Digital trace data collection through data donation

Laura Boeschoten

Dept. of Methodology and Statistics

Utrecht University

Team D3I

Researchers

Dr. T.B. (Theo) Araujo
Principal Investigator

Dr. L. (Laura) Boeschoten
Project Lead

Dr. N.C. (Niek) de Schipper
Research Engineer

Dr. mr. H.L. (Heleen) Janssen
Legal expert

Dr. B. (Bella) Struminskaya
Methodological expert

Dr. K. (Kasper) Welbers
Visualization expert

T.C. (Thijs) Carriere MSc.
PhD student

J. (Jonathan) Koop BSc.
Research assistant

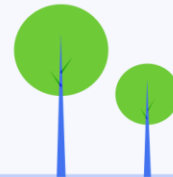**D. (Danielle) McCool MSc.**
Researcher
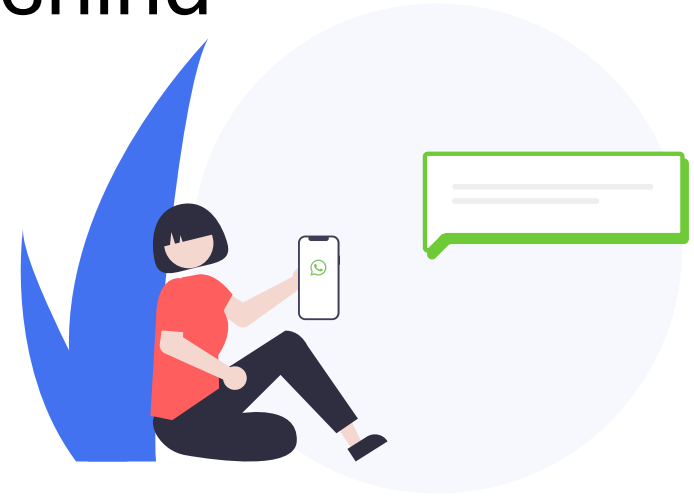
Funded by

Collaborations

Outline

- Introduction to data donation

- Software and infrastructure

- Ongoing work
  - Applications
  - Methodological
  - Software and infrastructure

- What if you want to do a study yourself?

What is data donation?

With almost everything we do, we leave digital traces behind

GDPR

Article 15 – Right of data access

- *Access to the personal data*

Article 20 – Right of data portability

- *receive the personal data in a structured, commonly used and machine-readable format ("Data Download Package")*
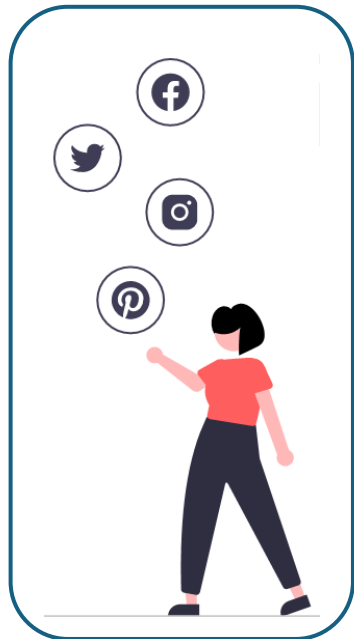
Data Donation

Donation of digital traces collected as Data

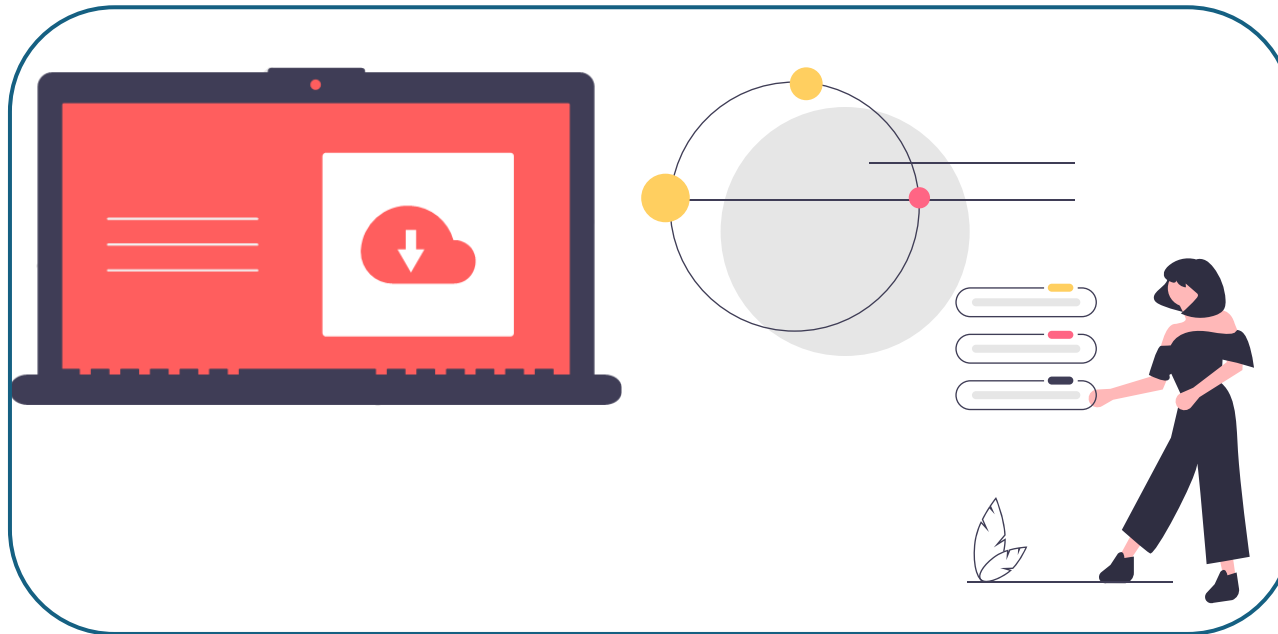Download Packages (DDPs) for research purposes.

The data donation workflow

Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, *4*(2), 388–423. https://doi.org/10.5117/ccr2022.2.002.boes
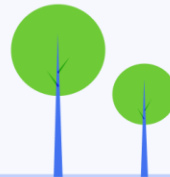
Online platform

Device of participant
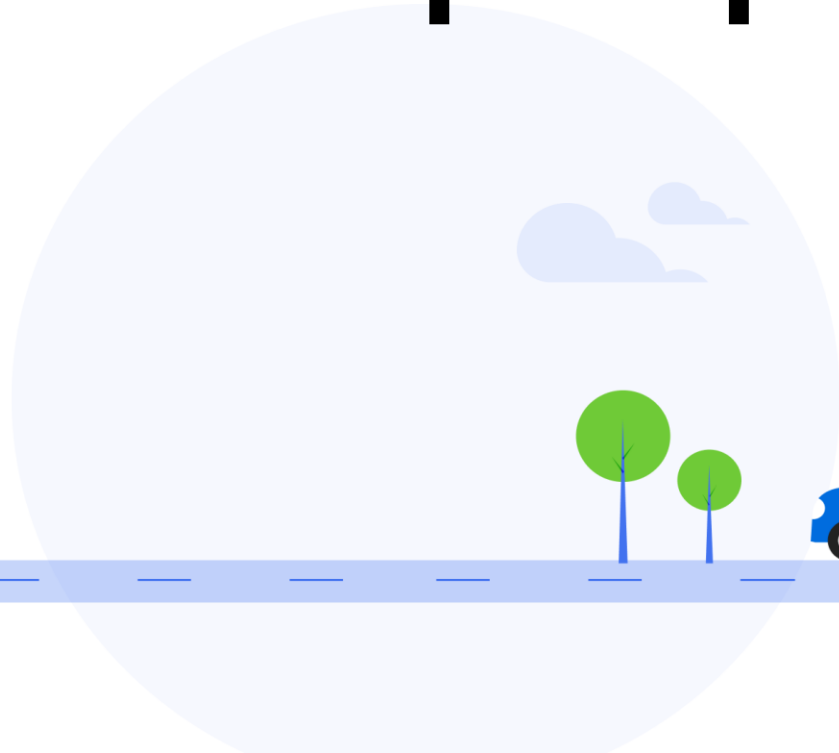
Researcher environment

Prepare your own data donation study using
Port

Boeschoten, L., de Schipper, N. C., Mendrik, A. M., van der Veen, E., Struminskaya, B., Janssen, H., & Araujo, T. (2023).
Port: A software tool for digital data donation. *Journal of Open Source Software*, *8*(90). https://doi.org/10.21105/joss.05596

Researcher perspective

Projects

Overview 2

New project >

Data donation demo

Data donation

1 item

...

Social Media Pilot Cambridge

Data donation

4 items

...

Project

Items 1

Add item >

Concept

Data donation demo TikTok

Assignment

participants | 0 donations

...

Data donation

1 Settings   2 Workflow   3 Monitor

Publish   Preview

- General:
    - Expected number of participants
    - Language (English, Dutch, German)
- Branding:
    - Title and subtitle
    - Organization logo
    - Header image
- About page
- Privacy statement
- Consent form
- Helpdesk contact details

- Panel integration
- Data storage

...nisation that is responsible for
...ut.

Subtitle

University of Cambridge

Organization logo

Next

- My console
- **Projects**
- To-do **0**

- Helpdesk
- My profile

Data donation

① Settings ② **Workflow** ③ Monitor

Publish **Preview**

Workflow

Add tasks from the library to build a custom workflow for participants.

Use the arrows to order the tasks

Questionnaire

⌄ Expand

Request manual ✓

⌄ Expand

Download manual

⌄ Expand

Donate ✓

Library

Choose which tasks to add to the workflow.

Donate

Enables participants to donate data.

Add

Questionnaire

Redirects participants to an online questionnaire.

Add

Request manual

Instructions for participants on how to request digital trace data.

Next

- My console
- Projects
- To-do **0**

1 Settings   2 Workflow   **3 Monitor**

Publish   **Preview**

Donate TikTok TikTok

● **39 finished**   ● **16 started**   ● **345 expected**

Donate Instagram Instagram

● **47 finished**   ● **19 started**   ● **334 expected**

Final Questionnaire

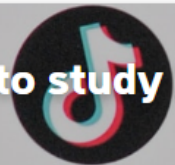● **400 expected**

- Helpdesk
- My profile

Participant perspective
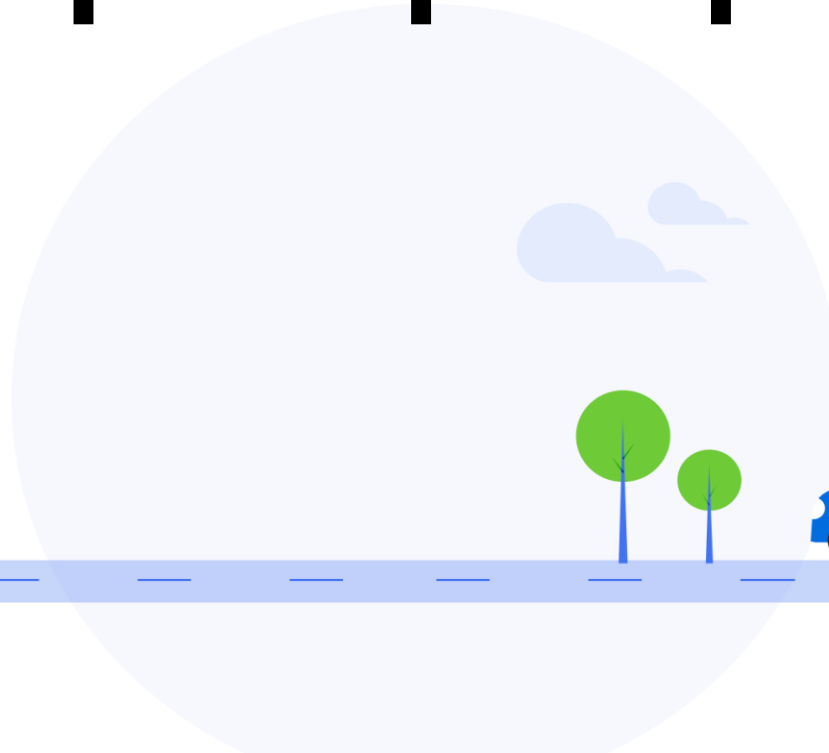
Harnessing digital data to study 21st century adolescence
University of Cambridge

Consent

Dear Participant,

You will find three assignments once you press 'Continue'.
1. If you are comfortable with sharing both your TikTok and Instagram data with us, please select the first assignment that says 'Data Donation – TikTok and Instagram'.
2. If you are comfortable with sharing only your TikTok data with us, please select the second assignment that says 'Data Donation – TikTok'.
3. If you are comfortable with sharing only your Instagram data with us, please select the second assignment that says 'Data Donation – Instagram'.

Privacy and confidentiality:
- Your data will remain on your local PC until you choose to press 'Donate' on the Next platform.
- All uploaded data will be securely stored and anonymised for analysis purposes.
- Only authorised members of our research team will have access to the uploaded data, and stringent confidentiality measures will be upheld at all times.

Contact Information:
If you have any questions, concerns, or encounter difficulties during the data download or upload process, please don't hesitate to contact our research team at dmh-group@mrc-cbu.cam.ac.uk.

Thank you once again for your participation!

Do you agree to the above terms and conditions?

Yes, I agree   No

Harnessing digital data to study 21st century adolescence
University of Cambridge

Tasks

1. Questionnaire
2. Request ♪
3. Download ♪
4. Donate ♪

Questionnaire

Start

Harnessing digital data to study 21st century adolescence
University of Cambridge

Tasks

1 Questionnaire

2 Request ♪

3 Download ♪

4 Donate ♪

Donate

Click 'Start' to go through a series of steps to extract data from your TikTok data package. During this process, your package does not leave your device. Before deciding on whether to donate the data, they will be displayed on the screen.

Start

Harnessing digital data to study 21st century adolescence
University of Cambridge

Tasks

1. Questionnaire
2. Request ♪
3. Download ♪
4. Donate ♪

TikTok

Click 'Choose file' to choose the file that you received from TikTok. If you click 'Continue', the data that is required for research is extracted from your file. This may take a while, thank you for your patience.

Choose a file   **Choose file**

Note: The process to extract the correct data from the file is done on your own computer. No data is stored or sent yet.

Continue

Harnessing digital data to study 21st century adolescence

University of Cambridge

Tasks

1 Questionnaire

2 Request

3 Download

4 Donate

Open

<< 2023_Cambridge > demo

Search demo

Organize ▾   New folder

TikTo

Click 'Choose

from your file

Choose a f

Note: The process t

Continu

Quick access

ownCloud

Dropbox

OneDrive - Universiteit Utrecht

SURFdrive

This PC

Network

Linux

Name

instagram-lauraboeschoten-2024-03-04-a6lZHwlu

release_hotfix_insta

release_insta

release_tiktok

TikTok_Data_1709294680

extracted

ile

File name: TikTok_Data_1709294680

Custom Files

Open   Cancel

Harnessing digital data to study 21st century adolescence
University of Cambridge

Tasks

1. Questionnaire
2. Request 🎵
3. Download 🎵
4. **Donate** 🎵

TikTok

Click 'Choose file' to choose the file that you received from TikTok. If you click 'Continue', the data that is required for research is extracted from your file. This may take a while, thank you for your patience.

TikTok_Data_1709294680.zip   **Choose file**

Note: The process to extract the correct data from the file is done on your own computer. No data is stored or sent yet.

Continue

Harnessing digital data to study 21st century adolescence
University of Cambridge

Tasks

1. Questionnaire
2. Request 🎵
3. Download 🎵
4. Donate 🎵

TikTok

Decide whether you would like to donate the data below. Carefully check the data and adjust as required. Your donation will contibute to the reasearch project that was explained at the start of the project. Thank you in advance.

If you DO NOT want to donate any of the information in the table below, you can select the row and delete it from your data donation in the table below

Summary information

Search

🔲 2 columns, 9 rows   🔍 Hide table

	Description	Number
🔲	Followers	0
🔲	Following	91
🔲	Likes received	0
🔲	Videos posted	0
🔲	Likes given	13
🔲	Comments posted	0
🔲	Messages sent	0

🗑 Delete   ‹‹ ‹ 1 › ››

Comments and likes

Search

Harnessing digital data to study 21st century adolescence
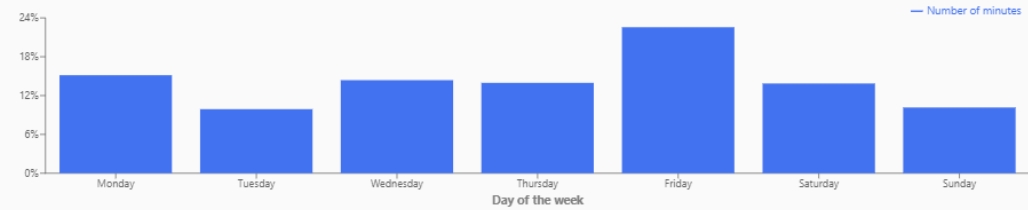University of Cambridge
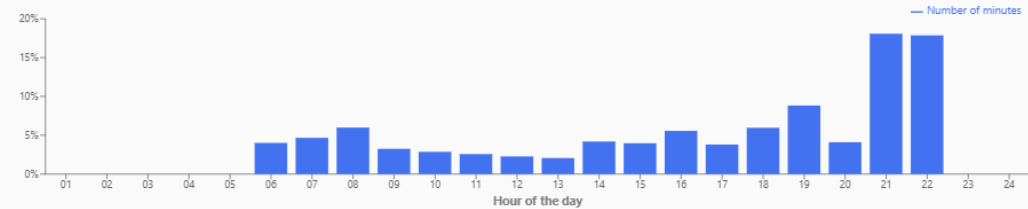
Tasks

1. Questionnaire
2. Request 🎵
3. Download 🎵
4. Donate 🎵

Month

Average time spent on TikTok per day of the week

— Number of minutes

Day of the week

Average time spent on TikTok per hour of the day

— Number of minutes

Hour of the day

Direct Message Activity

This table contains the times at which you sent or received direct messages. The content of the messages is not included, and user names are replaced with anonymous IDs.
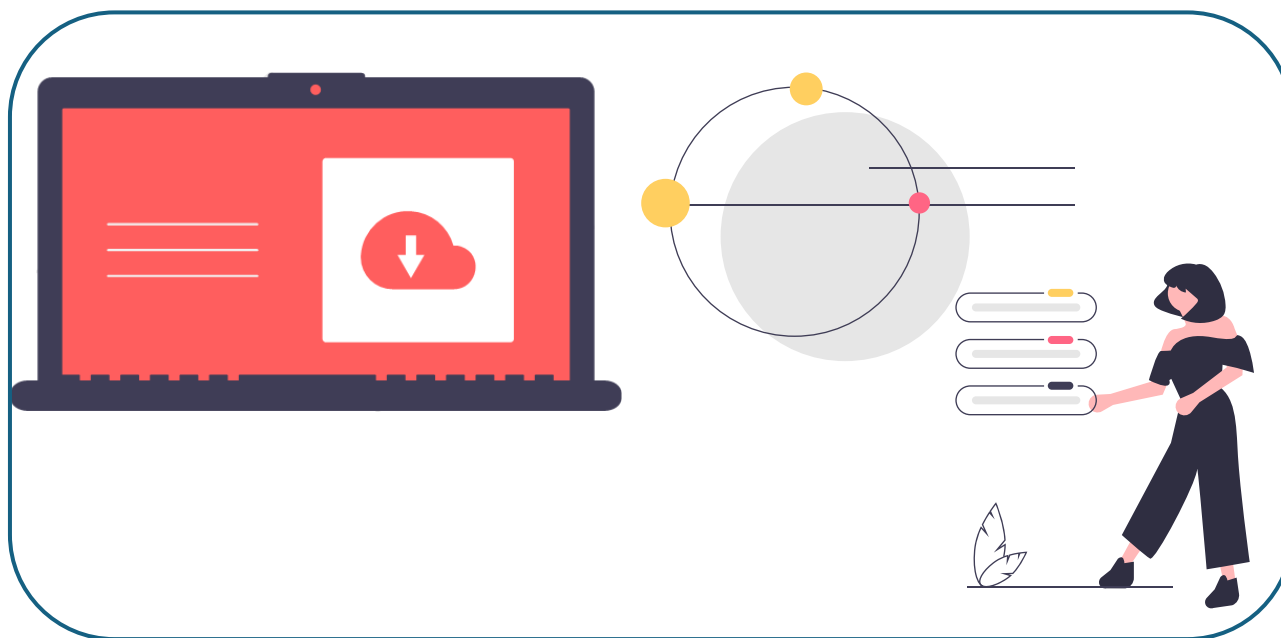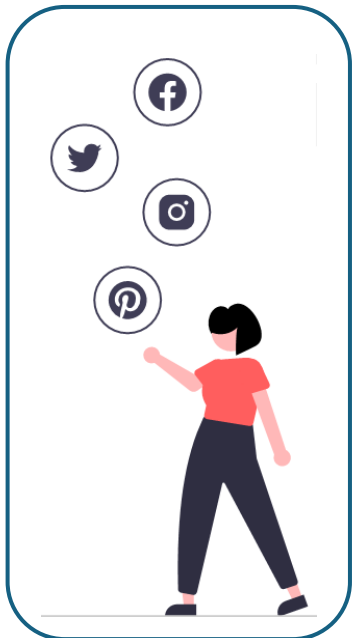
▦ no data

Do you want to donate the above data?

[Yes, donate]   No

How to do it yourself?

Consider separately:

1. The Next platform for configuring your data donation **_Participant Flow_** (SaaS by Eyra).

2. The **_Donate Task_** within the participant flow (Python script).

1. The Next platform

- Available open source through: https://github.com/eyra/mono

- Can be used directly:
  - SaaS solution by Eyra
  - SURF Research Cloud
  - Or do it yourself

- Arranging participant recruitment and data storage can be related to how you use Next.

2. The donate task

- As each study considers a different platform, and is interested in different parts of the DDP, a custom Python script needs to be written, see: https://github.com/d3i-infra/data-donation-task

- Python script consists of the following elements:
  1. Ask the participant to submit a zip file
  2. Perform validation on the submitted zip file, if not valid return to step 1
  3. Extract the data from the submitted zip file
  4. Render the extract data on screen in a table
  5. Send the data to the data storage upon consent
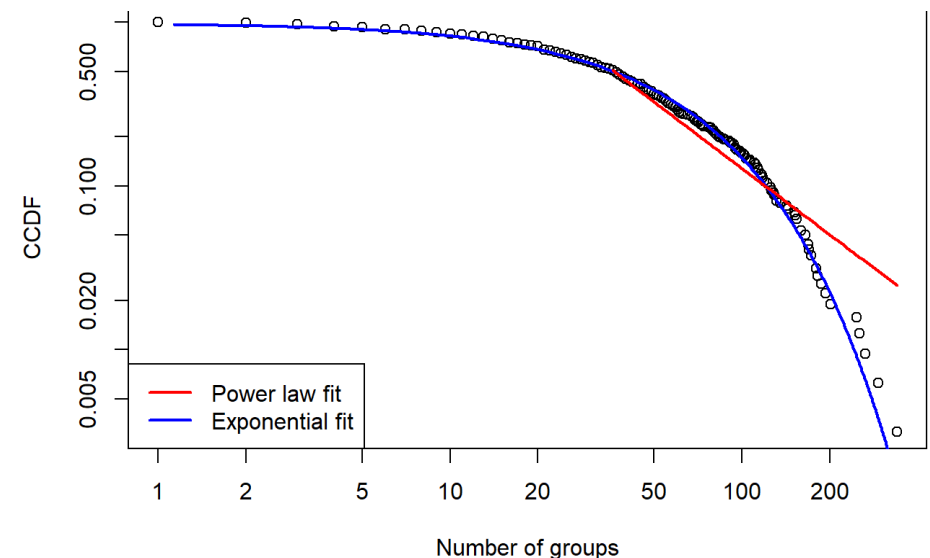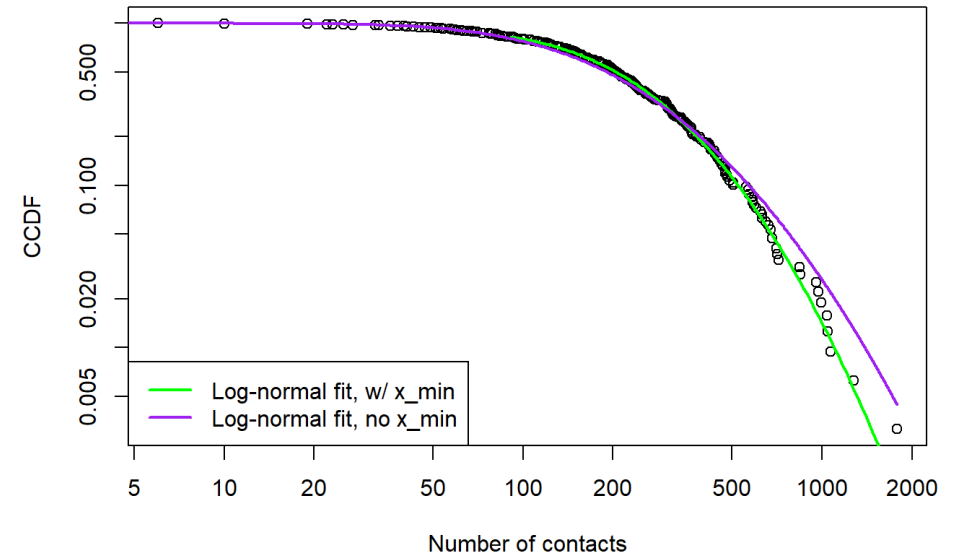
Current work
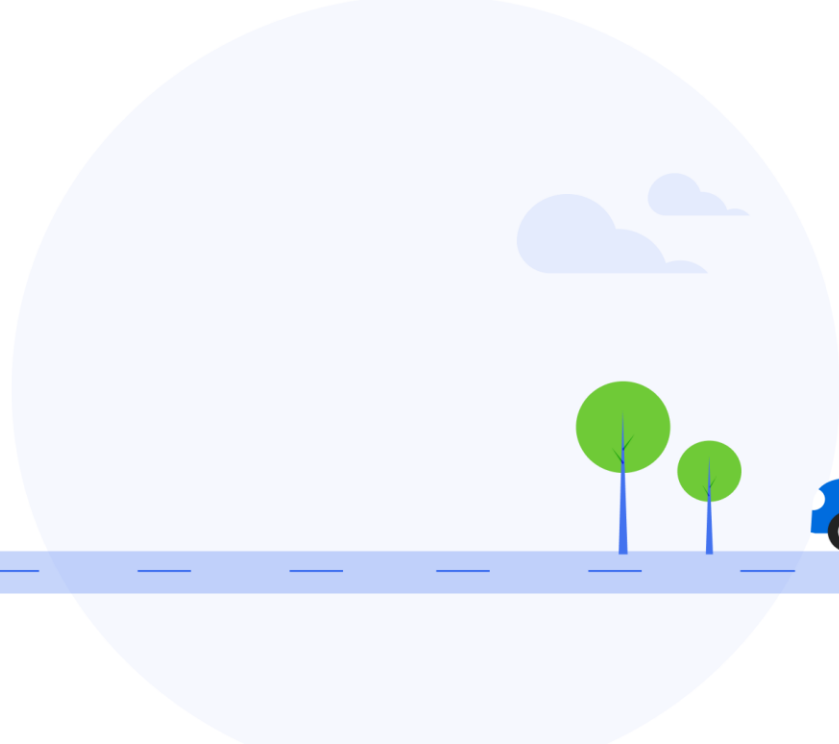
1. Applications

Some examples from 2023 & 2024
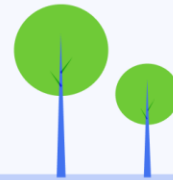
Results: Social network structure

Work by [Corten et al. (2024)](#)

- Donation of WhatsApp account data in the LISS panel.

- Understanding of topological features of instant messenger networks.

- Degree distribution of contacts and of group membership.

Results: Debunking Netflix myths

Work by Van Es et al. (2024)

- Use data donation to critically assess claims made by streaming platforms.

- 126 donations of Netflix data through the I&O Research panel

- Myth example: "Binge-watching is the new normal"

	2 Watched	3 Watched	4 Watched	5 Watched	6 Watched
# users	116 (92%)	108 (86%)	88 (70%)	69 (55%)	53 (42%)
% total sessions	32.2	14.8	8	4.7	2.8

Results: Election campaigns and YouTube media exposure
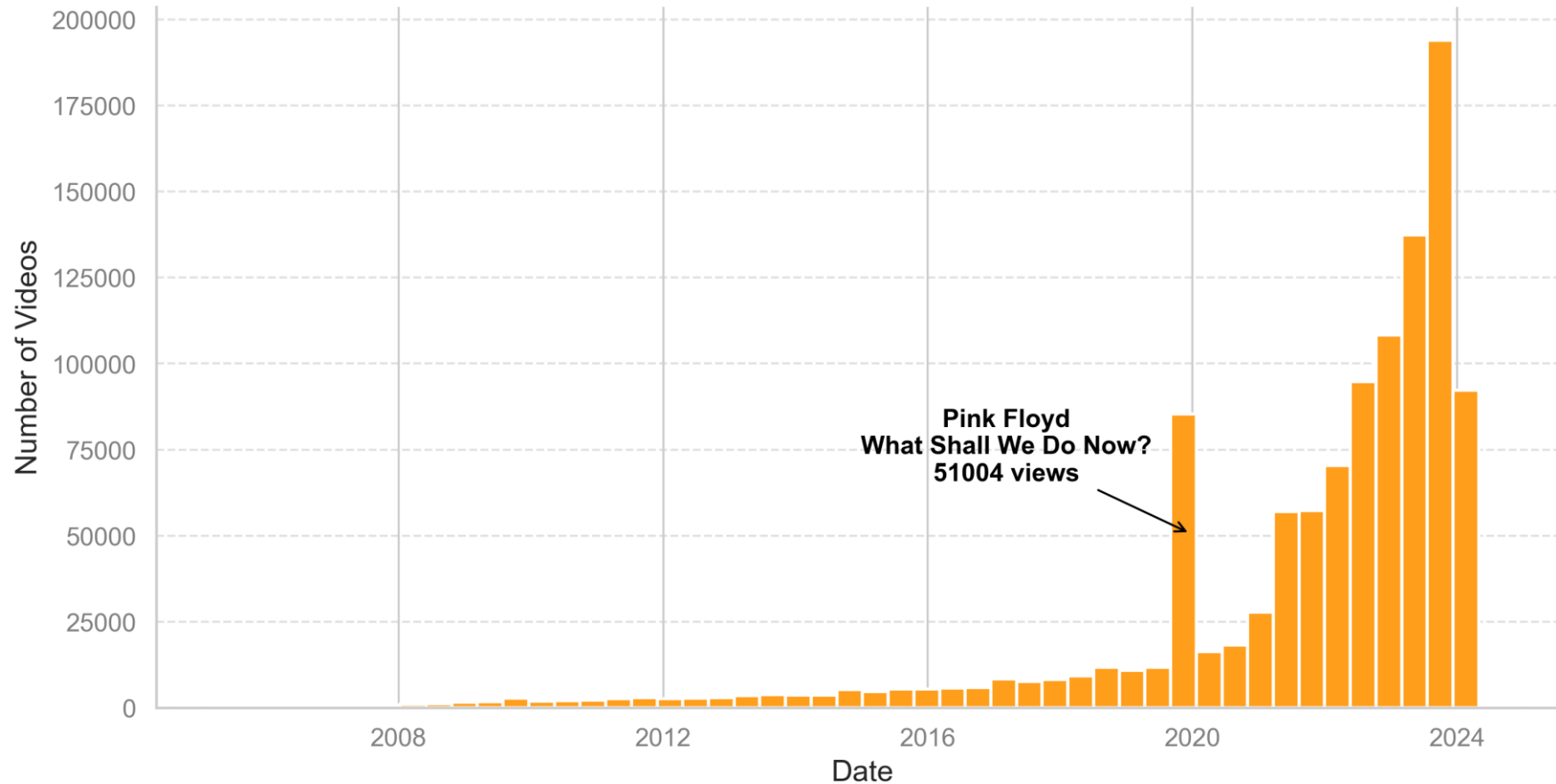
Work by Loecherbach et al. (2024)

Age Distribution Comparison Across Datasets
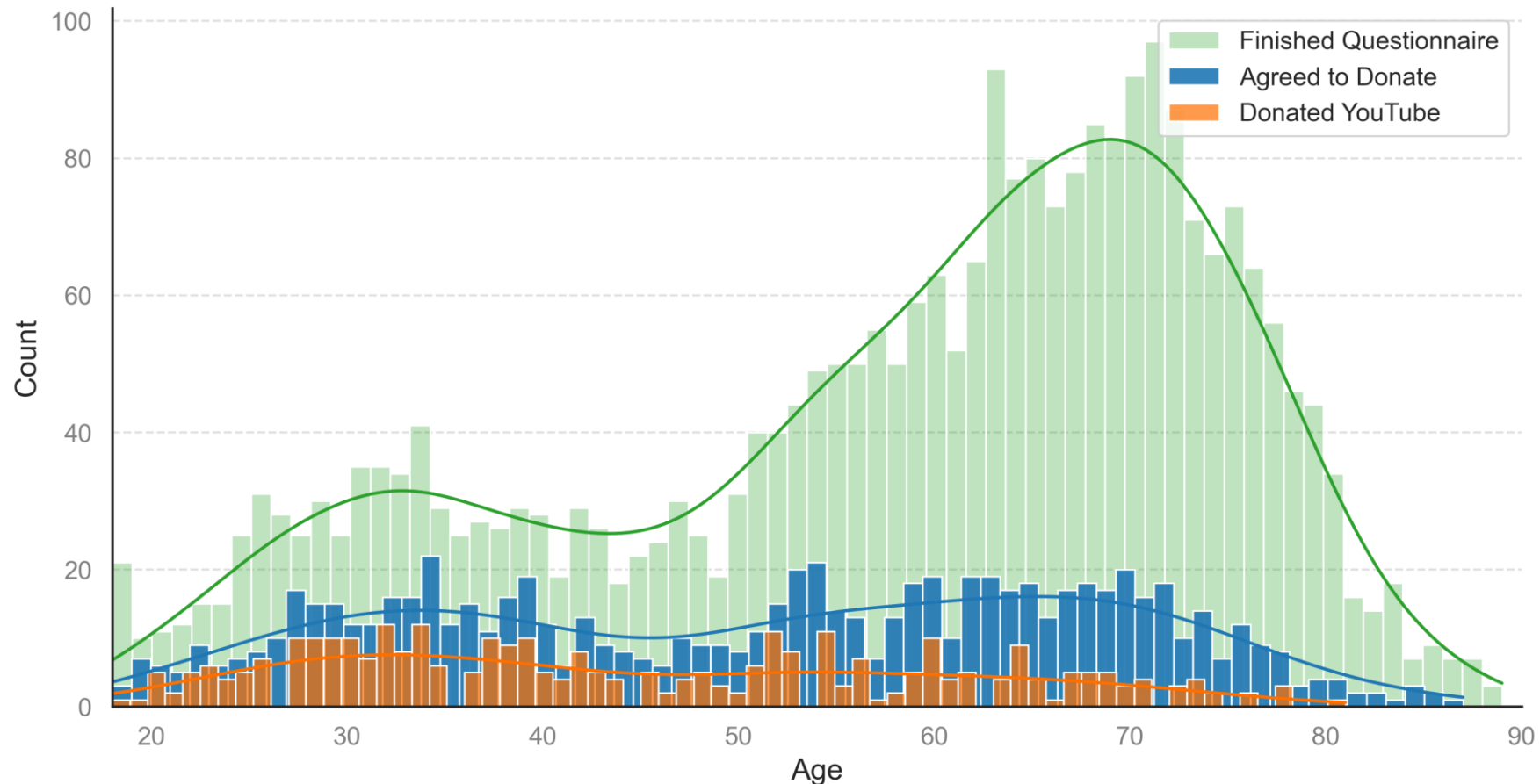
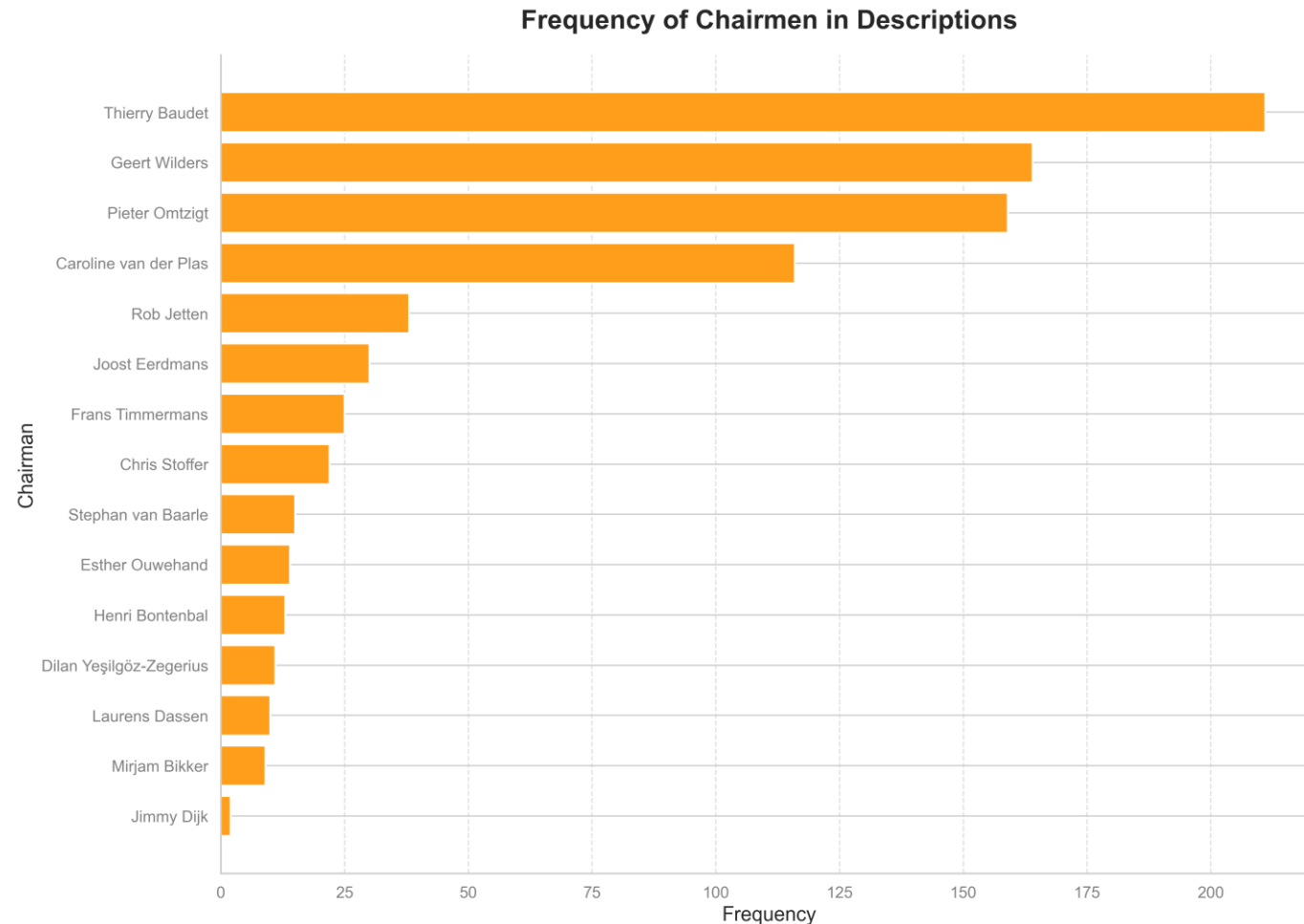# Results: Election campaigns and YouTube media exposure

Work by Loecherbach et al. (2024)

Publication Date Donated Videos

Results: Election campaigns and YouTube media exposure

Work by Loecherbach et al. (2024)

Frequency of Chairmen in Descriptions
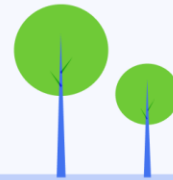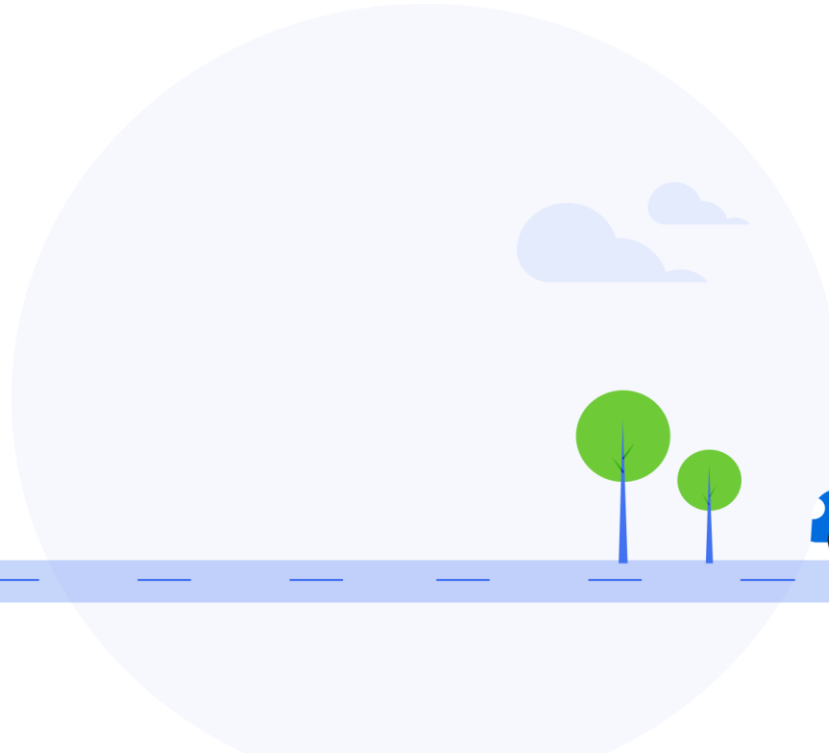
Some other examples

- How are adolescent social media use and well-being related? (TikTok, Instagram)

- How physically active are elderly people?(Google Location History, Apple Health, Fitbit)

- How do people interact with GenAI? (ChatGPT)

- How connected are employers outside working hours? (Slack)

2. Understanding representation issues

Reasons for non-donation

Ongoing work by Hase et al. (2024)

Reasons participants provide for declining an invitation to donate in a study in the Netherlands (N=1,163) and Germany (N=2,039)

- Concerned about their privacy:
  - Netherlands 53%
  - Germany 40%

- Considered their data "irrelevant" for the study:
  - Netherlands 4%
  - Germany 26%

Understanding the data donation task

Ongoing work by Struminskaya et al. (2023)

Statements asked to respondents	Correct %	Incorrect %	Don't know %
You are asked to download information from Google. TRUE	48.8	19.8	31.4
The software implemented in the survey will extract the information on the number of hours you cycle, walk, take public transport, travel by car. TRUE	62.3	6.1	31.2
Information on all the locations you visited will be shared with Centerdata. FALSE	39.2	31.4	29.4
Google collects information on location about everyone. FALSE	24.8	46.6	28.5
From the data you will provide, the information can be traced back to you. FALSE	45.3	22.2	32.5
You will be able to inspect the data before sending it to Centerdata. TRUE	59.0	7.8	33.1
It is impossible to identify you as an individual from the data that you provide. TRUE	43.4	19.6	37.0

Understanding the data donation task

Ongoing work by Struminskaya et al. (2023)

- **5.5% had everything correct**
- Mean correct: 3.23, median = 4 (out of 7 questions)
- **People with more correct answers more likely to be willing & to donate:**
  - 4.54 correct statements for willing
  - 2.56 correct statements for non-willing
  - OR = 1.572, p <.001

  - 5.33 correct statements for donated
  - 3.94 correct statements for not donated
  - OR = 1.795, p <.001

Understand how participants experience data donation

Work by Groot-Kormelink et al. (2024)

- 20 participants donated data from three platforms of choice in a lab and carried out a think-a-loud protocol and in-depth interviews.

- Ages between 18 and 72

- Highly varying digital skills

# Understand how participants experience data donation

Reasons to participate

- Insight in their own media use.

- Monetary compensation.

- Altruistic reasons: "helping" the researchers.

- Importance of the study for society at large.

DDP inspection

- Surprised and impressed reactions to their own data.

- Understanding how to navigate tables and figures.

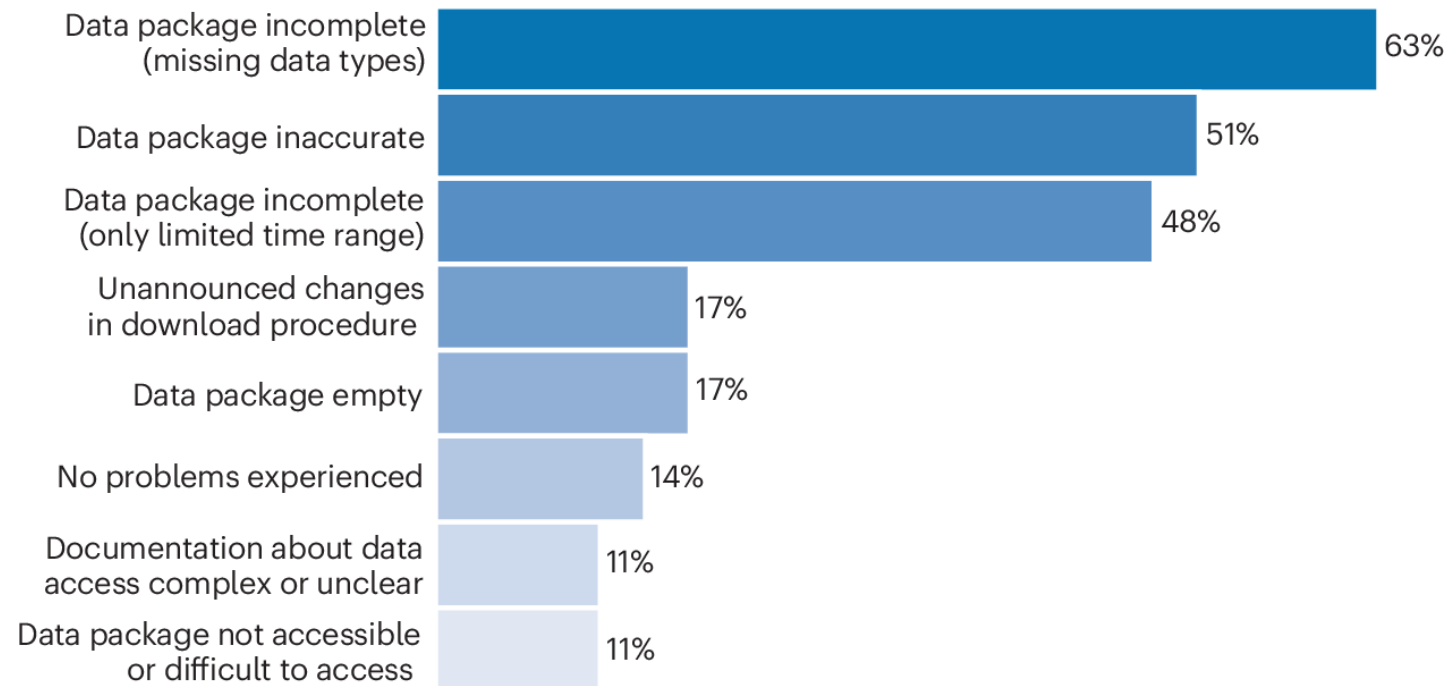- Quickly scrolling down to the donate button.

3. Monitor platforms
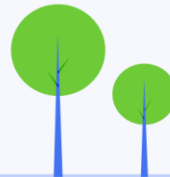
What issues do DD researchers encounter?

Work by Valkenburg, P.M. et al., (2024).

Common problems in platform data donation experienced by researchers

Identify the issue of interpretive entrepreneurship

Work by Janssen, H. et al., (2024).

- Conducted two studies:
    1. Evaluate privacy policies and DDPs on GDPR elements.
    2. Completeness of DDPs in light of the privacy policies.

- Data used:
    - Facebook, Google, Instagram, Netflix, Spotify, TikTok, WhatsApp & X(Twitter).
    - New accounts created and used for six months to investigate DDPs + matching privacy policies.
    - Data used from May 2023.

Identify the issue of interpretive entrepreneurship

Work by Janssen, H. et al., (2024).

- Platform practices in terms of access rights indicate (room for) interpretive entrepreneurship:
  - Interpretation of the law according to a platform's preference or interest.
- Platforms sometimes don't even comply with their own privacy policies!

- This is harmful for individuals, but also researchers.
- We call for regulator to reconsider compliance and enforcement.

Suggestions to improve legislation

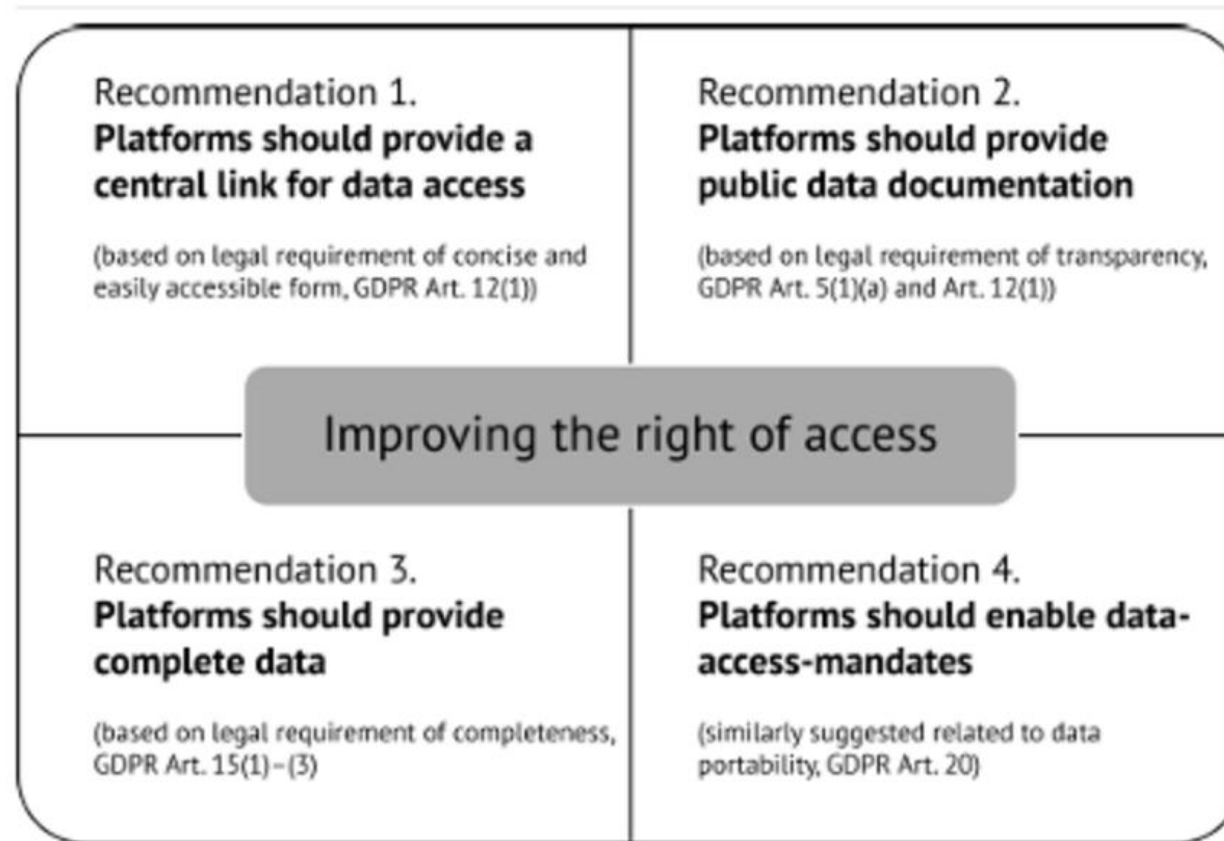*Work by Hase, V. et al., (2024).*

Figure 1: Key recommendations.

4. Integrate with ODISSEI facilities

Plans in the SSHOC-NL roadmap:

1. Integrate with the LISS panel.

2. Facilitate findability of data donation datasets through the ODISSEI portal.

3. Facilitate re-use of data donation datasets through SANE.

5. Make it easier to do studies

1. Maintain Python scripts for commonly used platforms.
2. Create a library of commonly used functions.
3. Exploring new functionalities
    - E.g. local model training

What if you want to do a study yourself?

1. **Try out the platform you are interested in.**
    1. Does it collect the data that you had in mind?
    2. How easy is it to collect the data?
2. **Determine who is in your team.**
    1. Do you work with an existing panel?
    2. What are your technical skills?
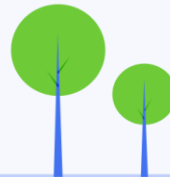   3. Do you have IT support?
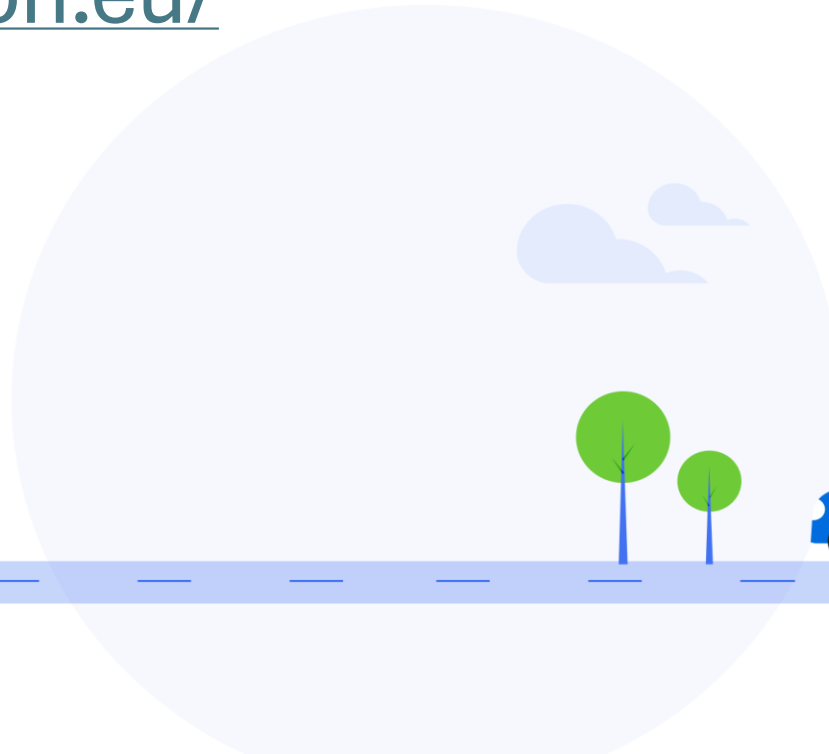3. **Determine your budget.**
    1. Is there for example budget available for SaaS?
    2. Will you reimburse your participants?
4. **Reach out to me to discuss options!**

Questions?

l.boeschoten@uu.nl
https://datadonation.eu/

References

- Boeschoten, L. et al. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, *4*(2), 388–423. https://doi.org/10.5117/ccr2022.2.002.boes

- Boeschoten, L. et al. (2023). Port: A software tool for digital data donation. *Journal of Open Source Software*, *8*(90). https://doi.org/10.21105/joss.05596

- Corten, R. et al. (2024) Assessing Mobile Instand Messenger Networks with Donated Data, *SocArxiv.* https://doi.org/10.31235/osf.io/x4ku2

- Van Es, K. et al. (2024) Studying Global VoD Services with Data Donations: A View From Netflix, *Under review.*

- Groot-Kormelink, T. et al. (2024) A user-centric approach to digital trace data: How participants experience and understand data donation, *Under review.*

- Hase, V. et al. (2024). Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies?. *Internet policy review: Journal on internet regulation, 13*(3). https://doi.org/10.14763/2024.3.1793

- Janssen, H. et al. (2024) Interpretative entrepreneurship shaping data access under under the GDPR*, In preparation*.

- Loecherbach, F. et al. (2024) Dissecting Digital natives' Political Preferences: Election Campaigns and YouTube Media Exposure*, In preparation*

- Valkenburg, P.M. et al., (2024) It is time to ensure research access to platform data. *Nature Human Behavior.* https://doi.org/10.1038/s41562-024-02066-5