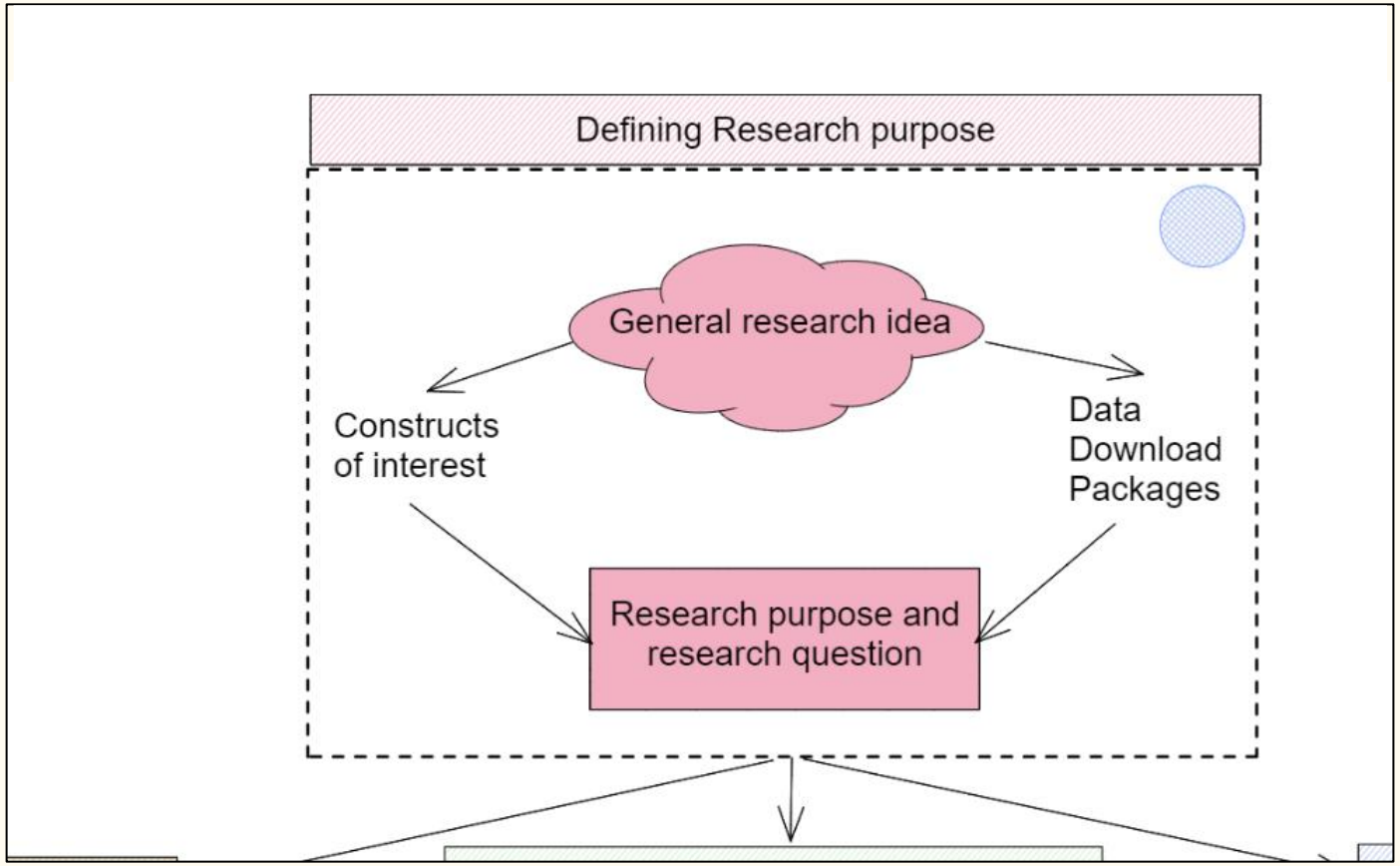


Example study

Corten et al. (2023) – Data donation with WhatsApp data

Collected data:
- Number of groups
- Number of contacts
- Summary statistics of single group chat

1. Defining research purpose



1. Defining research purpose

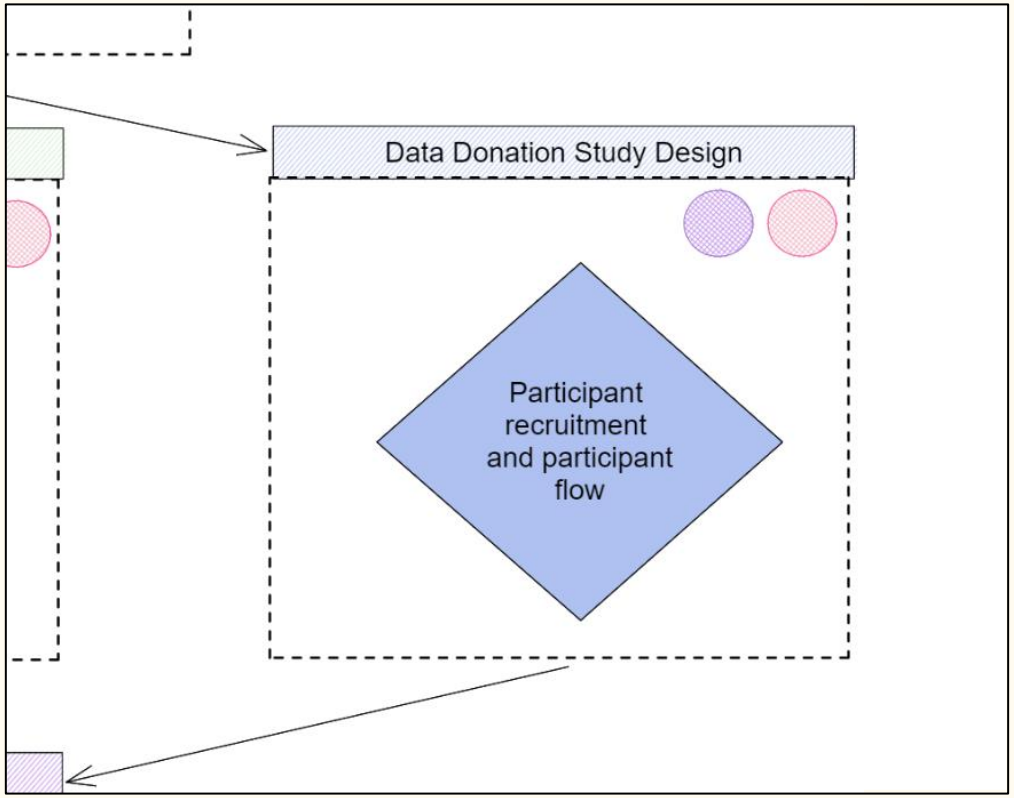
- Clear choices in this step facilitates further study process.
- Decisions on platforms of interest:
  - What is used in target population?
  - What data is available?
  - What drawbacks do platforms have?
- Match between digital traces and study construct.

Step accounts for *Construct invalidity* & *Measurement error*

WhatsApp study:

- Interested in construct 'replying behavior'
- Indicator: number of messages sent directly after someone else
- Drawbacks of indicator: time aspects & reply to other message
  → (no quoting-feature data)

2. Study design



2. Study design

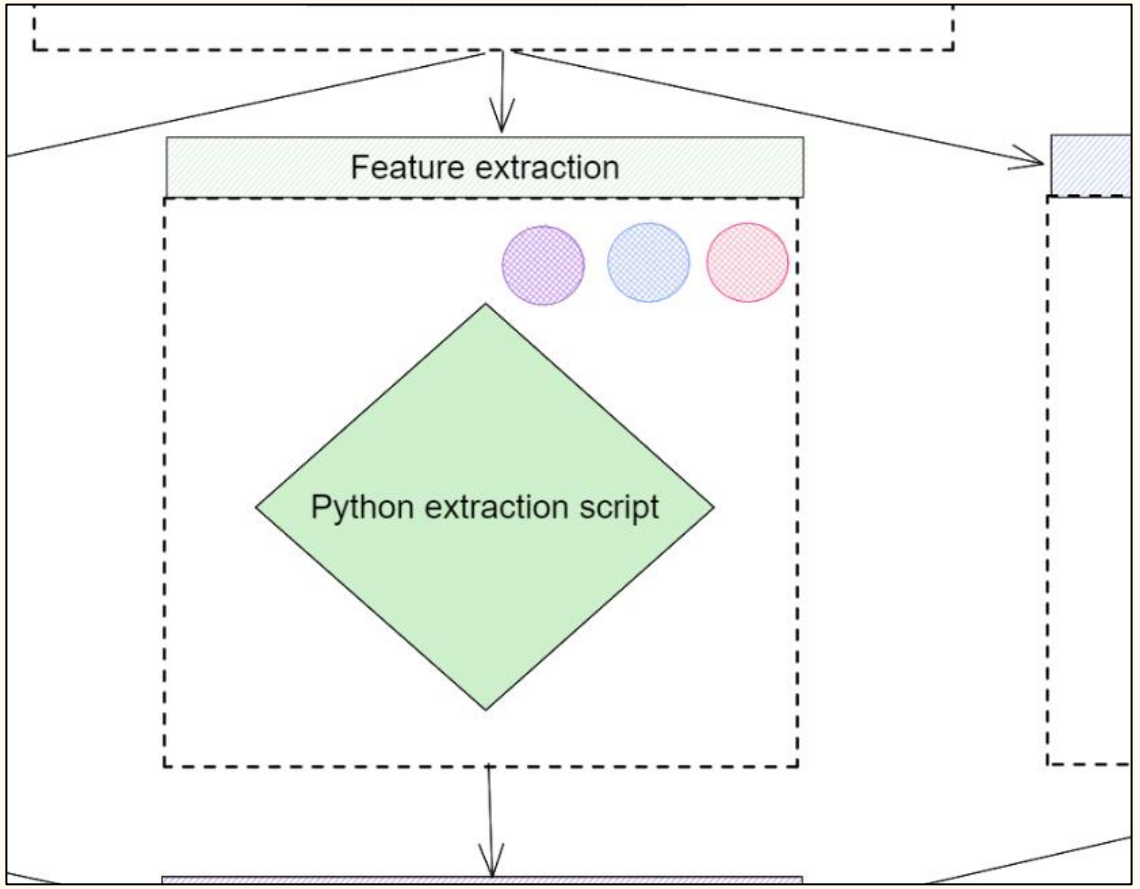
- Include alternative data collection methods (e.g. Surveys).
- Ensure privacy preservation to participants.
- Make DDP download instructions tailored and clear.
- Account for DDP waiting time and expiration time.

To retain response rates:

- Make the study flow easy to follow for participants.
- Include incentives

Step accounts for
Representativity errors
(non-response error; compliance error; consent error)

3. Feature extraction

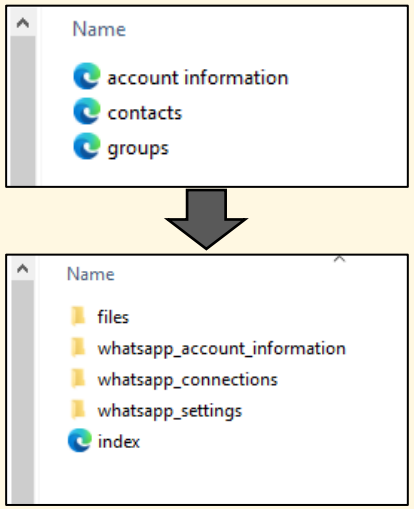


3. Feature extraction

Extraction script should be flexible and tested extensively.

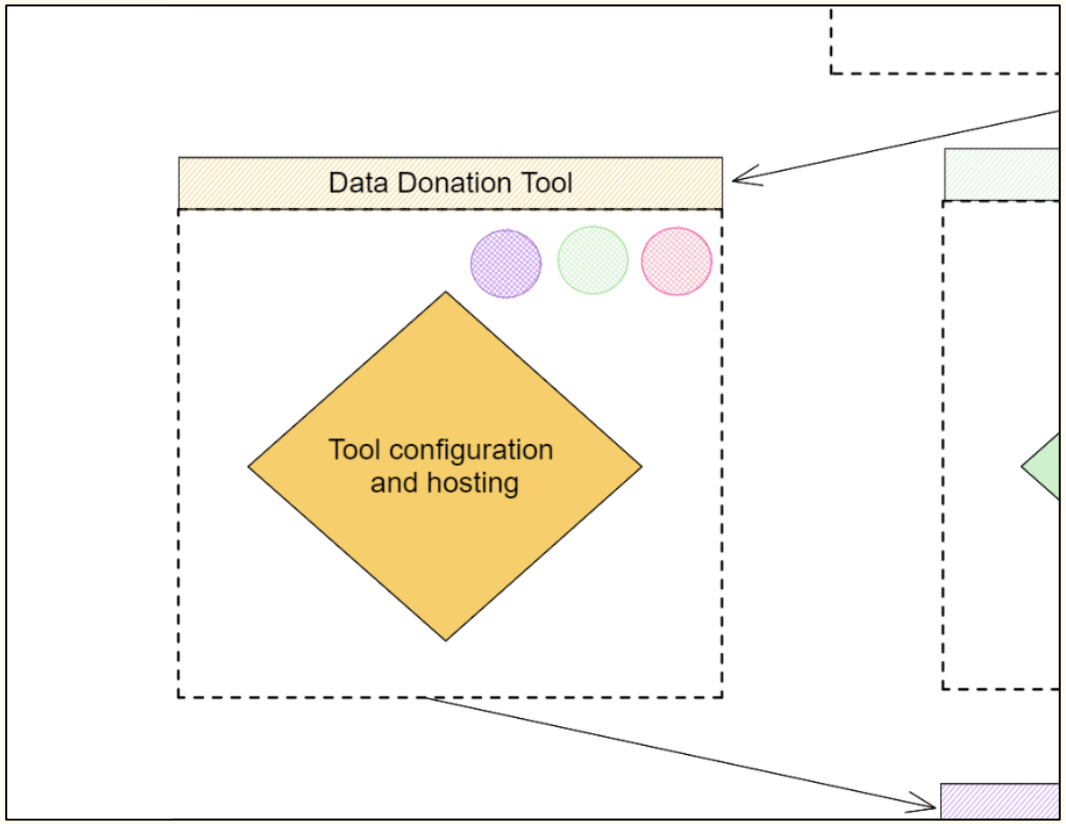
Step accounts for *extraction error* and *algoritmic error.*

- Take into account volatility in DDPs: e.g. Type of files in DDP / Notation.
- Causes
  - Language settings of system
  - Operating system
  - Change over time




```
Date-time formats:

30.07.22, 14:58 - Name:

17/07/2017, 12:02 - Name:

05-12-2021 21:44 - Name:

[7/20/18, 07:34:04] Name:

[7/3/22, 8:25:07 PM] Name:
```

4. Data donation tool



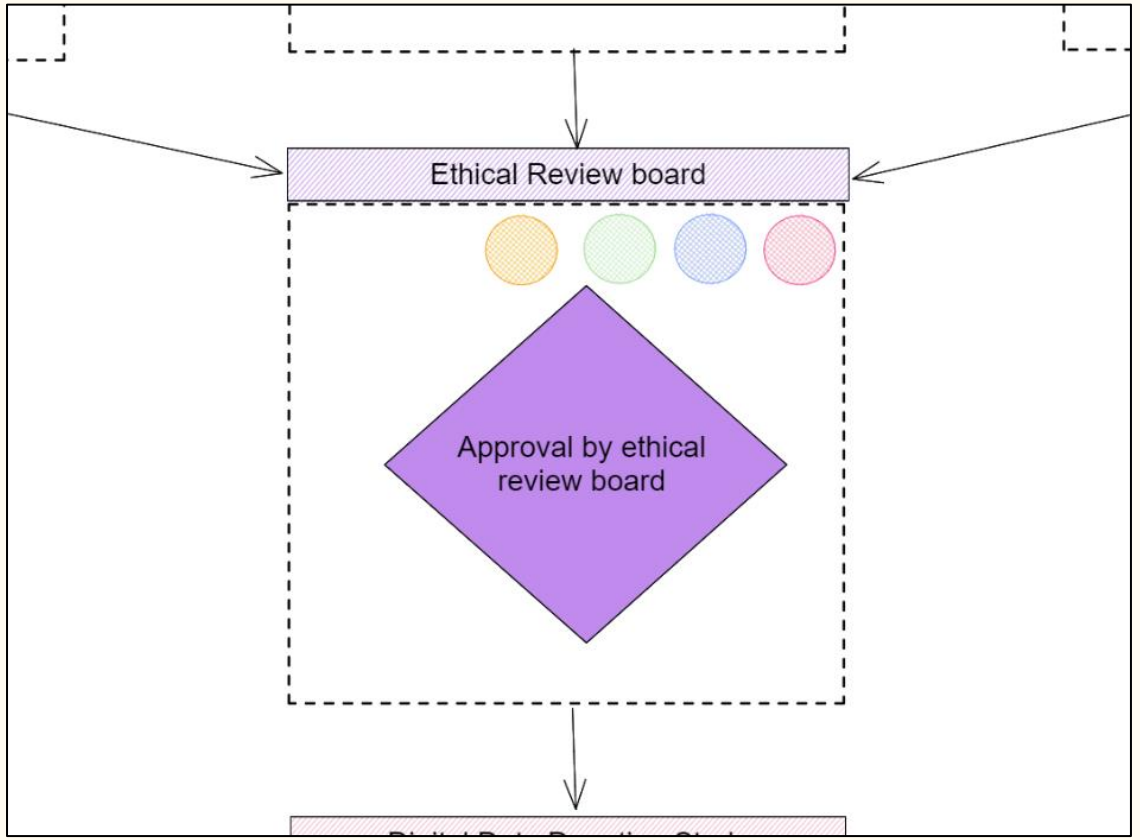
4. Data donation tool

- Choice between outsourcing or using own software.

  → Privacy should be considered

- Initiatives as D3I (*Araujo, Boeschoten et al., 2022*) or Data Donation Module (*N. Pfiffner, P. Witlox, & T. Friemel, 2023*).

5. Ethical aspects

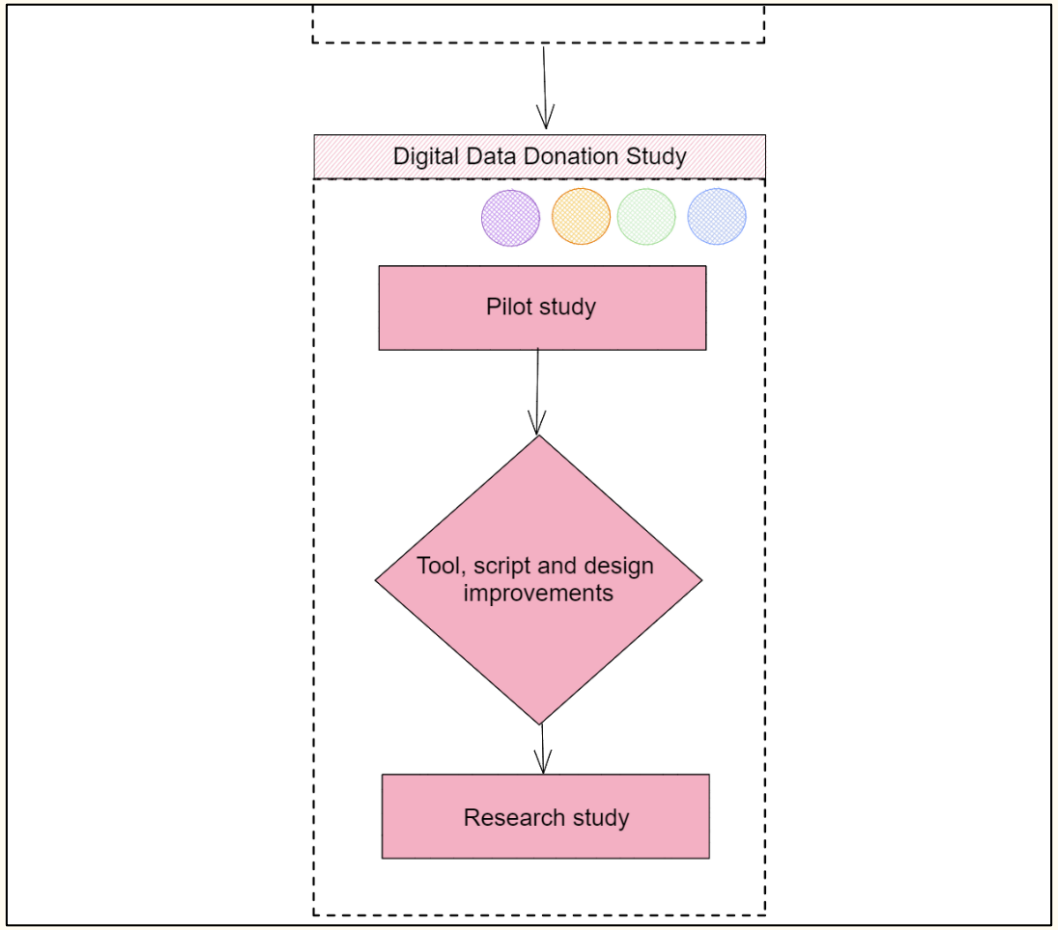


5. Ethical aspects

- GDPR legislation relevant (PI now data controller as well).
- Highly advised to have help from legal expert from institution conducting the study.
- Privacy should be communicated very clearly and should be apparent all study long.

Step might help with *consent error.*

6. Pilot iterations



6. Pilot iterations

Pilot important:

- First time elements interact.
- Spot challenges in participant steps.

First opportunity to link data sources.

→ Survey data and donated data might come in separately

Step accounts for *integration error* and *compliance error.*

Concluding remarks

- Data donation viable collection method for digital trace data.
- The proposed workflow guides researches new to data donation in setting up their studies.
- Improvements needed to make the study flow easier (and less burdensome) for participants.

  → Infrastructures as D3I help

- More substantive studies & studies in new domains helpful for more insight in methodological challenges.

Utrecht University

Sharing science,
shaping tomorrow

Thank you for your attention!

---

t.c.carriere@uu.nl

This presentation is part of the Digital Data Donation Infrastructure (D3I) project (https://datadonation.eu), made possible by the Platform Digitale Infrastructuur SSH (https://www.pdi-ssh.nl)