

# Digital Data Donation Infrastructure – D3I

Data Donation Day, 11.03.22

*Theo Araujo*

# Agenda

- The project
- Deliverables
- How we will get there
- Current experiences
- Getting involved



# The project

This project is supported by the Platform Digitale Infrastructuur SSH (<http://www.pdi-ssh.nl>)

## Key aim

Create a (national) infrastructure that allows SSH researchers to work with individuals willing to donate their data to academic research

*Explicit focus:*  
Social Sciences and Humanities research

# Funding programme

- Project runs from 2022 to early 2025
- Platform Digital Infrastructure Social Sciences and Humanities
- Funding provided for *infrastructure* (i.e., technical) development
- Infrastructure to be offered nationally during/after the project

# Team

- Bella Struminskaya (Faculty of Social and Behavioural Sciences, UU)
- Claartje ter Hoeven (Erasmus School of Social and Behavioural Sciences, EUR)
- Daniel Oberski (Faculty of Social and Behavioural Sciences, UU)
- Enny Das (Center for Language Studies, RUN),
- Henk van den Heuvel (Centre for Language and Speech Technology and Head of the Humanities Lab, RUN)
- Jef Ausloos (Faculty of Law, UvA)
- Judith Möller (Faculty of Social and Behavioural Sciences, UvA)
- Karin van Es (Faculty of Humanities, UU)
- Laura Boeschoten (Faculty of Social and Behavioural Sciences, UU)
- Maarten van Gompel (KNAW Humanities Cluster & Centre for Language and Speech Technology, RUN)
- Marieke Fransen (Behavioural Science Institute, RUN)
- Marjolijn Antheunis (Tilburg School of Humanities and Digital Sciences, TU)
- Mirko Schaefer (Faculty of Humanities, UU)
- Theo Araujo (Faculty of Social and Behavioural Sciences, UvA)
- Wouter van Atteveldt (Faculty of Social Sciences, VU)
- Wyke Stommel (Center for Language Studies, RUN)

# Consortium



**Erasmus University Rotterdam**



**Radboud University**



**Universiteit  
Utrecht**





# Deliverables

This project is supported by the Platform Digitale Infrastructuur SSH (<http://www.pdi-ssh.nl>)



## Overall flow

D3I enables researchers to:

1. **Connect** the researchers' own data collection tools and processes (e.g., online surveys, interviews) to an infrastructure through which respondents can donate their data.
2. **Instruct** willing individual respondents on how to download their own data (Data Download Packages: DDPs) from a set of platforms.
3. **Process** DDPs to ensure that only the strictly necessary data are shared with researchers (data minimisation).
4. **Visualize** the DDP data such that it fosters individuals to exert agency on what (not) to donate, and thus to provide *meaningful informed consent*.
5. **Export** the processed donated data in formats that can be used by researchers across a variety of SSH fields.
6. **Inform** individuals who donate their data by providing insights on the research this generates, increasing awareness of the data that digital platforms have on them.

# I. Connect to researcher's own data collection tools and processes

## SSH Participant Recruitment Tools & Processes

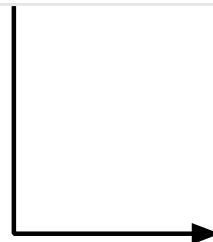
Examples: PDI-SSH Panel, LISS, universities' own participant pools, panel recruitment agencies etc.)

Researchers recruit individual participants for SSH studies

Participants provide self-reports (e.g., answers to online survey)

## Digital Data Donation Infrastructure (D3I)

Donation of individual platform usage data (linked to self-report)



## 2. Instruct individual respondents on how to download their own data

### Hoe kan je je Facebook data downloaden?

Facebook geeft gebruikers de mogelijkheid om delen van eigen data te downloaden. We vragen je om een deel van die data te downloaden waar jij alleen toegang tot hebt en die aan ons te doneren zodat wij die informatie kunnen gebruiken binnen ons wetenschappelijke onderzoek.

We leggen nu stap voor stap uit hoe dit werkt. Aan het einde van dat proces heb je een deel van je Facebook data gedownload als een .zip file of een folder.

Mocht je enige help willen tijdens het proces, kan je ook de Facebook Help Pagina bekijken: <https://www.facebook.com/help/212802592074644>

---

**Stap 1.** Ga naar [facebook.com](https://www.facebook.com) en volg de instructies zoals deze in de volgende plaatjes staan.

Als je een kopie van je gegevens op Facebook wilt downloaden, kun je de functie Je

3. Process DDPs to ensure that only the strictly necessary data are shared with researchers

**Figure 1.** Selection of information contained in a Facebook Data Download Package (DDP).

*Note: A research project would likely only use (and collect) a subset of these data*



4. **Visualize** the DDP data so individuals can exert agency on what (not) to donate

### Select file(s):

You can use the file selector to select the zipfile from your platform

You can also drag the folder into this box

No file chosen

## 5. Export the processed donated data in formats that can be used by researchers

### *Examples of use cases*

- Digital divide and digital literacy research.
- Investigations of causes, contents, and consequences of online interactions with individuals, groups or the media.
- Quantification of chronemic and linguistic aspects of online interactions and language evolution.
- Microanalysis of online/digital interactions to understand the native, embedded social and communicative behaviours of users.
- Studies on the role of platforms and algorithms in the digital society.
- Natural language processing research.

*Different use cases = potentially different requirements on how to collect, process and export data*

**6. Inform**  
individuals,  
increasing  
awareness of the  
data that  
platforms have  
on them.

- Provide relevant data visualisations to a participant so they can understand what data platforms have on them, and get insights on their own online activities and behaviour



# How we will get there

This project is supported by the Platform Digitale Infrastructuur SSH (<http://www.pdi-ssh.nl>)



# Project

- Integration of PORT (UU) and OSD2F (UvA-VU)
- Extension of the infrastructure to:
  - Improve data visualisation and user friendliness
  - New ways to process and combine data
- Legal and methodological blueprint – i.e., how can we make this type of research *feasible* in NL universities?
- Pilots to ensure the infrastructure can help diverse use cases for Social Sciences *and* Humanities researchers

# Work Packages

- **WP1. Toolkit to enable data donations to academic studies** (back-end)
- **WP2. User interface that ensures individual agency and informed consent** (front-end)
- **WP3. Privacy-friendly local processing and measurement** (front-end + local data analysis)
- **WP4. Baseline infrastructure for top platforms for SSH research** (pilots for Facebook\*, Google\*, Spotify, Netflix & Uber)  
\* And their properties (e.g., Instagram or YouTube)
- **WP5. Methodological and ethical & legal framework** (including legal discussion with partner universities)



How to get involved

How to get  
involved

We would **really** like to talk to you about your ideas – and needs – for data donation-related projects

