



Digital trace data collection through data donation

Dr. Laura Boeschoten

Postdoc



- 1. Background*
- 2. Workflow + Software*
- 3. Illustration*
- 4. Next steps*

Dr. Laura Boeschoten
Postdoc



1. Background

Dr. Laura Boeschoten

Postdoc

1. Background

Digital trace data

- Everywhere
- Everything
- Everyone

Measurement quality?



1. Background

Drawbacks current procedures

- Selective sample
- Limited access
- Aggregated format
- Only public info
- No informed consent

Alternative: Tracking apps





1. Background- How to overcome these issues?

What we want:

- Ability to draw a sample
- Access to (private) micro data
- With informed consent

GDPR

All data subjects have the...

- right of data access
- right of data portability

1. Background – Data Download Packages (DDPs)

All data controllers comply.

Data subjects can request
their complete personal
archive in a digital
transportable format.



**PRIVATE DATA
HANDLE WITH CARE**

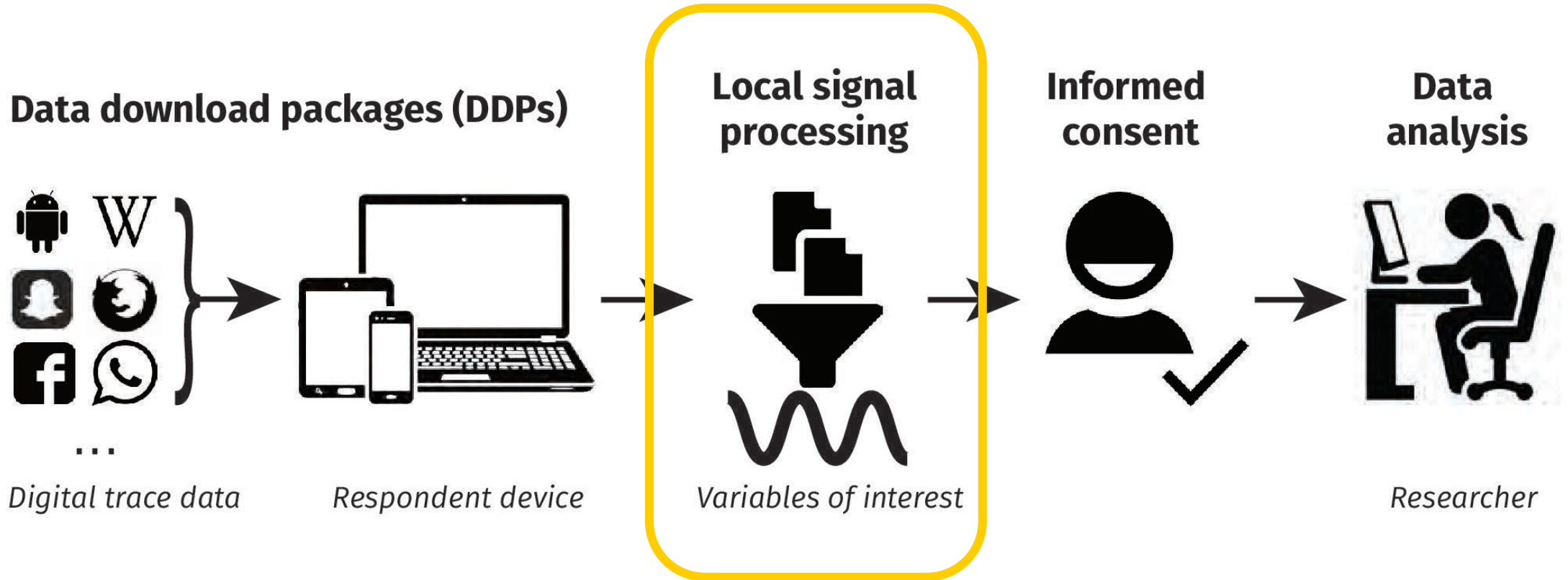
1. Background – Data Download Packages (DDPs)

- Deeply private
- Not willing to consent
- Not (all) needed for research question

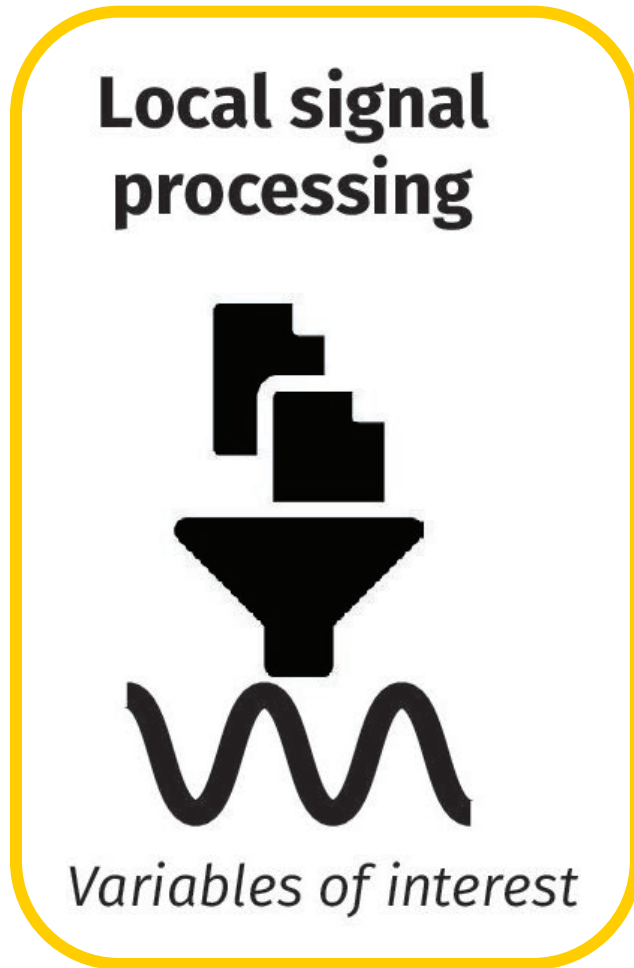
→ Alternative workflow

2. Workflow + Software

2. Workflow



2. Software



- Is used in webbrowser
- Creates a temporary system environment
- Runs a Python script that extracts info from DDP:
- **Specific for research question**
- **Specific for DDP**
- Extracted info is shown for consent
- After consent it is send to a server from the researcher
- System environment destroyed when browser is closed

3. Illustration

3. Illustration

Research Question:

Did the % of time you spent at home or at other locations change during a Covid-19 lockdown?

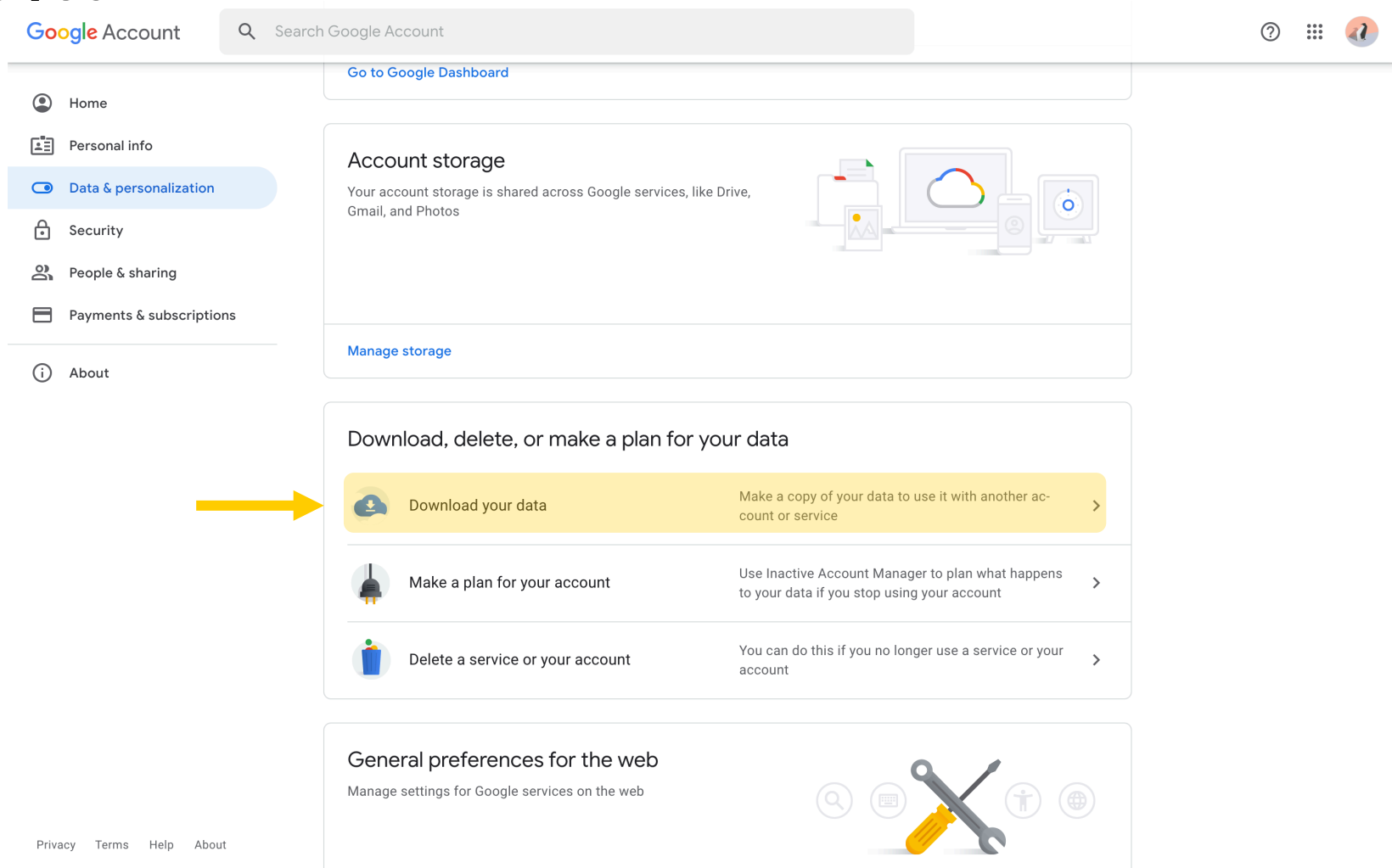
Which DDP to use?
Google Semantic
Location History

```
"location" : {  
  "latitudeE7" : 520844770,  
  "longitudeE7" : 51716420,  
  "placeId" : "ChIJ3ySEdJtoxkcRJo5l0lqqC1Y",  
  "address" : "Heidelberglaan 1\n3584 CS Utrecht\nNederland",  
  "name" : "Faculteit Sociale Wetenschappen",  
  "sourceInfo" : {  
    "deviceTag" : 1769097206  
  },  
  "locationConfidence" : 47.502987  
,  
  "duration" : {  
    "startTimestampMs" : "1551970749819",  
    "endTimestampMs" : "1551976816602"  
  },  
  "placeConfidence" : "MEDIUM_CONFIDENCE",  
  "centerLatE7" : 520845963,
```



3. Illustration

A. Request your DDP



3. Illustration


A. Request your DDP

Google Account



← Google Takeout

1 Select data to include


Multiple formats

 **Keep**
Notes and media attachments stored in Google Keep. [More info](#) ☐


Multiple formats

 **Location History**
Your Location History data collected while opted-in to Location History. ☒ 


Multiple formats

 **Mail**
Messages and attachments in your Gmail account in MBOX format. User settings from your Gmail account in JSON format. [More info](#) ☐

Multiple formats ☐ All Mail data included

 **Maps**
Your preferences and personal places in Maps ☐

Multiple formats ☐ All Maps data included

 **Maps (your places)**
Records of your starred places and place reviews. [More info](#) ☐

Multiple formats

3. Illustration

B. Download DDP to device

Google Account



← Google Takeout

Your account, your data.
Export a copy of content in your Google Account to back it up
or use it with a service outside of Google.

YOUR EXPORTS

Your latest export

Location History on July 1, 2021



✓ Download

+2 more

Manage exports

CREATE A NEW EXPORT

1 Select data to include

Products

Deselect all



Android Device Configuration Service

Android device attributes, performance data, software versions, and account
identifiers. [More info](#)



HTML format

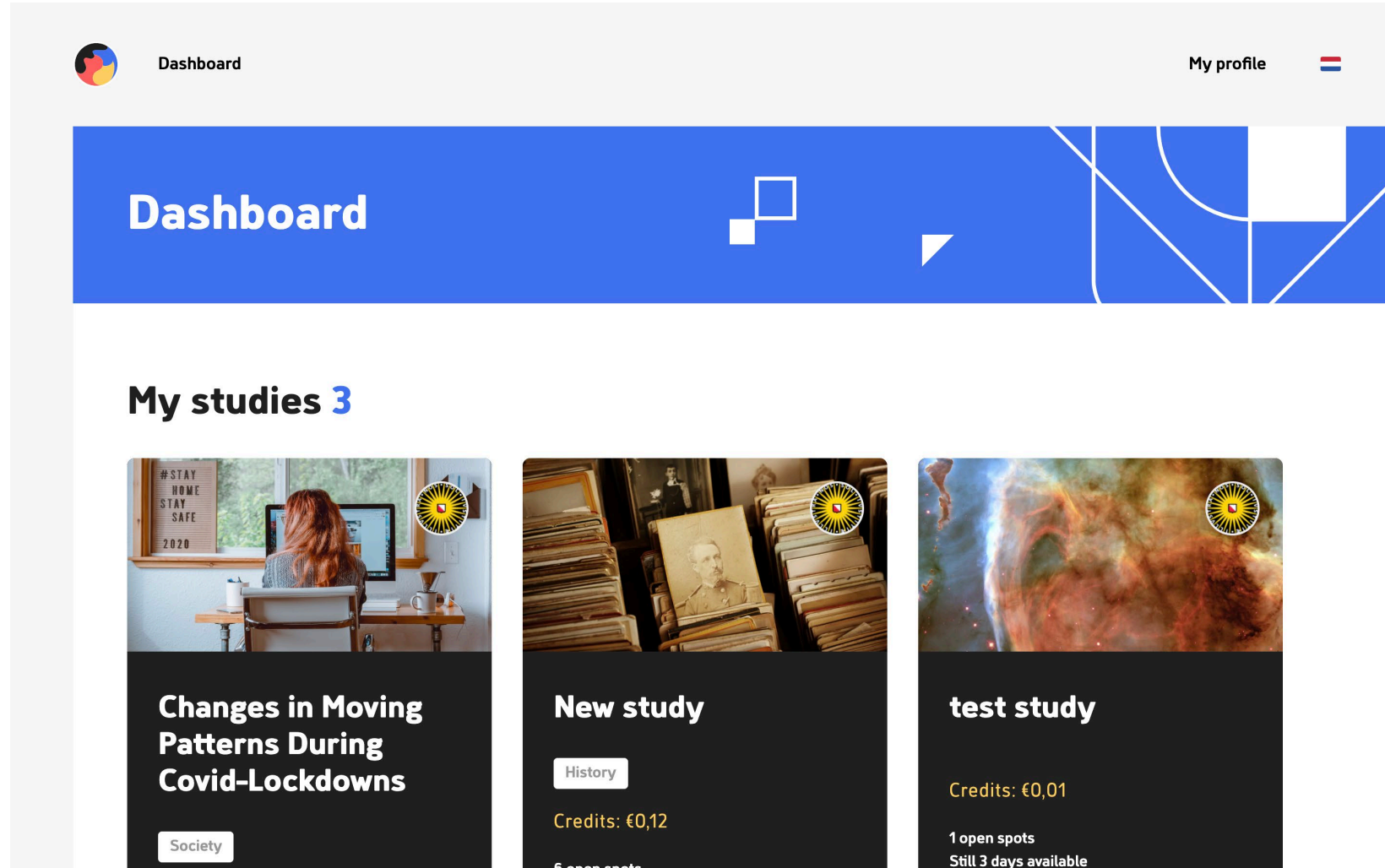


Arts & Culture

Favorites and galleries you've created on Google Arts & Culture

3. Illustration

C. Local processing → Go to PORT website



3. Illustration


C. Local processing → Read about the study

What can participants expect from the survey?

We will examine your Google semantic Location History data of January 2019, 2020 and 2021. To be precise, per month the total number of visited places are extracted, and the number of days spend at each place for the three most visited places. Furthermore, we will extract the number of days spend in places, the number of days spend travelling, and the travelled distance in km. We respect your privacy. Your data donation is anonymous.

Estimated duration

This study examines the change in travel behavior during the COVID-19 pandemic. The research question of this study is: Does the % of time people spent inside and outside their home changes when there is a covid-lockdown?



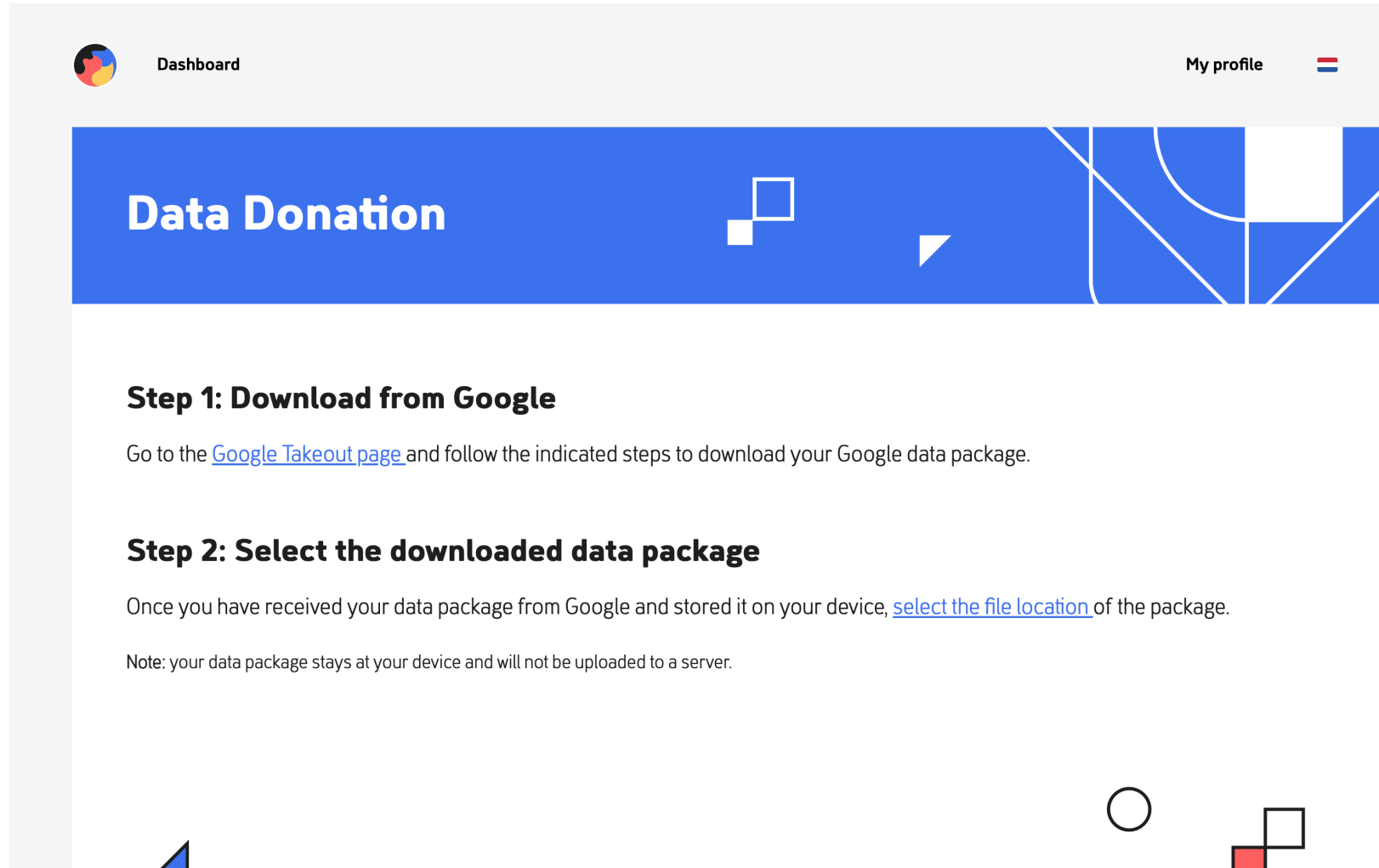
Researcher
Postdoc at UU
[Bezoek website](#)

[Donate your data](#)

[< Back to overview](#)

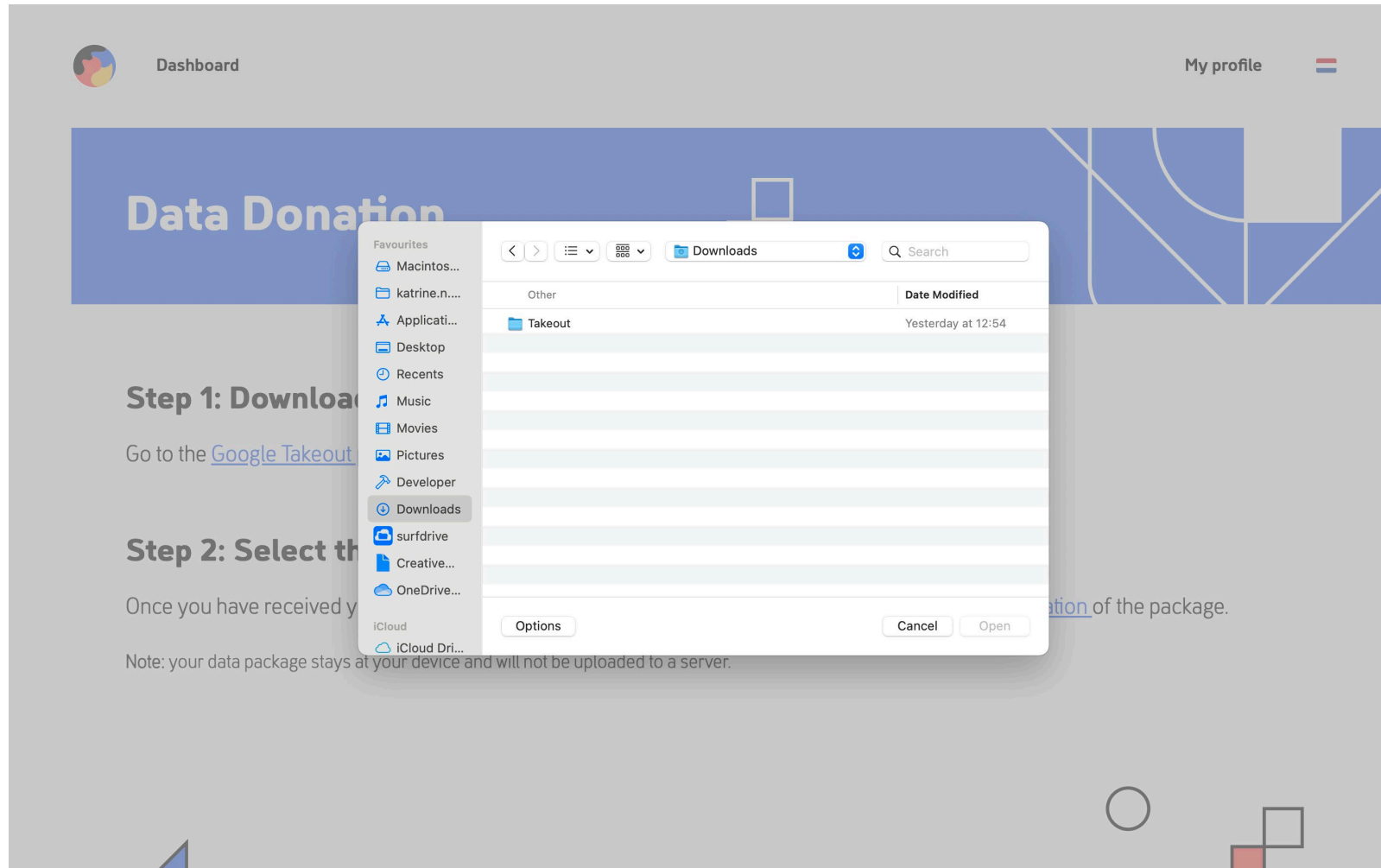
3. Illustration

C. Local processing → Donate your data!



3. Illustration

C. Local processing → Donate your data!



3. Illustration

C. Local processing → Donate your data!

Step 2: Select the downloaded data package

Once you have received your data package from Google and stored it on your device, [select the file location](#) of the package.

Note: your data package stays at your device and will not be uploaded to a server.

Step 3: Extract data

By clicking the button below, the data that is relevant for this research, will be extracted from your data package. During this process the data package will not leave your device and no data is stored on a server. The extracted data will be shown at step 4 for your consent. For your reference, the script that is used to extract the relevant data from your data package, is shown at the bottom of this page.

Process data package

Script

The script that is used to extract the relevant data from your data package

```
import json
import itertools
import re
import zipfile

import pandas as pd
```

3. Illustration

C. Local processing → Donate your data!

Step 4: Donate extracted data

The data that was extracted from your data package is shown below. Make sure, you check this data carefully before pressing the donate button below. If you have checked the extracted data and consent with donating this data for research, press the donate button.

This study examines the change in travel behaviour during the COVID-19 pandemic. We therefore examined your Google semantic Location History data for January in 2019, 2020, and 2021. To be precise, we extracted per month the total number of visited places, and the number of days spend per place for the three most visited places. Also, we extracted the number of days spend in places and travelling, and the travelled distance in km.

Year	Month	Number of Places	Places	Duration [days]	Activity	Duration [days]	Activity	Distance [km]	Place 1 (days)	Place 2 (days)	Place 3 (days)
0	2019 JANUARY	48	24.8	01	6.20			1492.873	9.722	7.986	0.843
1	2020 JANUARY	48	24.8	03	6.20			1569.261	10.664	6.597	1.290
2	2021 JANUARY	18	29.4	49	1.55			340.939	22.618	1.178	0.707

By pressing the donate button you agree to the following [terms and conditions](#).

Donate extracted data

4. Next steps

4. Next steps

Near future:

- *Integration with participant recruitment platform*
- *Integration with environment for safe data storage*

Far away future:

- *Integration with surveys*
- *Experiments for evaluation of measurement quality*
- *Expanding the Python extraction scripts*

Want to test PORT? Email:

l.boeschoten@uu.nl

d.l.oberiski@uu.nl

Manuscript workflow

(including TE framework):

<https://arxiv.org/pdf/2011.09851.pdf>

Github code PORT:

<https://github.com/eyra/port-poc>





Although every effort has been made to ensure that all information in this presentation is correct and up to date, Utrecht University cannot be held liable for any false, inaccurate or incomplete information presented herein.