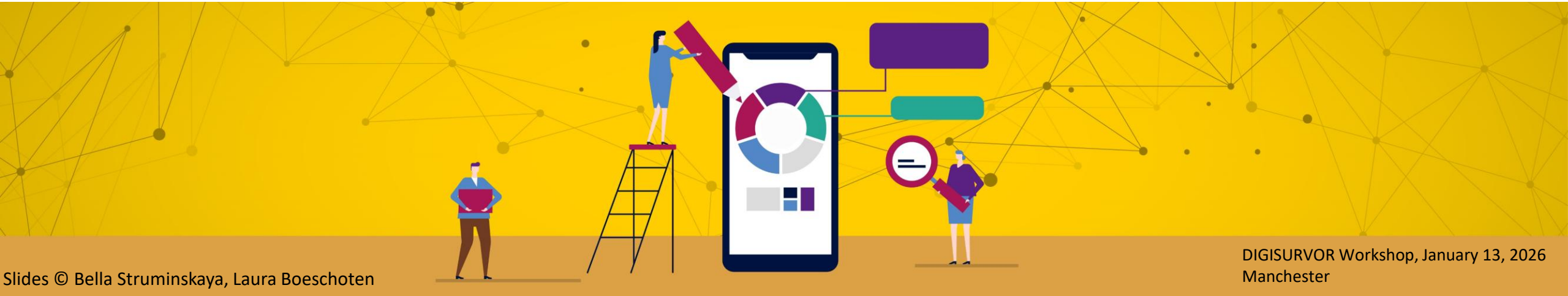
Utrecht University

Building sustainable software-centered research infrastructures to support digital data collection

Bella Struminskaya
Utrecht University



DIGISURVOR Workshop, January 13, 2026
Manchester

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

Funders

SSHOC NL social sciences & humanities open cloud

PDI SSH

Team Research App

Researchers

Bella Struminskaya
PI

Thijs Carriere
PhD candidate

Niek de Schipper
Research engineer

Jonathan Koop
Research assistant

Dion van Harten
Research assistant

Collaborators

Eyra

SURF

CENTERDATA research institute

Funders

NWO

Data Donation (D3I)

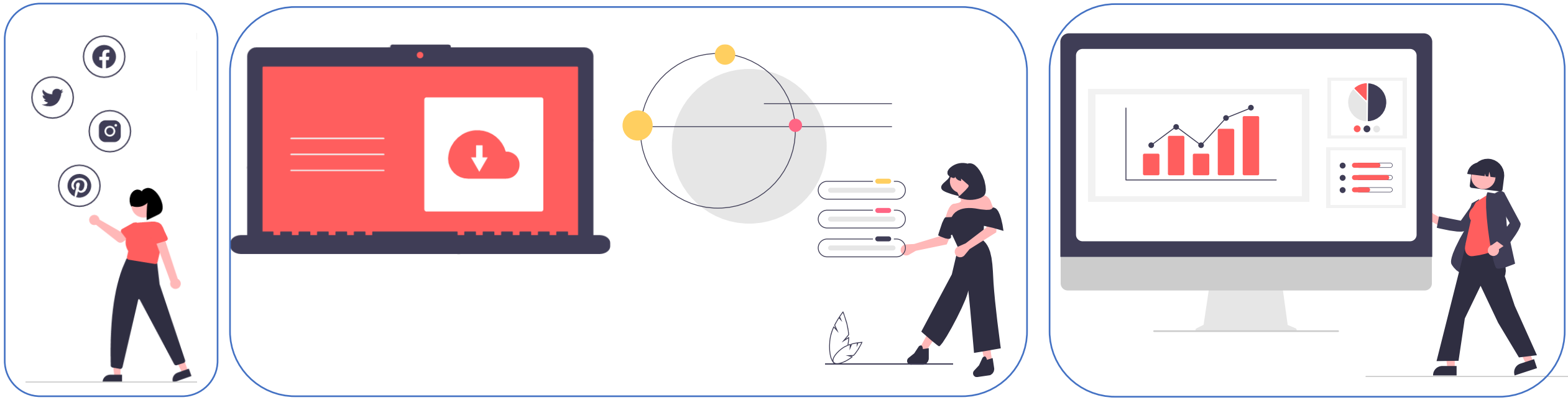
The data donation workflow

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

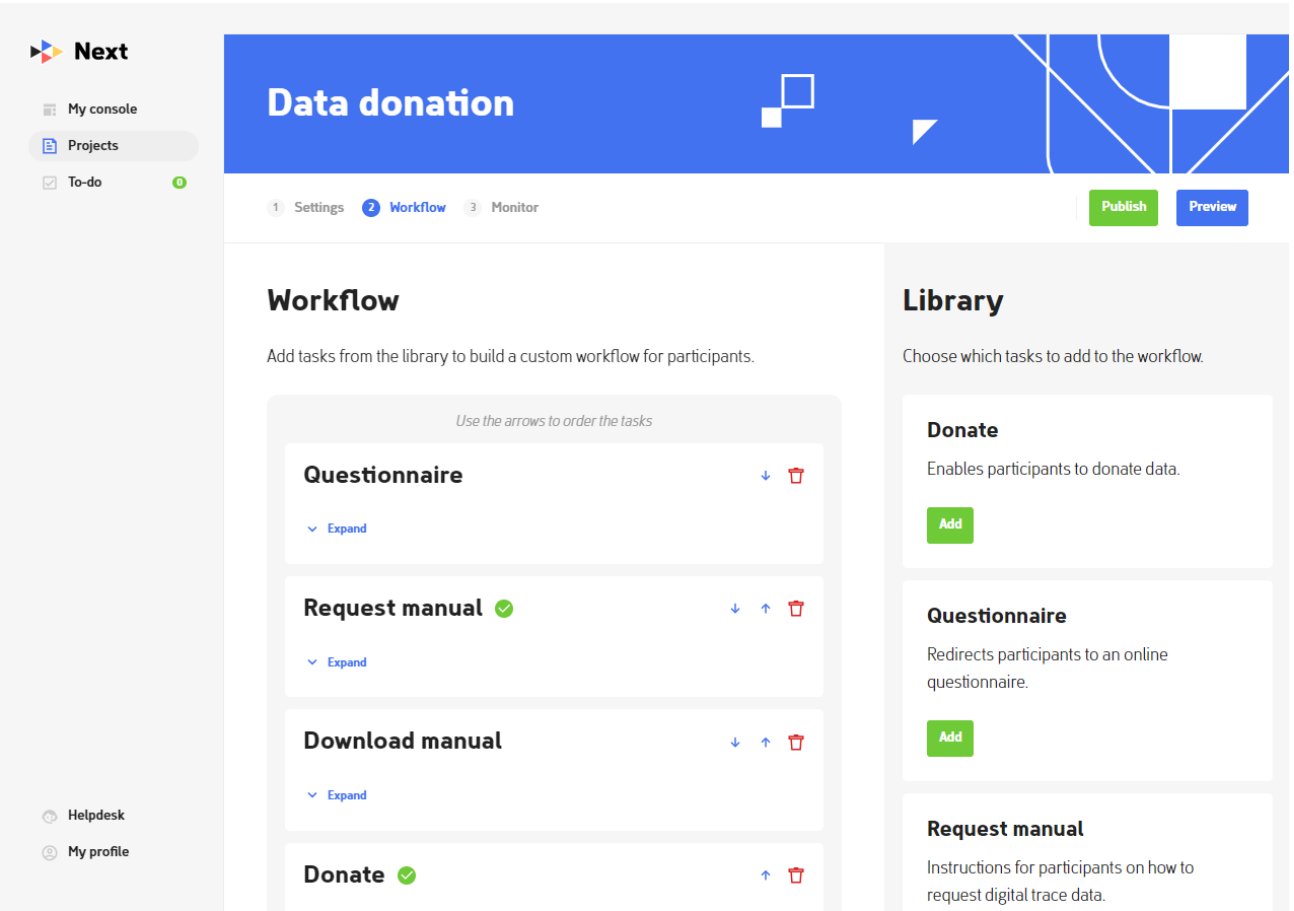


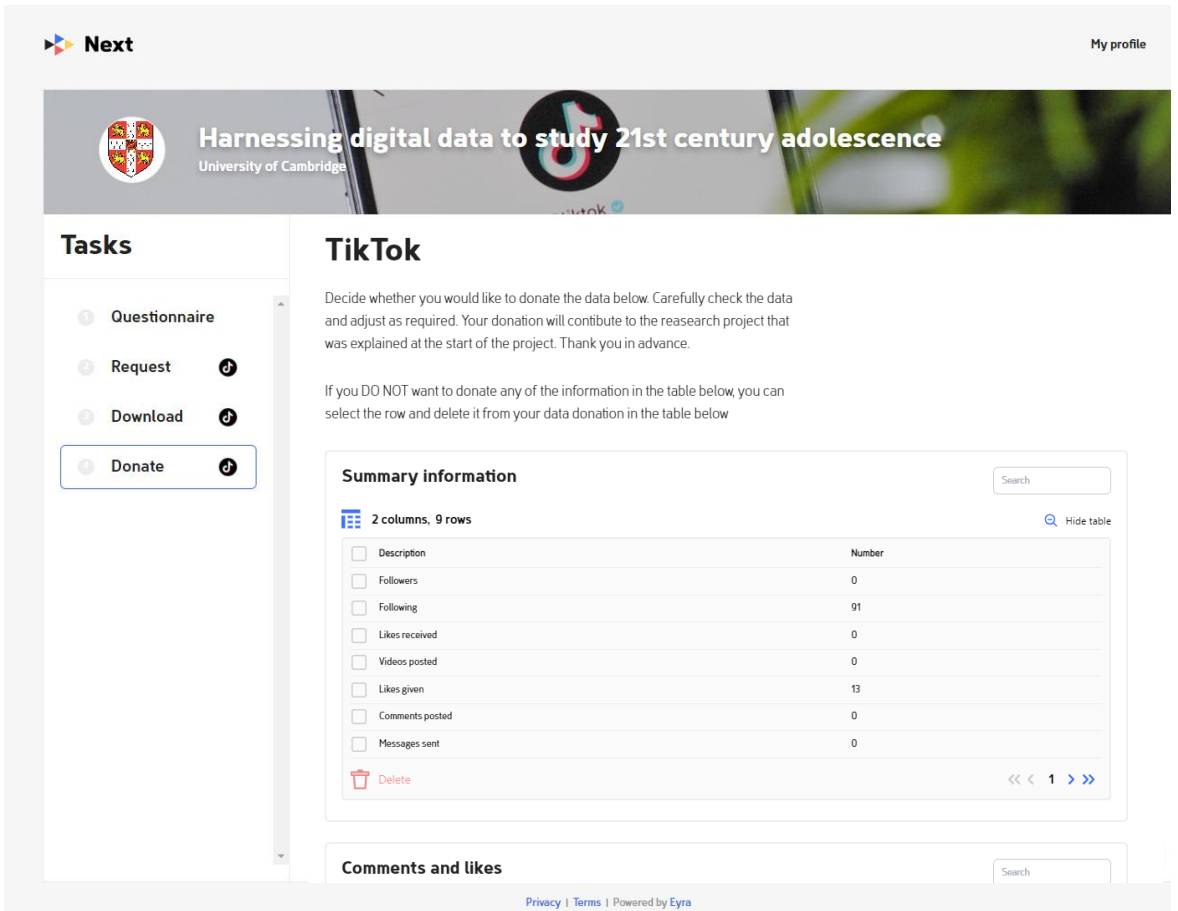
Receiving personal data in structured, commonly used, and machine-readable format ("Data Download Package"; DDP) & transmitting data to another data controller

Images: https://vecteezy.com

Port – Software for data donation

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





How to use Port?

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 DIY
- Arranging participant recruitment and data storage can be related to how one uses Next.

2. The Donate Task

- To account for different platforms & DDP parts, a custom Python script needed (https://github.com/d3i-infra/data-donation-task)
- Elements of the Python script:
    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 extracted data on screen (in a table)
    5. Send the data to the data storage upon consent

Success stories – completed projects

datadonation.eu

Data donation | About the project | Our software | Prepare a study | Community

Study design

Privacy and consent

Look at completed projects

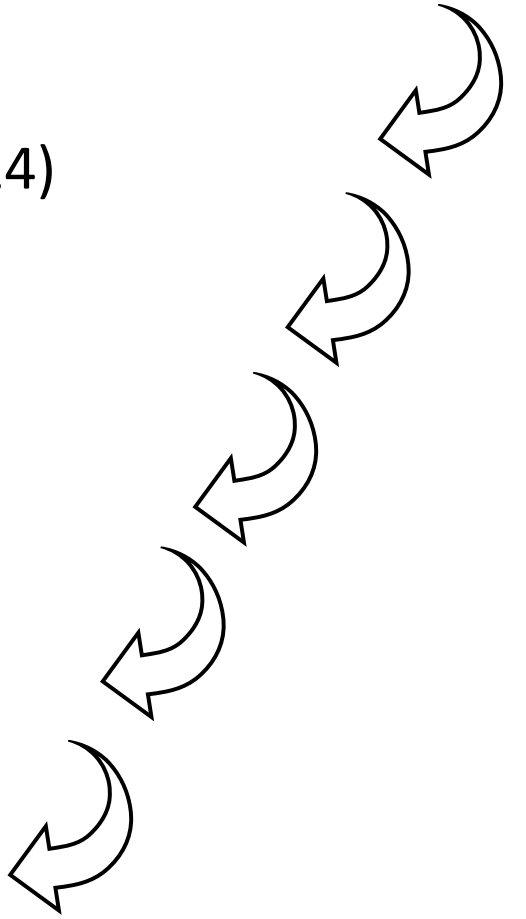
Look at completed projects

Below, you can find a selection of projects that have been completed in the last years by making use of Port.

Behind the screens: Exploring Netflix via Data Donations

A note on funding

- OSD2F (Araujo et al.; An Open-Source Data Donation Framework, University of Amsterdam)
- NWO Vidi Project Daniel Oberski (2020-2024)
- PDI-SSH-funded D3I project (2022-2025)
- TDCC “RIGHTS” (2025-2027)
  Thematic Digital Competence Centers
- SSHOC-NL (2024-2028)
  Social Science and Humanities Open Cloud for the Netherlands
- Structural funding D3I (2025-permanent)



Research App

Privacy concerns

↓

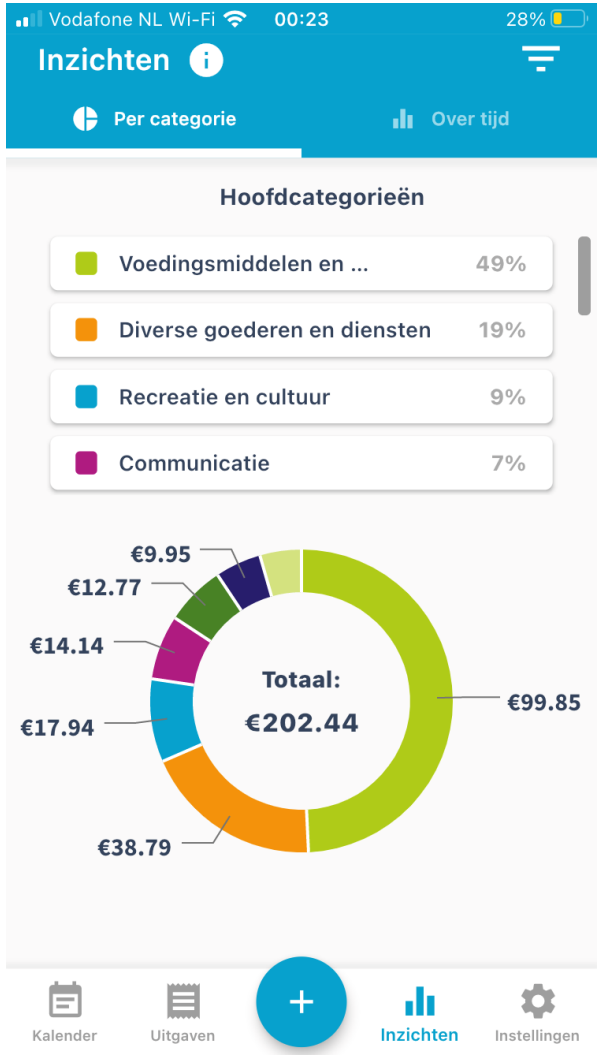
Selective participation

↓

Biased results

Struminskaya et al. 2020; 2021; Keusch, Struminskaya et al. 2019; 2020; Struminskaya & Keusch 2019

Insights for participants in app-based studies



Survey Research Methods (2025)
Vol. 19, No. 1, pp. 25-41
https://doi.org/10.18148/srm/2025.v19i1.8263
European Survey Research Association

Nonresponse and Dropout in an App-Based Household Budget Survey: Representativeness, Interventions to Increase Response, and Data Quality

Evelien Rodenburg[1] · Barry Schouten[2] · Bella Struminskaya[1]
[1]Utrecht University, Department of Methodology and Statistics
[2]Statistics Netherlands

Household budget surveys struggle with low response and participation rates, and lower data quality, in part due to a high respondent burden. App-assisted budget surveys may provide solutions to both these problems. This cross-country study carried out in the Netherlands, Luxembourg, and Spain, investigates the use of an app-based diary for collecting household expenditure data compared to a web-based method. We report the results of two randomized experiments: 1) using personalized feedback and 2) interviewer-assisted versus mail recruitment in terms of influence on response and participation rates. The app-based household budget survey yields slightly higher registration, activity, and completion rates compared to the web-based household budget survey that we use as a reference. We find disproportionate representation of certain groups in the app-based sample, but no substantial differences in the overall representativeness between the app-based and web-based samples. Providing households with personalized feedback does not affect registration or activity in the app. Using interviewers for recruitment does increase registration and activity rates, although this negatively affects the representativeness of the sample. Neither providing personalized feedback nor using interviewers for recruitment significantly affects dropout during the study or data quality. We also find no substantive differences between the quality of web-collected expenditure data and data collected in the app. Overall, using an app could be suitable for collecting expenditure data especially in combination with the use of interviewers for recruitment. However, this may come at a cost to representativeness.

Keywords: app-based surveys; household budget surveys; participant feedback; interviewers, data quality

Rodenburg, Schouten, Struminskaya, 2025, *Survey Research Methods*

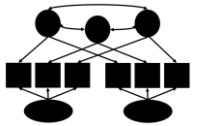
Research app for the (social) sciences

Participant in focus: fundamental research on **privacy concerns & personalized feedback**

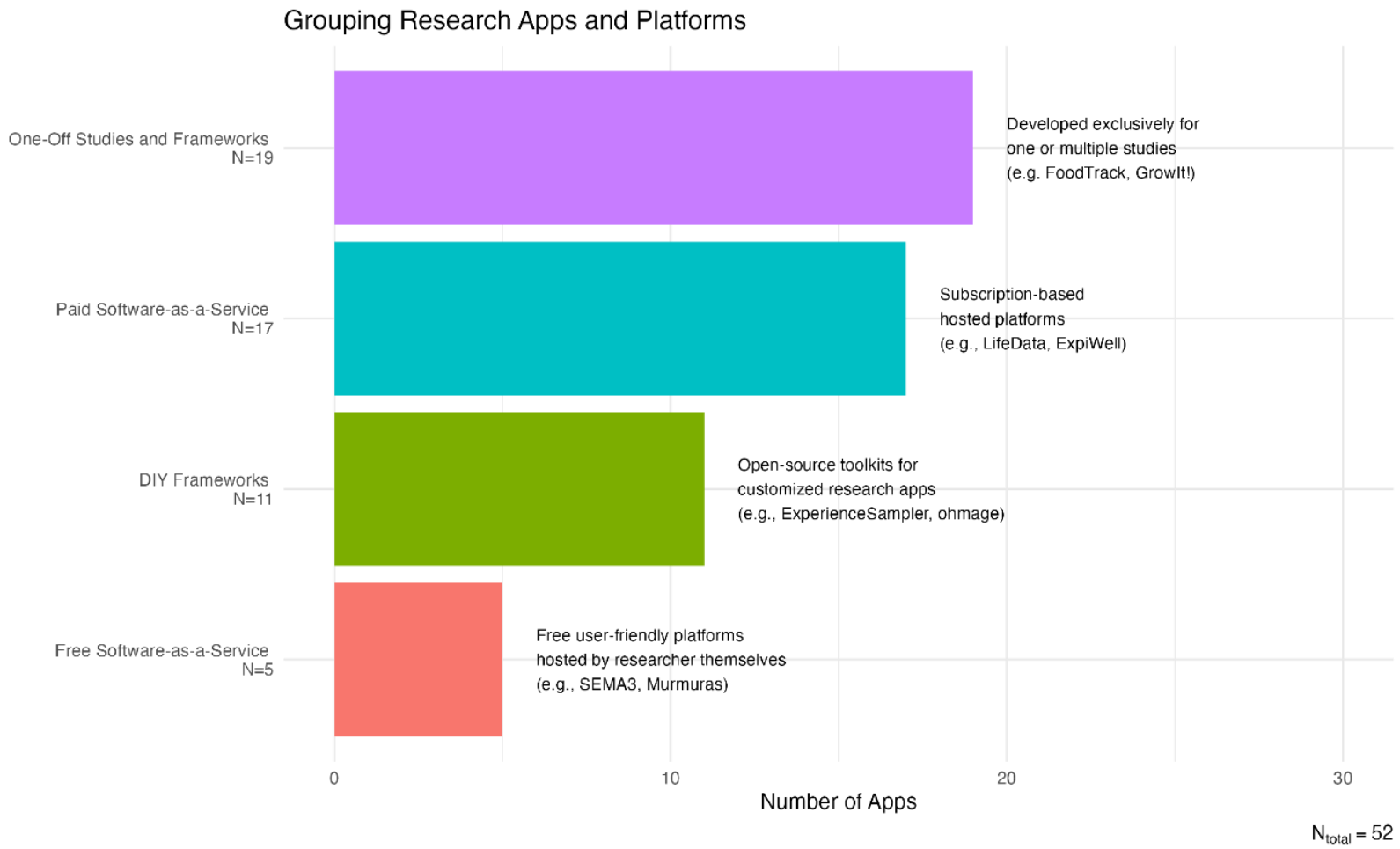
Open-source **research app**: developed for and with social scientists

Application studies: **social science pilot** studies

Ensuring **valid measurement** for self-report & app data



Apps landscape (avoiding the 'graveyard' of apps)



- Work in progress
- Lit search in 2024
- Publications on technical specs of an app/platform, or w/ description of the app development in a substantive paper
- Extracted info on functionalities, options to control data collection, view collected data, feedback, nudging, gamification, triggering of EMA, research area, open source, maintained, developed for a specific project etc.

Vision: maintained, sustainable, part of larger infrastructure

For (early career) researchers from the social sciences and related disciplines
Who wish to collect survey data (including EMA), geolocation data, app usage data, and psychical activity data using a mobile research app through the sensors of an Android or iOS smartphone.
The research app
Is an (affordable) open-source research app platform
That allows participants to have control over their data, addressing privacy concerns and provides personalized insights.
Unlike other apps that are either proprietary/not open source, expensive, provide limited functionalities, are no longer maintained, are not designed to be re-used by the developing team or other researchers, or are isolated (not integrated in larger research infrastructures).
Our solution is an app that is configurable by researchers to collect smartphone sensor data, based on methodological research about apps, has a user-friendly back-end to configure the app for the participant and is part of the larger digital data collection infrastructure of the Netherlands.

Utrecht University

NWO

Eyra

CENTERDATA research institute

Key considerations (discussion)

- Open source vs. proprietary
- Software development in-house vs. externally
- Financing software development through external funding
- Scalable vs. non-scalable software elements
- User-friendliness
- Responsible software use
- Restructuring software with development progression
- Long-term maintenance strategy
- Outreach/community engagement

Funding acknowledgements

Some work in this presentation is part of the project "Improving the methods of digital behavioral data collection for the social sciences" with file number VI.Vidi.221S.031 of the research programme NWO Talent Programme Vidi SSH - Social Sciences which is financed by the Dutch Research Council (NWO).

Digital Data Donation Infrastructure Project (D3I) was made possible by the Platform Digitale Infrastructuur SSH (PDI SSH).

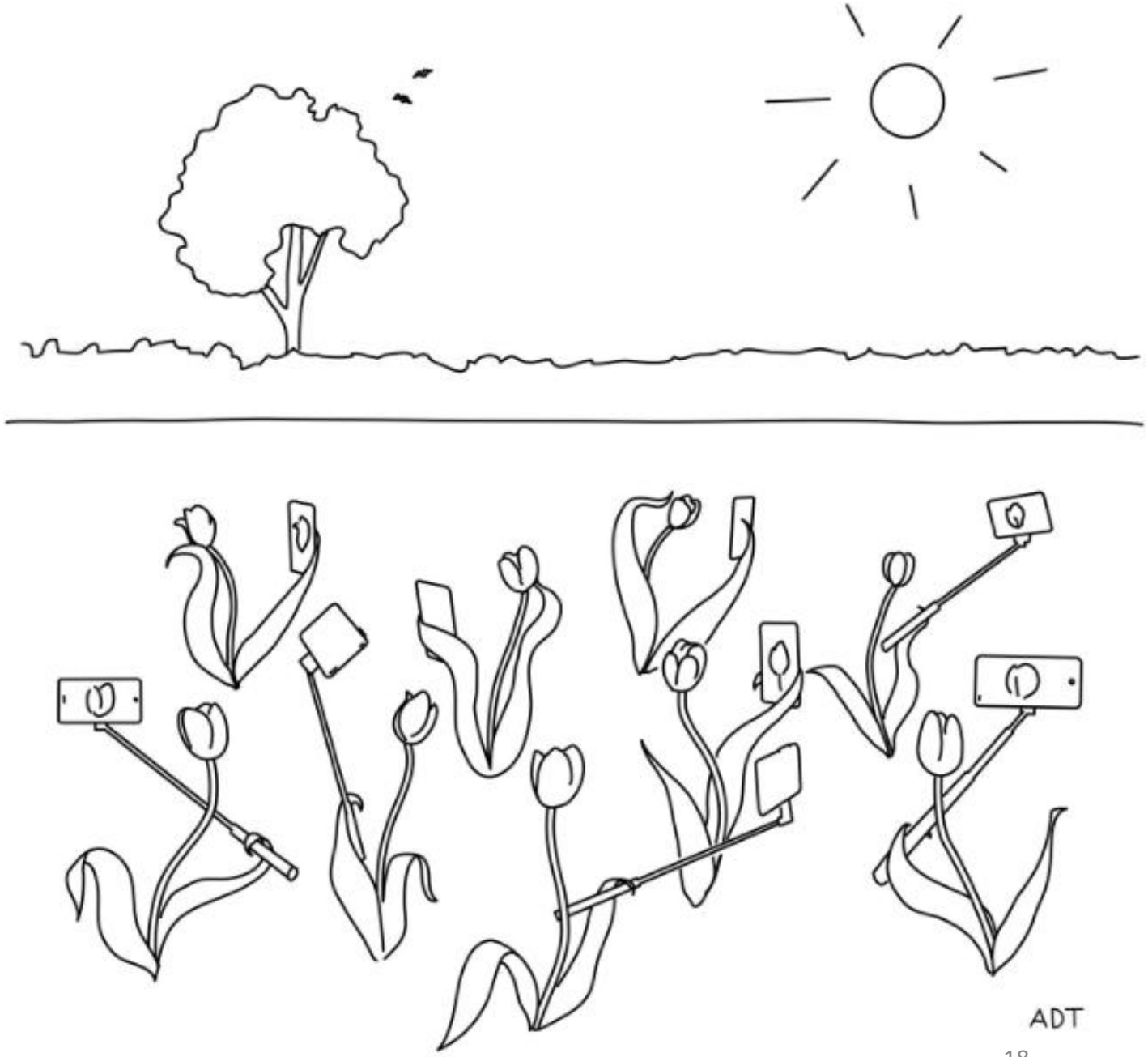
Thank you!

Contact:

b.struminskaya@uu.nl

https://bellastrum.com

@bellastrum



A.D. Thompson, The New Yorker