

User-centric approaches

Lecture 2

Laura Boeschoten

Assistant professor

Department of Methodology and Statistics

Literature

- Chapter 1: “Introduction” in Data collection with Wearables, Apps and Sensors (2023), Florian Keusch, Bella Struminskaya, Stephanie Eckman & Heidi Guyer
Sections 1.1 - 1.4
- Ohme, J. Araujo, T. Boeschoten, L., Freelon. D. Ram, N., Reeves, B.B. & Robinson, T.N. (2024) Digital Trace Data Collection for Social Media Effects Research: Apis, Data Donation and (Screen) Tracking, Communication Methods and Measures, 18(2) 124-141.



Goals of this lecture

1. Distinguish between platform- and user-centric approach to digital trace data collection.
2. Understand how commonly used user-centric approaches work and how they can be used for research.
3. Understand the advantages, challenges and ethical considerations of user-centric approaches.
4. Provide examples of studies and explain how research questions can be answered using a user-centric approach.

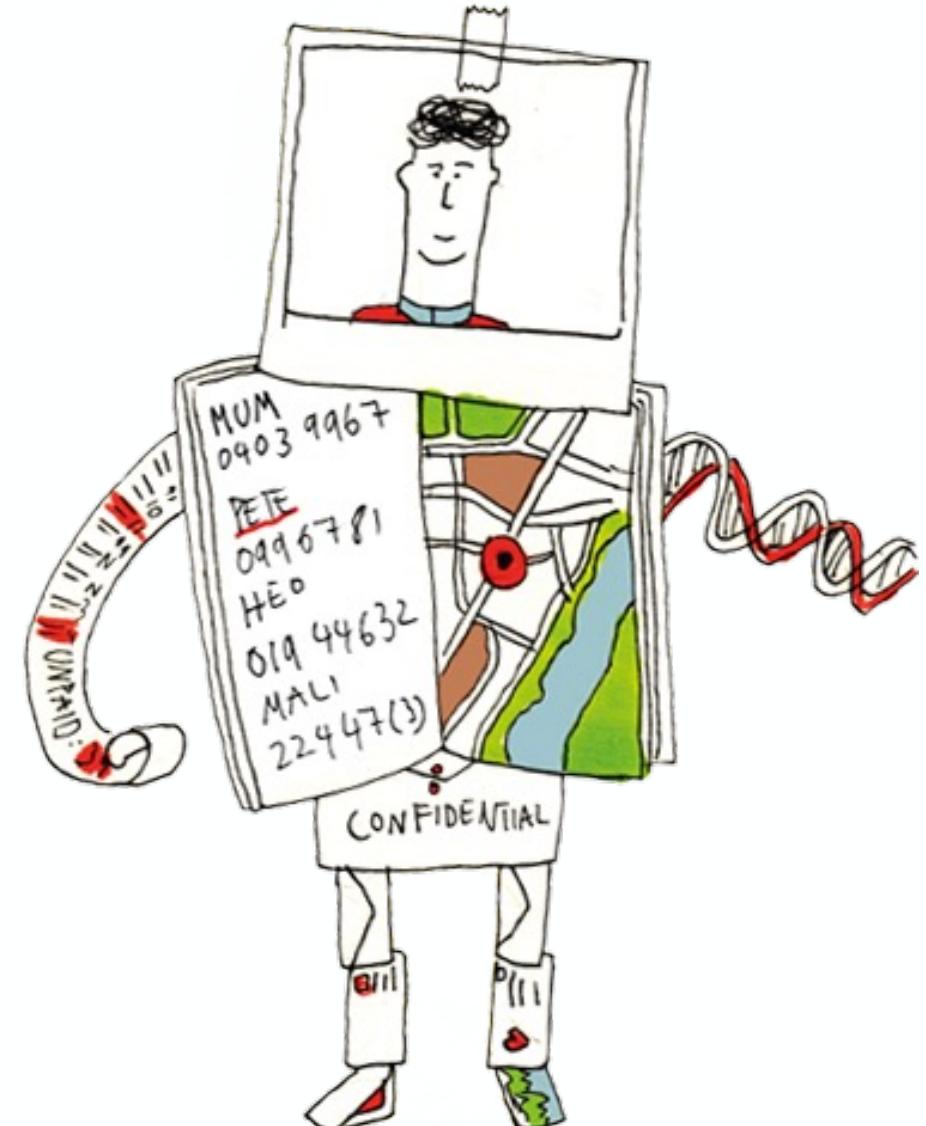
1. Digital Trace Data Collection:

User-centric and platform-centric
approaches

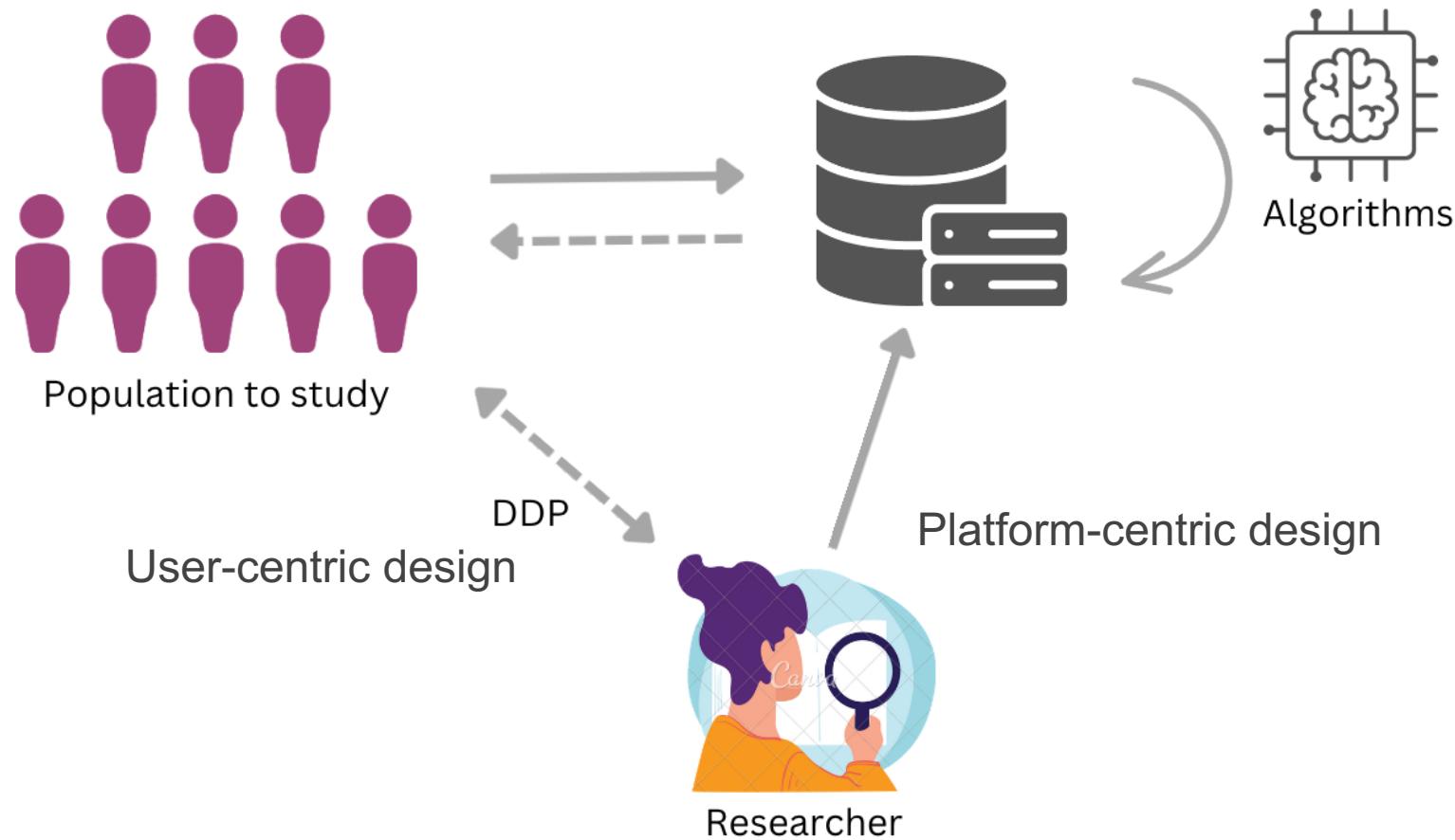
Digital traces

Digital footprint; Digital shadow

One's unique set of digital activities, actions, and communications that leave a data trace on the internet or on a computer or other digital device and can identify the particular user or device.



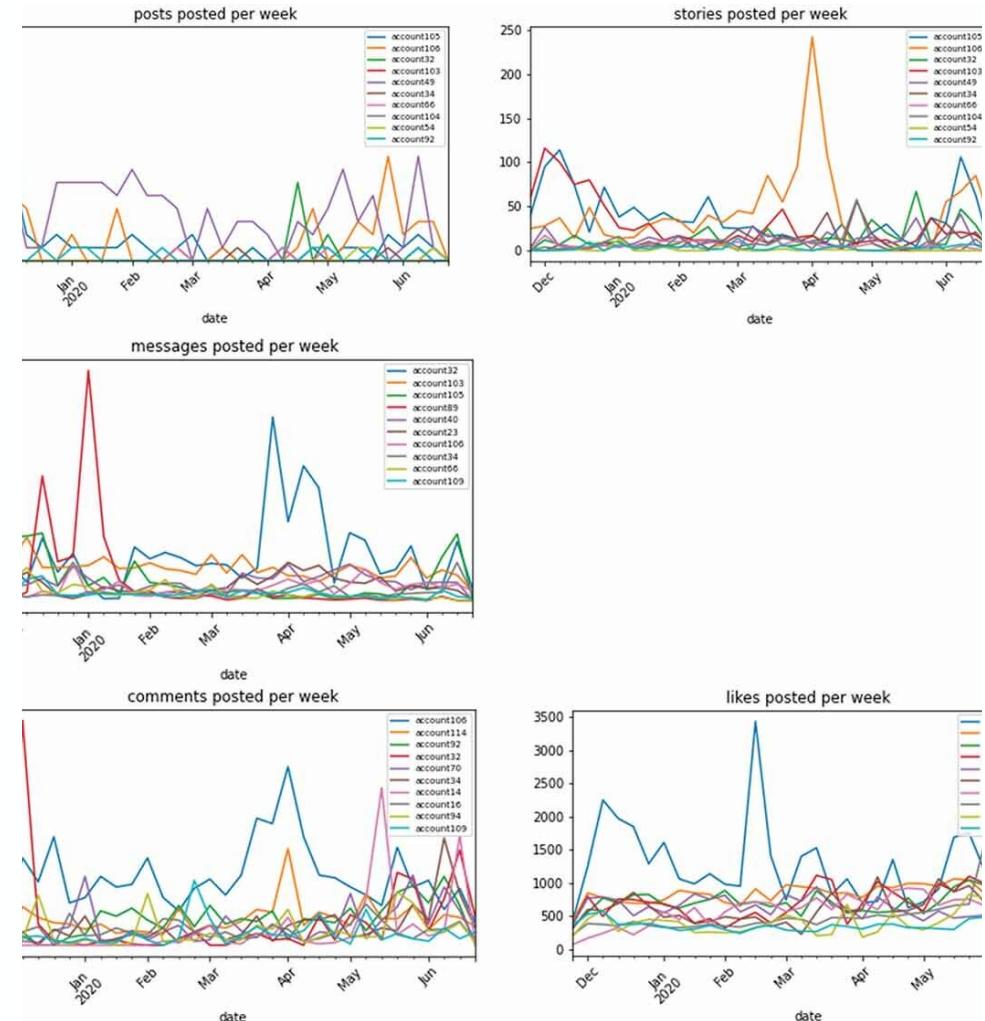
User-centric versus platform-centric



Why take a user-centric approach?

Studying person specific effects

- A **user** is any individual who generates a digital trace on a platform by making use of its functionalities.
- Selection mechanisms and potential biases shape the social contexts on which interactions on **platforms** take place.
- Platform-centric research methods allow for the study of **aggregated** or **public** social media effects, while user-centric methods allow for studying **person-specific** effects.



Fluctuations in posting, updating stories, direct messaging, commenting, and liking over time from the top 10 accounts engaging in each activity.

Supplement digital traces with more info

Measurement Burst Design (Nesselroade, 1991)

N = 388 (age 13-15)

T > 250



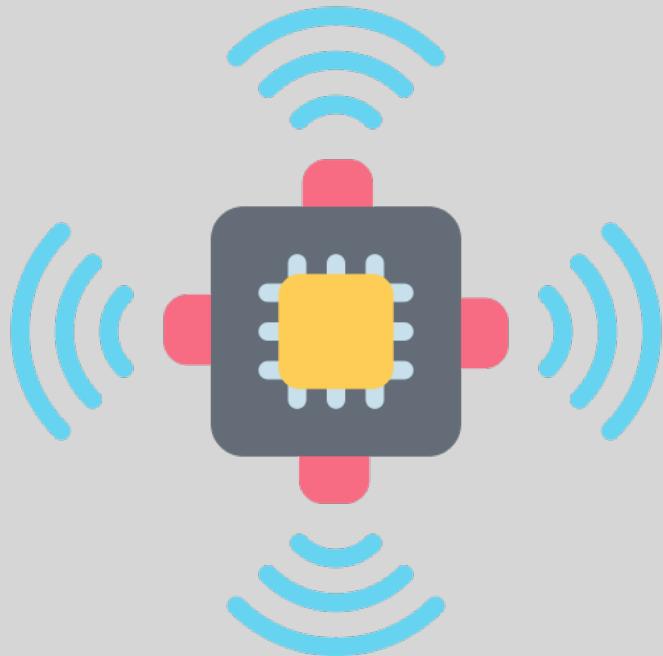
2. What user-centric approaches are out there?

**A. Sensor
approaches**

**B. Capture /
scraping
approaches**

**C. Data donation
approaches**

A. Sensor approaches



What sensors can you use?



What sensors can you use?

NFC

Thermometer

Wi-Fi

GPS

Cellular network

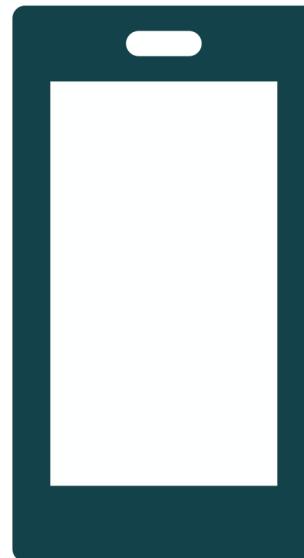
Fingerprint sensor

Barometer

Accelerometer

Gyroscope

Bluetooth



Air humidity sensor

Proximity sensor

Microphone

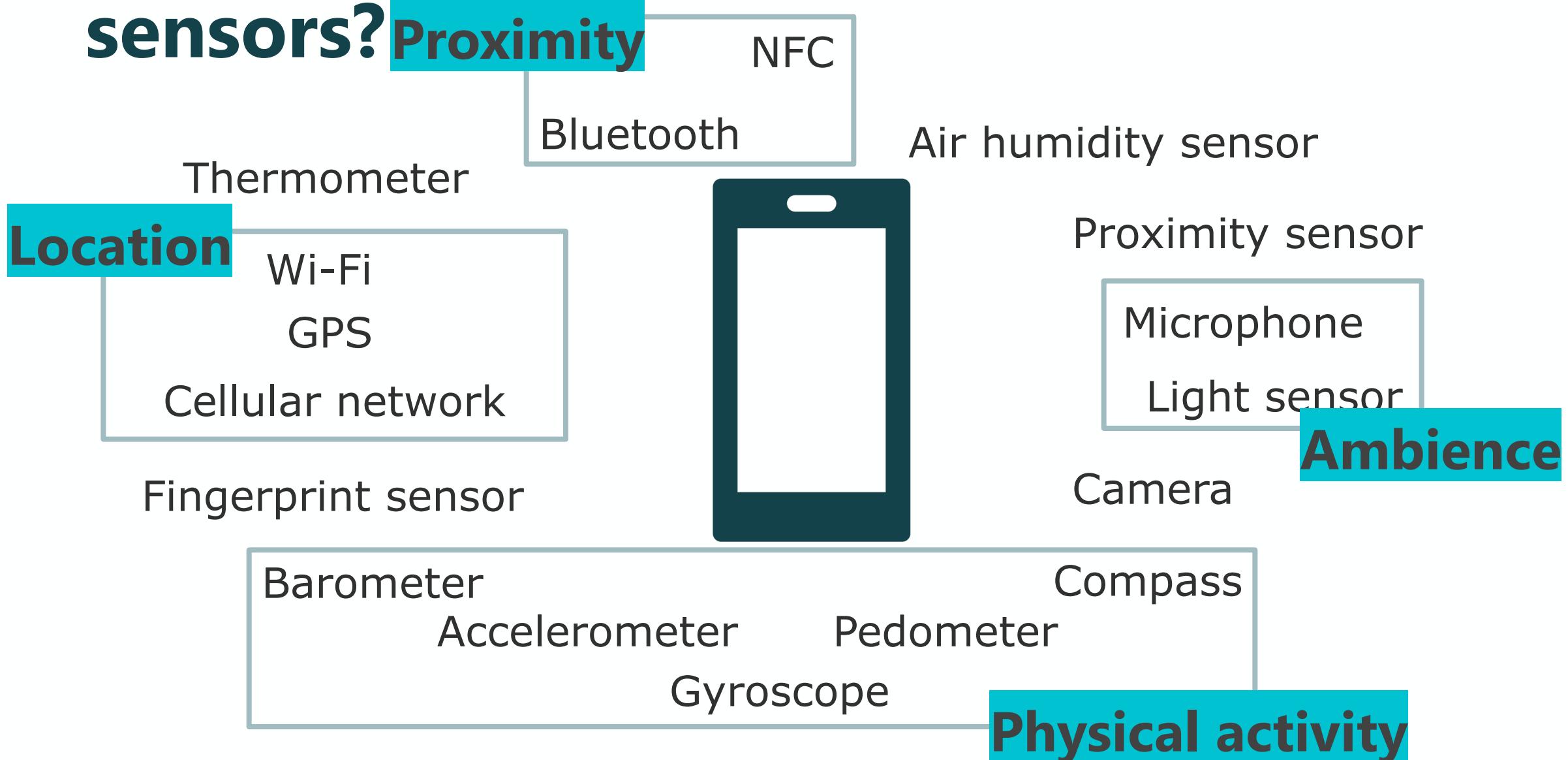
Light sensor

Camera

Compass

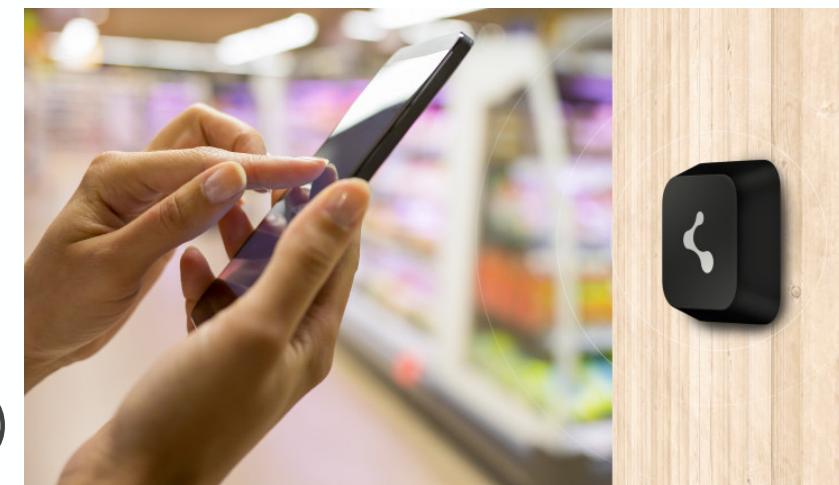
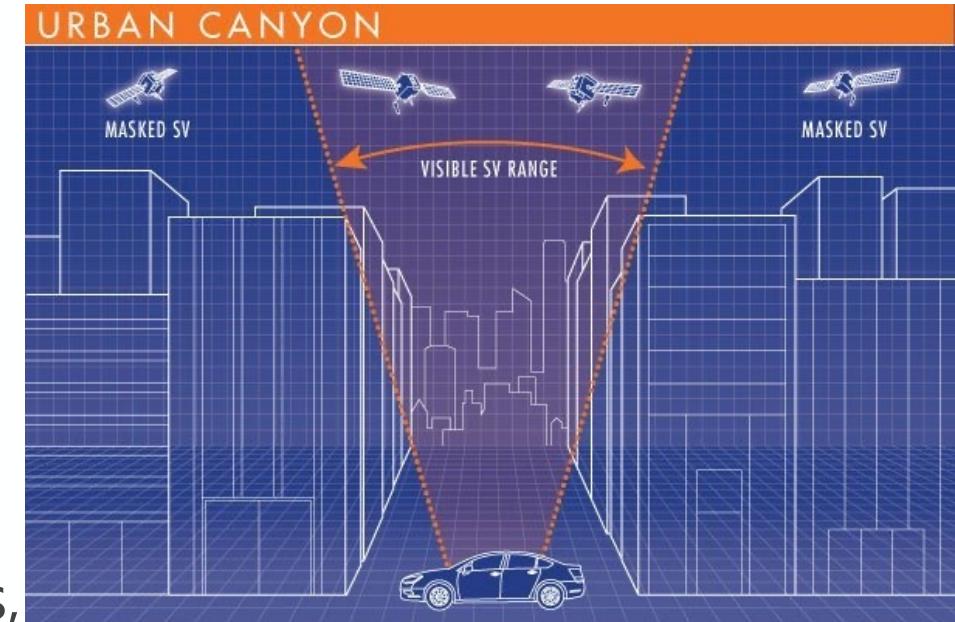
Pedometer

What can you measure with these sensors?



Location sensors

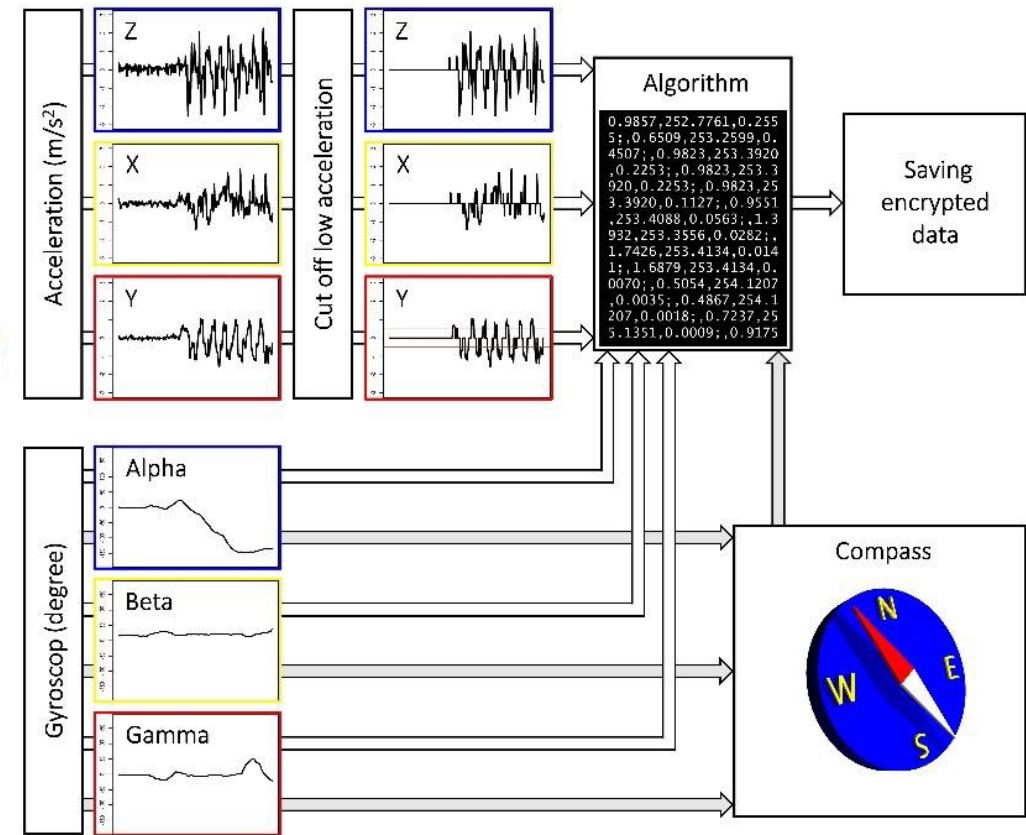
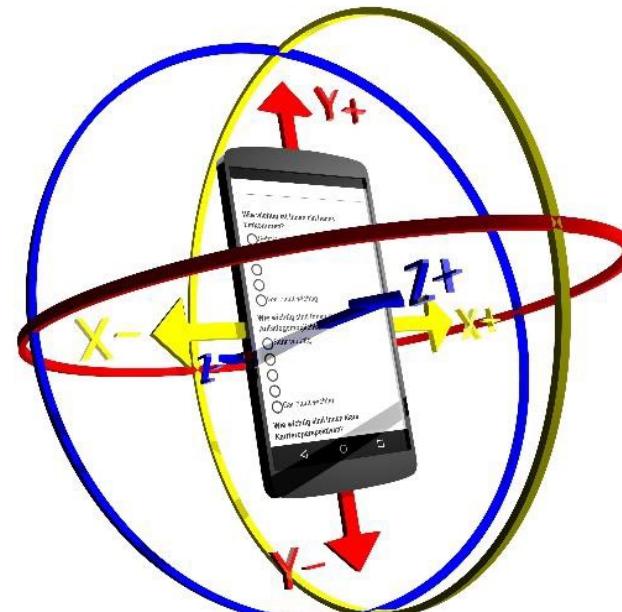
- GPS
 - Coordinates in longitude & latitude
 - High accuracy (newest generation 30 cm)
 - Works without cell/Internet connection
 - Performs worse in 'urban canyons', indoors, and underground (pseudo accuracy!)
 - Can be battery-draining
- Cellular network
 - Multilateration of radio signals between (several) cell towers
- WiFi
 - Inferring location from Wi-Fi access points (AP)
- Beacons
 - Bluetooth transmitters for indoors



Images: <https://i2.wp.com/geoawesomeness.com/wp-content/uploads/2014/01/urbancanyon.jpg?fit=600%2C450&ssl=1>,
<https://locatify.com/wp-content/uploads/2015/03/beacon-wall-756x425.jpg>

Physical activity sensors

- Accelerometer
- Gyroscope
- Barometer
- Pedometer



Ambience sensors, proximity sensors

- Camera
 - photos, videos, scanning of bar codes
 - linear distance
- Microphone
 - active and passive (ambient noise) recording
- Light sensor
 - e.g., identify idle state
- Bluetooth
- RFID (radio frequency)
- NFC



How does it work exactly?

- As a researcher, you utilize the existing sensors on participants' devices.
- Three options:
 1. You ask participants to use the sensor directly.
 2. You develop a piece of software (such as an app or plug-in) that captures the traces you are interested in.
 3. You make use of an existing app or plug-in.
- You invite participants.
- Data collection is prospective; it collects the traces as they are being produced.

Some definitions

Wearables, apps and sensors

- **Wearable:** A lightweight device including one or multiple sensors that people can wear close to or on their skin, e.g., a fitness bracelet or smartwatch.
- **App:** A software program that runs on a smartphone or other smart device. Apps allow users to interact with a device. For researchers, apps provide the opportunity to collect, aggregate, and process data collected on the device.
- **Sensors:** A piece of hardware that converts a physical phenomenon into an electrical signal. For example, the accelerometer sensor measures the rate of change of velocity of an object.

Passive and active data collection

- **Passive data collection:** Data are collected, for example, by sensors, as a byproduct of everyday activity (e.g., accelerometer collecting data about walking).
- **Active data collection:** A participant deliberately creates data as a reaction to a stimulus designed by the researcher (e.g., responding to a survey question, taking a picture and sharing it with the researcher).

Designed and found data

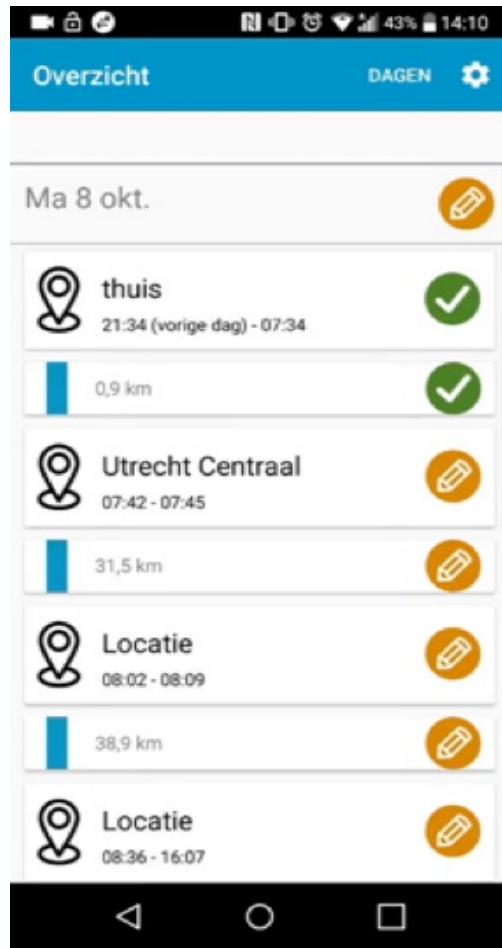
From the perspective of researchers, big data sources are "**found**". However, they are "**designed**" by someone.

Example 1

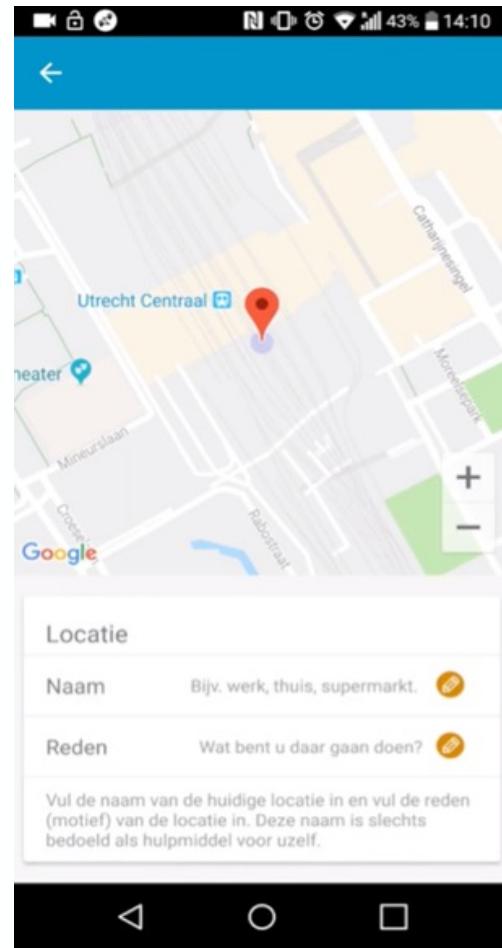
Measuring everyday mobility
using an app

How do people move in everyday life?

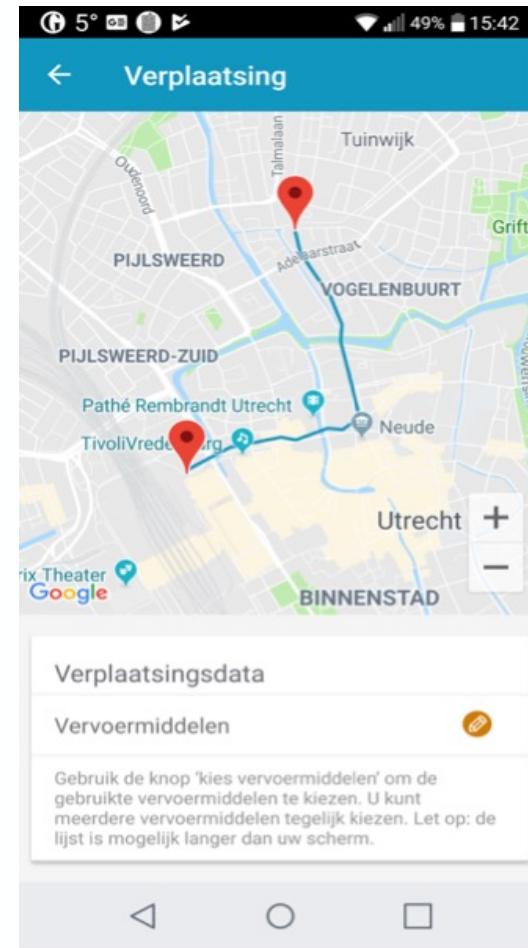
- Field test in the Dutch general population (Nov-Dec 2018)
 - Travel app of Statistics Netherlands ("TABI APP" Android & iPhone)
 - Data collection for 7 days
 - N = 1,902
- Sensing location per second (when moving) & per minute (when still):
 - GPS
 - Wi-Fi
- Respondents eagerly provide additional information that helps understand travel behavior (label stops and motives for travel).



Daily
overview



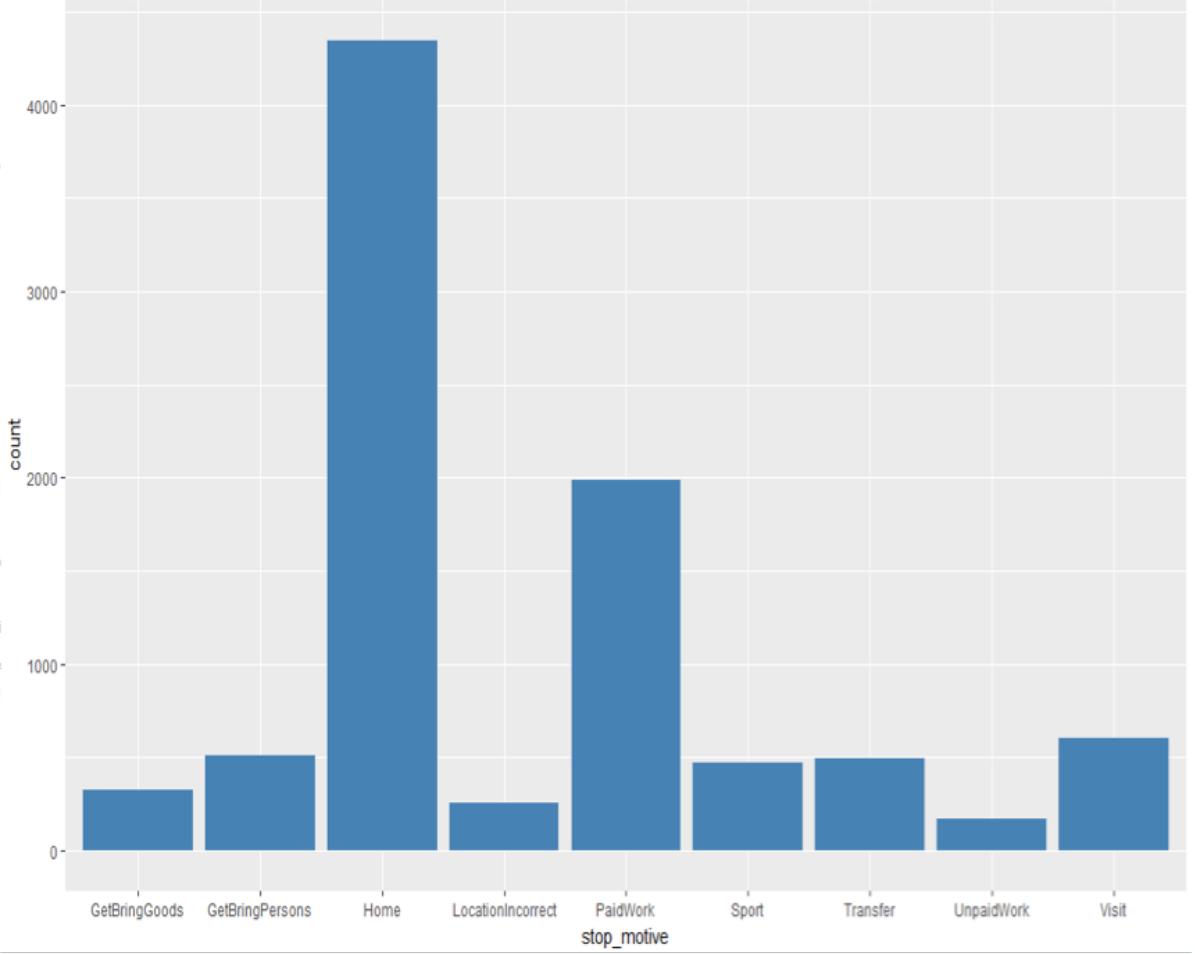
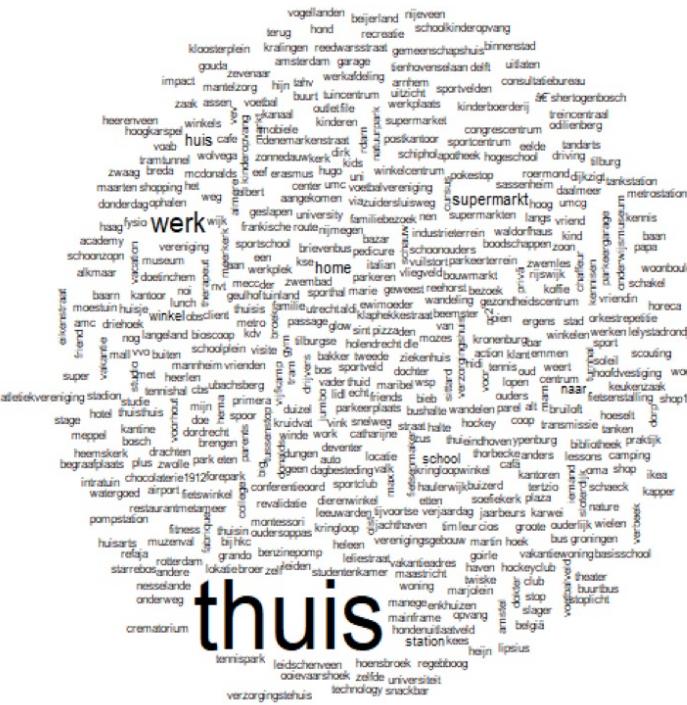
Questions about
stops



Questions about
trips

Motives and means of transportation

- 22,000 stops
 - 13,000 (60%) labeled
 - Overall 50% of respondents give complete/almost complete details



Example 2

Mobility and employment
using an app

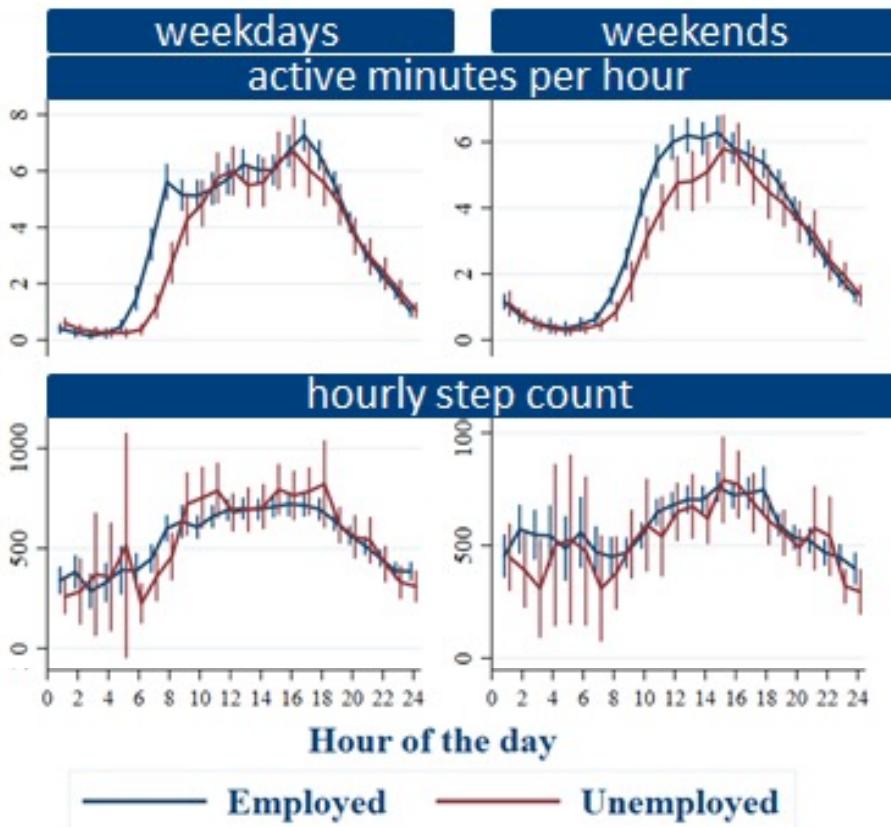
What are the effects of unemployment?

(Kreuter et al. 2018)

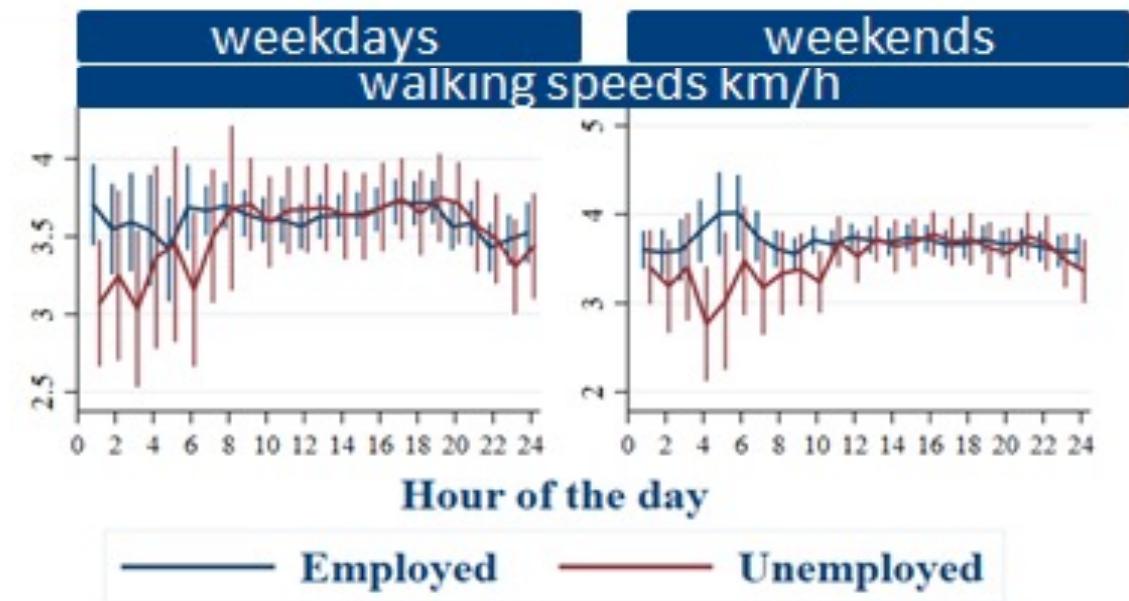
- 650 Android owners from German panel study "Labour Market and Social Security" (PASS) downloaded *IAB-SMART* app for 6 months
- EMA questions concerning both subjective (e.g., affective impact of daily smartphone use, Big 5 personality) and objective phenomena (e.g., employment and job search activities, use of smartphones in everyday life, memberships in professional and voluntary organizations)
- Five sensing modules:
 - Location using GPS, Wi-Fi, and cellular sensors every 30 minutes
 - Activity and means of transportation (e.g., walking, biking, riding in/on a motorized vehicle) using accelerometer and pedometer
 - Call and texting behavior using phone and SMS logs
 - Use of apps installed on smartphone
 - Social network characteristics from contact lists.

What are the effects of unemployment?

(Kreuter et al. 2018)



Predictive Margins with 95% confidence intervals.
Controls: Gender, age, hours smartphone is kept nearby.



Predictive Margins with 95% confidence intervals.
Controls: Gender, age, hours smartphone is kept nearby.

Exercise:

Of the two examples discussed:

- Was this data collection active or passive?
- Who designed the data and for what purpose?



Exercise

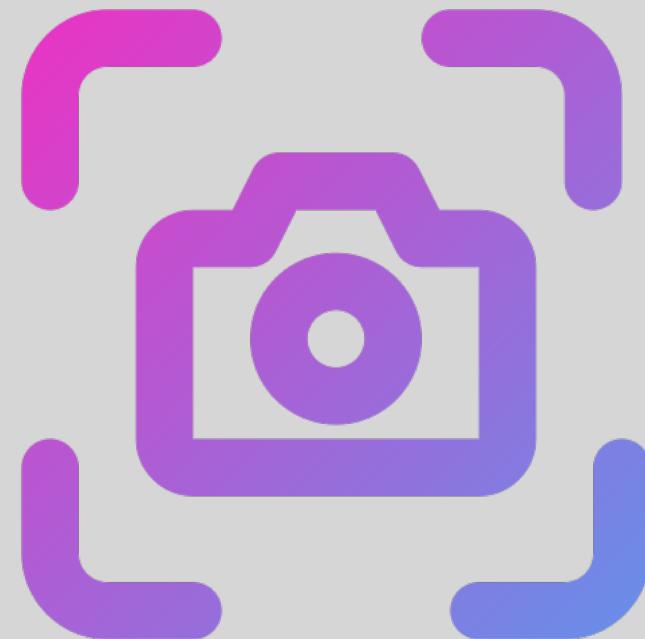
Example 1: Measuring everyday mobility using an app

- Passive / active?
- Who designed the data?

Example 2: Mobility and employment using an app

- Passive / active?
- Who designed the data?

B. Capture or scraping approaches



How does it work exactly?

- Like sensor approaches:
 1. You invite participants.
 2. You ask them to make use of an app or plug-in.
 3. Data collection is prospective.
- Main difference: Instead of phone sensors, you make use of scraping or screen capture tools.

Example 1

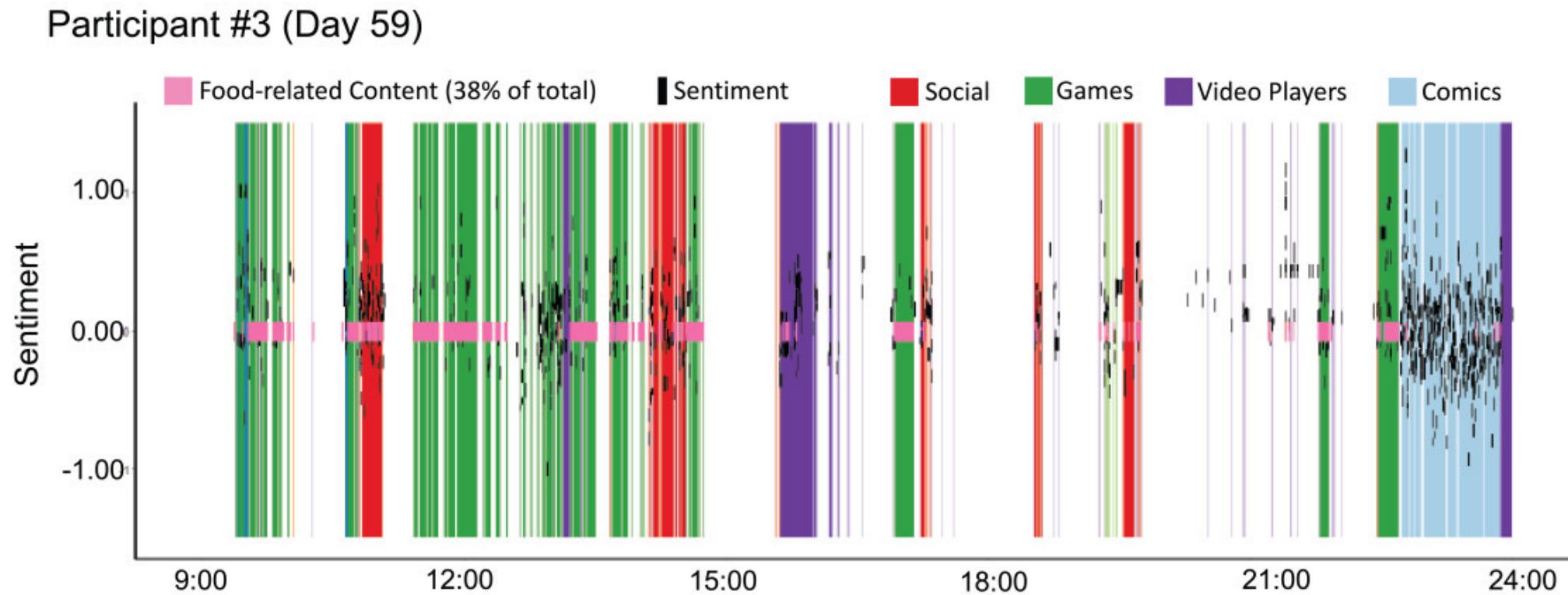
The human screenome project

The human screenome project

- The **Screenomics** software records, encrypts and transmits screenshots automatically and unobtrusively **every 5 seconds**, whenever a device is turned on.
- When it is deployed on **multiple devices** at once, the screenshots from each one are synced in time.

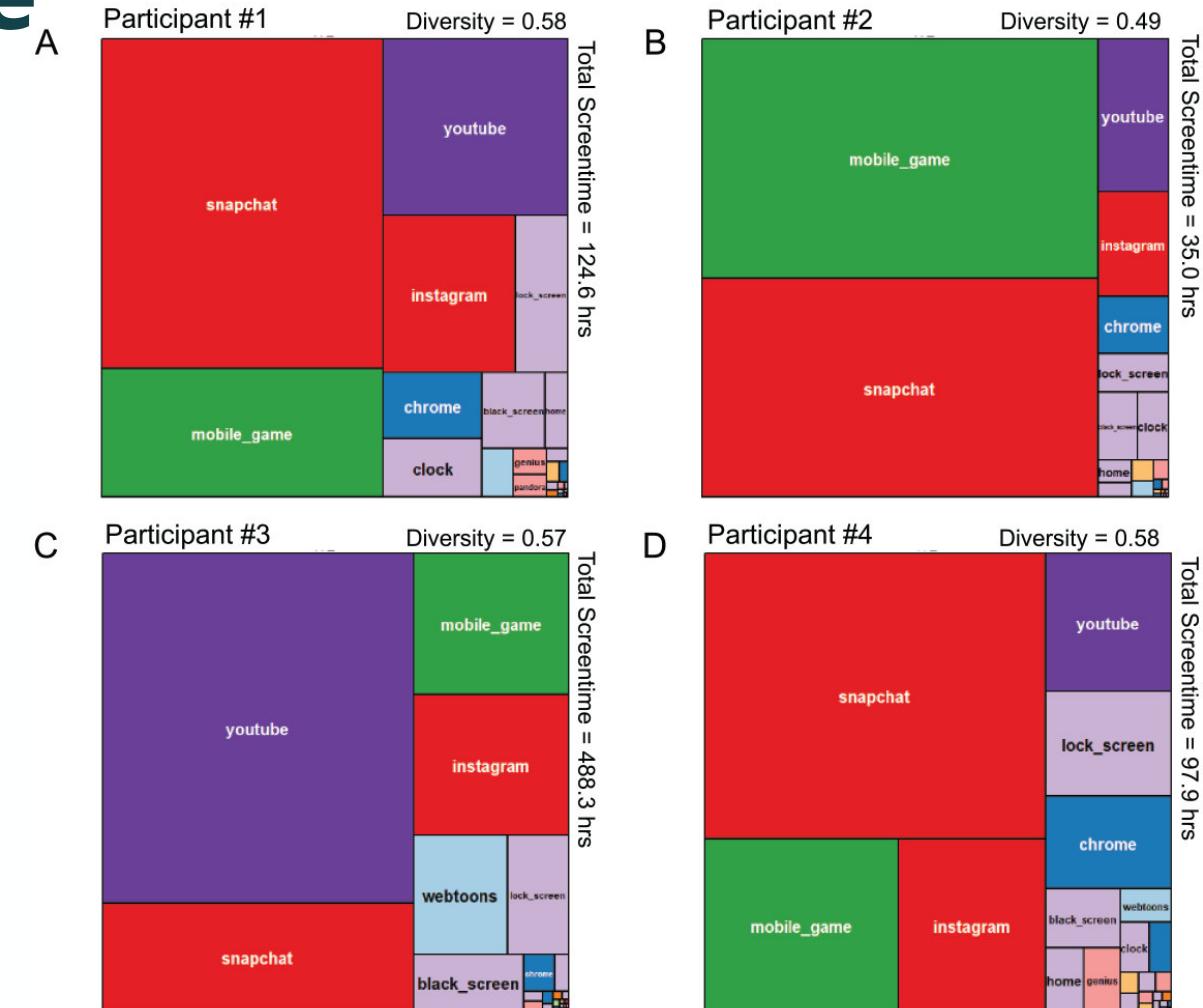


Example data from the Screenomics app



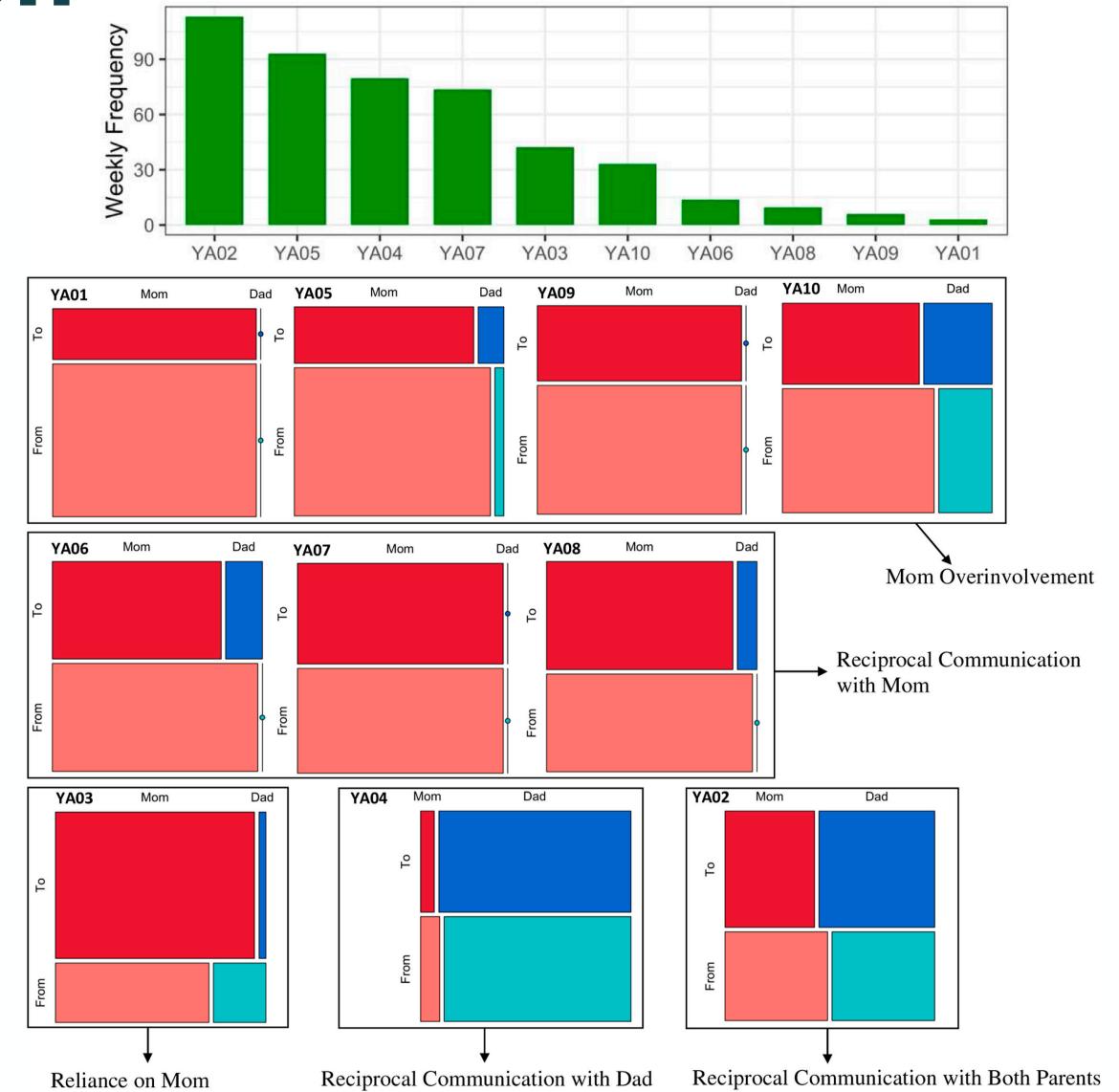
Example data from the Screenomics app

- Activities on smartphones of four different participants
- Blocks represent different apps/activities
- Notice the difference in screentime!



Screnconomics application

- Connectedness and independence of young adults and parents in the digital world:

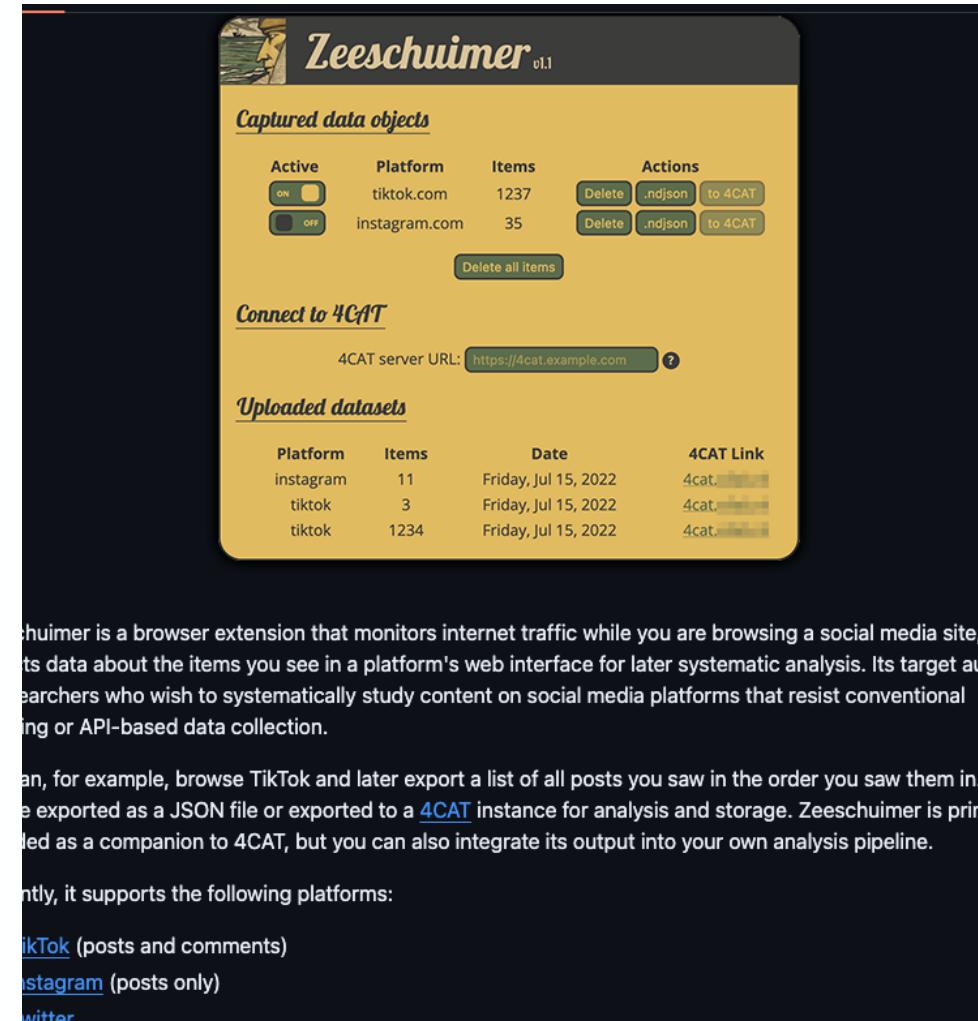


Example 2

Capture social media website visits
using a plug-in

Browser plug-in Zeeschuimer

- You install a browser plugin developed for a specific platform (for example TikTok).
- When you use the platform of interest (like TikTok) it captures the video's that you see and stores these in a list.
- You share that file with the researcher.



The screenshot shows the Zeeschuimer extension interface. At the top, there's a header with the logo and version (v1.1). Below it, a section titled "Captured data objects" displays two rows of data. The first row has "Active" status set to "ON" for both platforms, "Platform" as "tiktok.com" and "Items" as 1237, with "Actions" buttons for "Delete" and ".ndjson to 4CAT". The second row has "Active" status set to "OFF" for both platforms, "Platform" as "instagram.com" and "Items" as 35, with similar "Actions" buttons. A "Delete all items" button is also present. Below this is a "Connect to 4CAT" section with a "4CAT server URL" input field containing "https://4cat.example.com" and a help icon. The final section, "Uploaded datasets", lists three entries: "Platform" (Instagram, TikTok, TikTok), "Items" (11, 3, 1234), "Date" (Friday, Jul 15, 2022), and "4CAT Link" (links to 4cat.example.com). At the bottom, there's descriptive text about the extension's purpose and supported platforms.

Zeeschuimer is a browser extension that monitors internet traffic while you are browsing a social media site, capturing data about the items you see in a platform's web interface for later systematic analysis. Its target audience is researchers who wish to systematically study content on social media platforms that resist conventional crawling or API-based data collection.

You can, for example, browse TikTok and later export a list of all posts you saw in the order you saw them in. This list can be exported as a JSON file or exported to a [4CAT](#) instance for analysis and storage. Zeeschuimer is primarily designed as a companion to 4CAT, but you can also integrate its output into your own analysis pipeline.

Currently, it supports the following platforms:

- [TikTok](#) (posts and comments)
- [Instagram](#) (posts only)
- [Twitter](#)

Application: Assess TikTok videos from news outlets

News #foryou on TikTok: A Digital Methods-Based Study

I-27
© 2025 AEJMC
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10776990251328623
<http://journals.sagepub.com/home/jmq>



Jonathan Hendrickx¹ 

Abstract

TikTok is rapidly establishing itself as an important platform for contemporary digital journalism but explorations on its transnational journalistic usage thus far remain limited in size and scope. Hence, this explorative study adopts a digital methods approach to collect and assess 26,473 TikTok videos posted by 91 European news outlets between 2019 and 2022. Rooted conceptually in affordance and hybridity theory and methodologically in digital methods, the study theorizes digital production trends by drawing on a proposed typology of visual, hashtags, and auditory affordances. News outlets studied adhere to visual and hashtag affordances, but much less so to auditory ones.

Exercise:

Of the two examples discussed:

- Was this data collection active or passive?
- Who designed the data and for what purpose?



Exercise

Example 1: Connectedness of young adults and parents through Screenomics

- Passive / active?
- Who designed the data?

Example 2: Migration of academic twitter using Zeeschuimer

- Passive / active?
- Who designed the data?

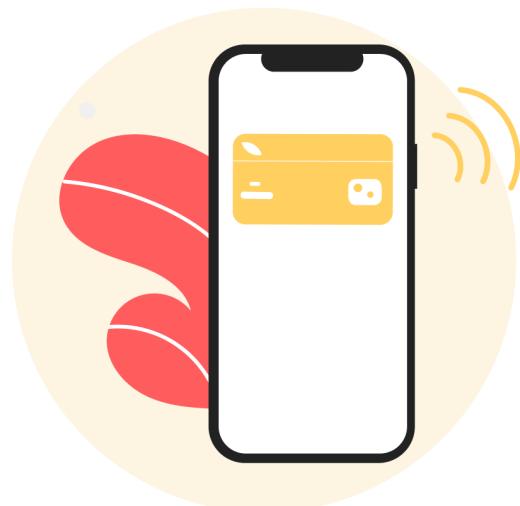
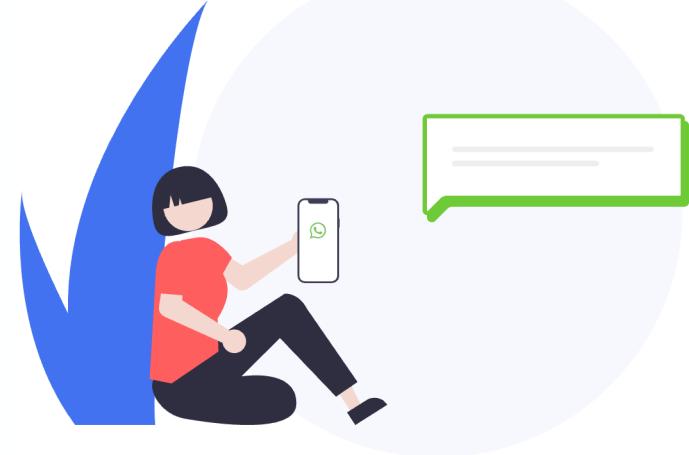
Coffee Break



C. Data donation approaches



With almost everything we do, we leave digital traces behind



GDPR

Article 15 – Right of data access



Right to obtain from a data controller:

- Confirmation whether personal data are being processed
- Access to the personal data
- Access to information regarding data recipients and sources and data derived from your personal data

GDPR

Article 20 – Right of data portability

Grants data subjects the right to

- *receive the personal data in a structured, commonly used and machine-readable format ("Data Download Package")*
- *transmit those data to another data controller*

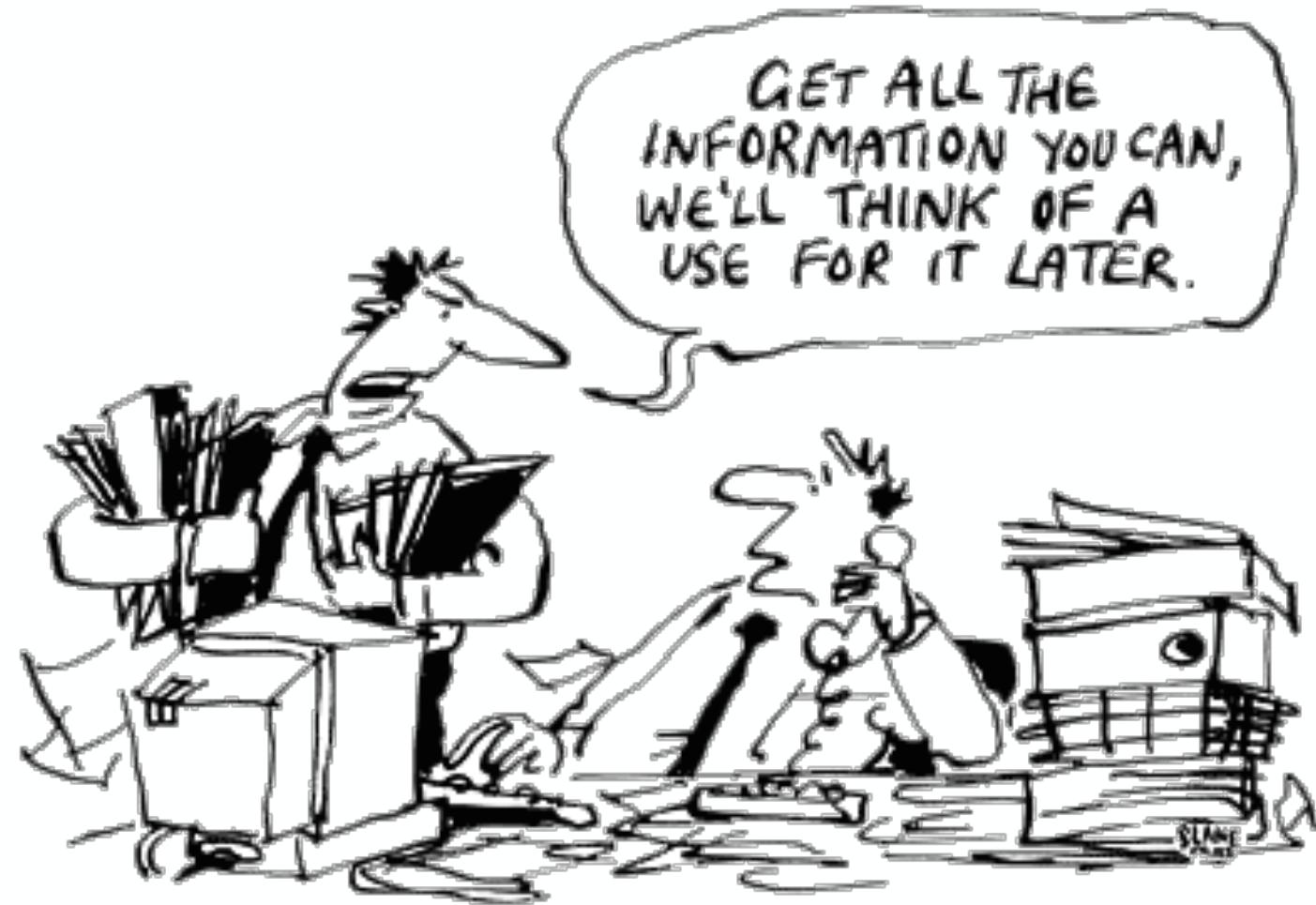


Data donation

Donation of digital traces collected as Data

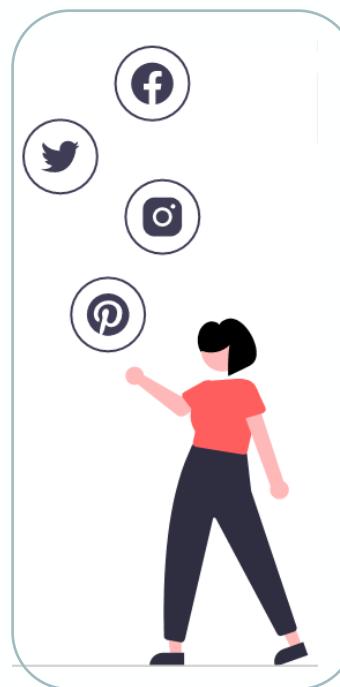
Download Packages (DDPs) for research purposes.

See: <https://datadonation.eu/>

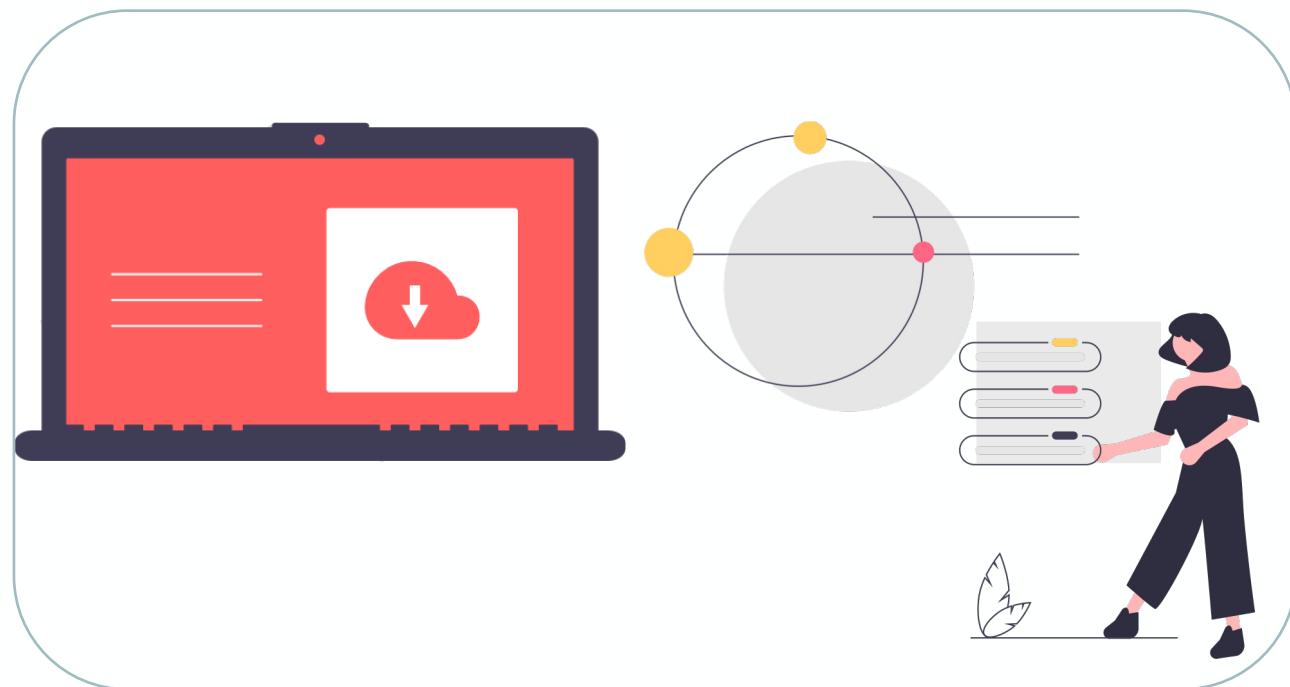


Data donation workflow

Online platform



Device of participant



Researcher environment



Two approaches to acquire the donations

1. Donation of complete DDPs.
 - Often, de-identification protocols are in place as soon as the data comes in.
 - Particularly adopted in early data donation studies.
2. Donation of digital traces after local processing.
 - Currently most used approach.

GDPR Definitions

- **Personal data:** Information relating to an identified or identifiable natural person.
- **Data subject:** The natural person that the personal data refer to.
- **Data controller:** The person or organization responsible for processing personal data. The controller decides which data will be processed, how and why.

Data donation definitions

- **Data Download Package (DDP):** Because of the right of data access, data subjects are always allowed to retrieve their personal data from data controllers. Data controllers are obliged to comply with such requests and because of the right of data portability, provide the requested data in a machine-readable format. To comply with these regulations, large data controllers such as social media platforms typically provide data subjects with a .zip file containing the personal data requested.
- **Local processing:** Locally at the device of a research participant, data is extracted from a DDP that is relevant for a particular research question under investigation.

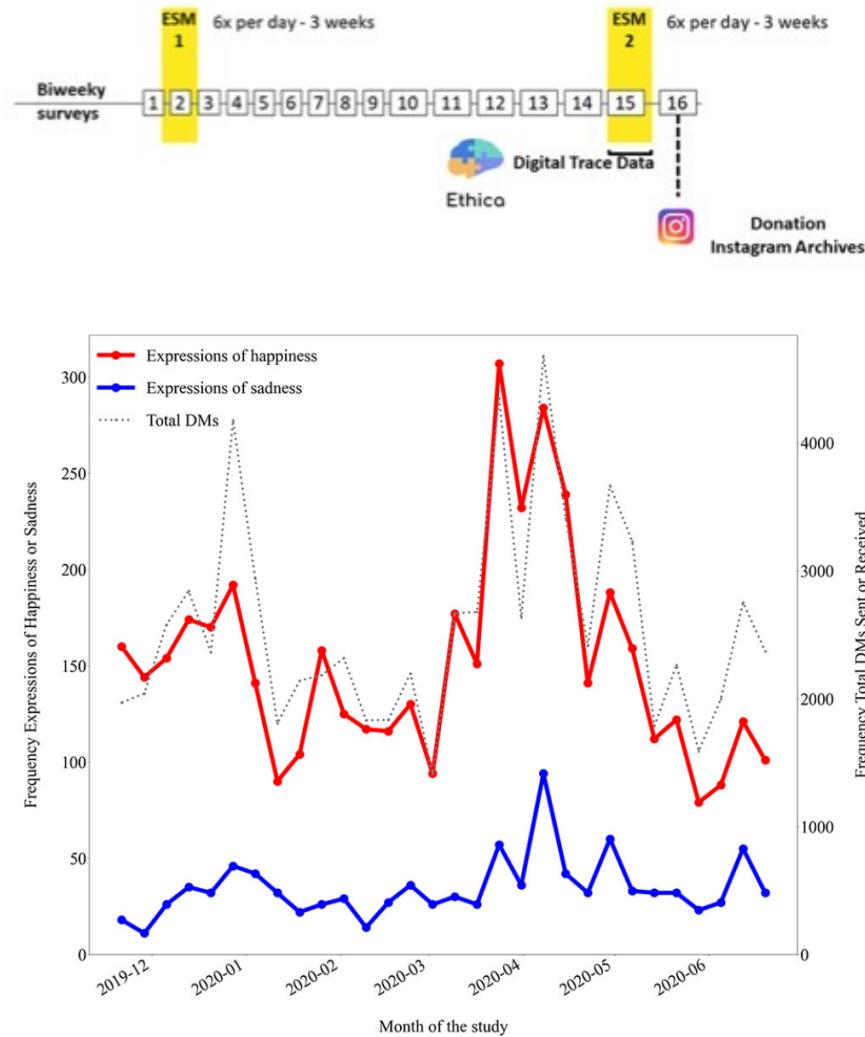
Example 1

Happiness and sadness in adolescent's
Instagram direct messages

Measurement Burst Design (Nesselroade, 1991)

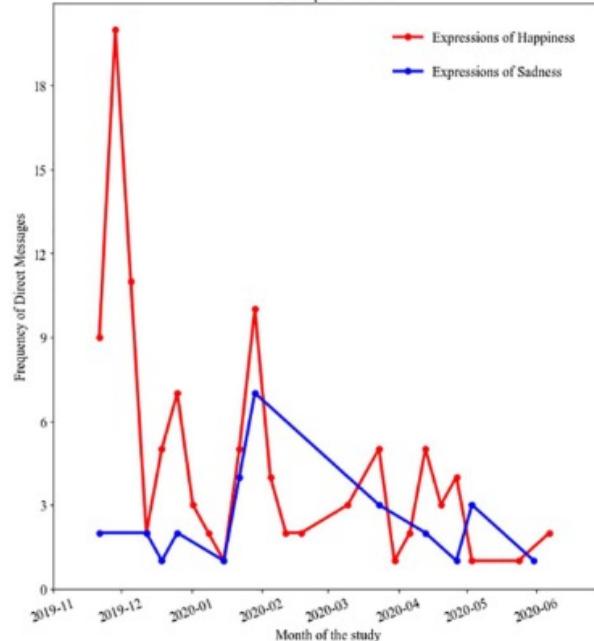
N = 388 (age 13-15)

T > 250



Temporal Trends in Expressions of Happiness/Sadness for Two Adolescents

Participant ID 22



Participant ID 30

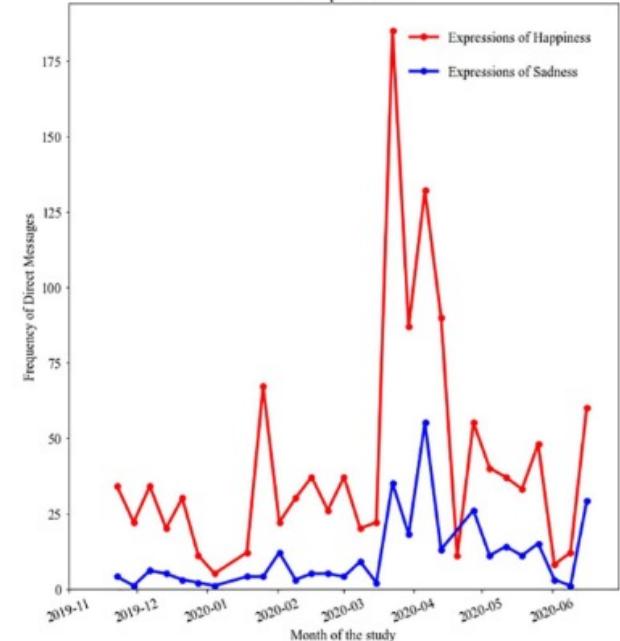
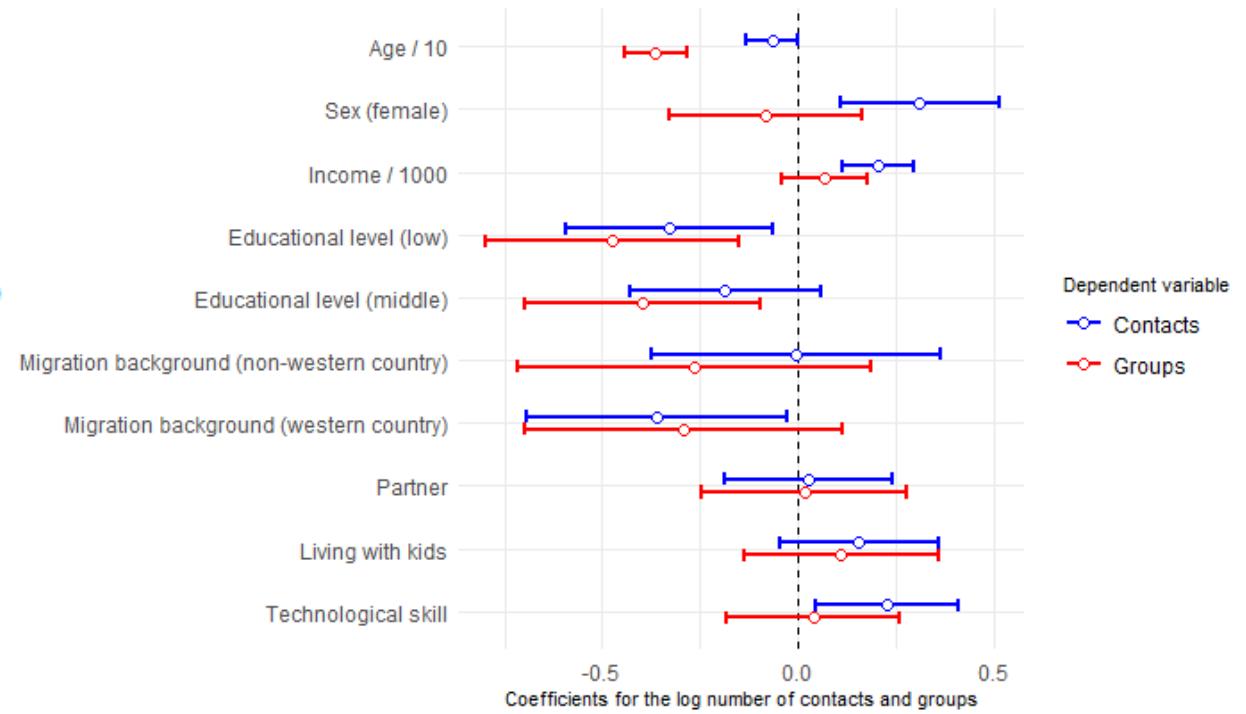
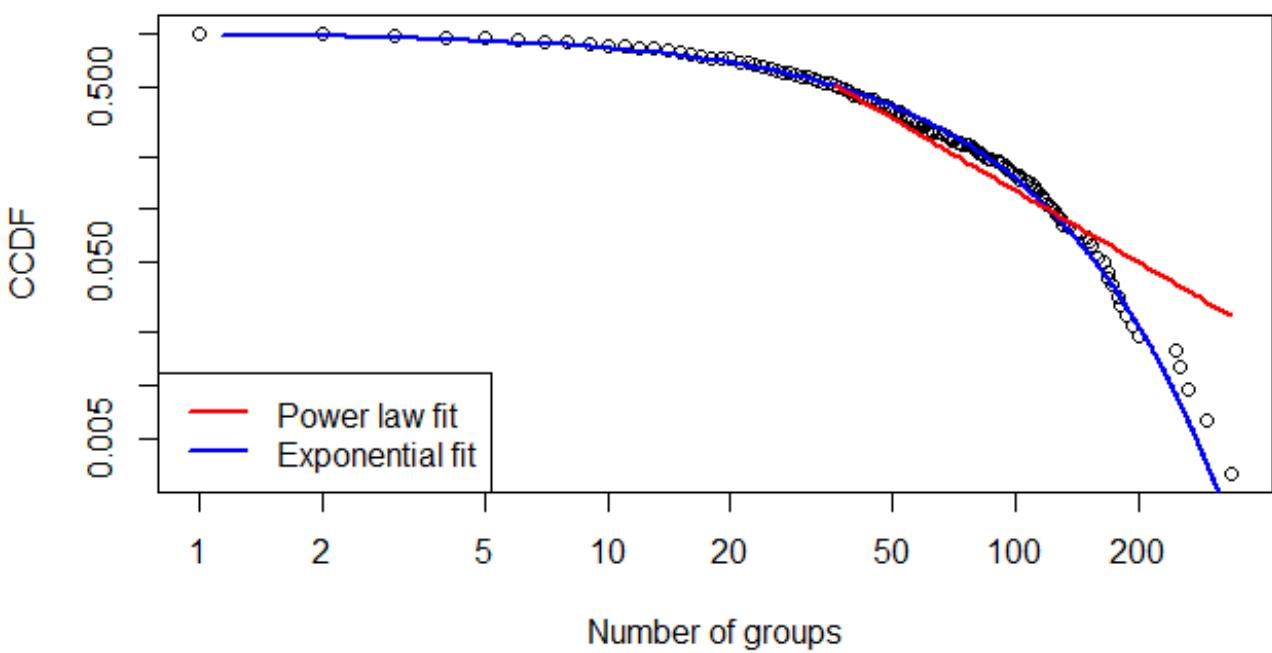
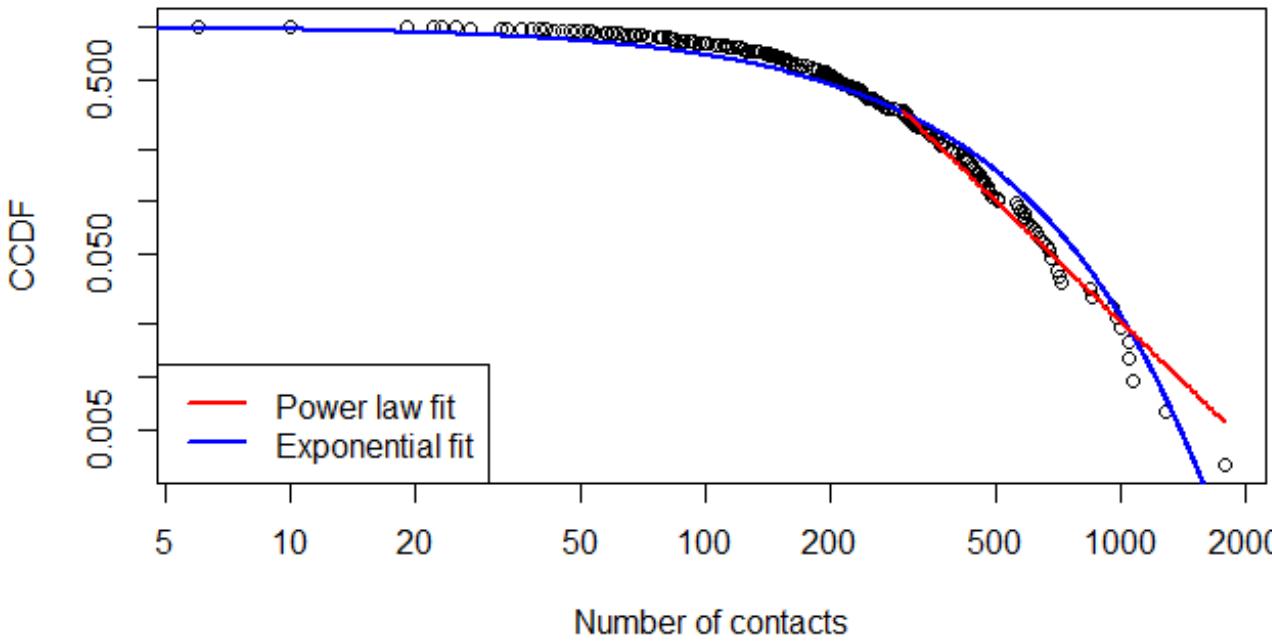


Figure 4. Temporal trends in the frequency of direct messages containing expressions of happiness (red line) and sadness (blue line) for two different participants.

Note. The y-axis differs per plot.

Example 2

Composition and structure of an
Instant Messaging network



Example 3

Sensitivity and Intimacy with Google
Assistant Users



Figure 7: Tipping point in the course of data donation process and data-centric interviews.



Exercise:

- Of the examples discussed:
 - Was this data collection active or passive?
 - Who designed the data?

Exercise

Example 1: Happiness and sadness in Instagram DMs

- Passive / active?
- Who designed the data?

Example 2: Composition and structure of an Instant Messaging network

- Passive / active?
- Who designed the data?

Example 3: Sensitivity and intimacy with Google Assistant Users

- Passive / active?
- Who designed the data?

Advantages, challenges and ethical considerations

Advantages

All	Sensor	Capture / scraping	Data donation
Naturalistic context	Participant effort relatively small	Participant effort relatively small	Data is collected retrospective
Large scale			Data is not influenced by participants "feeling seen"
Detailed data			Relatively cheap to set up
Concepts that were difficult to measure			
Can be supplemented with other sources of data			

Challenges

All	Sensor	Capture / scraping	Data donation
Not everyone can participate	App development is expensive	App development is expensive	Substantive effort from participants is required
A lot of data processing is needed for it to make sense	Expensive to maintain	Expensive to maintain	Every study needs a custom design
Non-participation due to privacy concerns			
Participants need to have digital skills			

Ethical considerations

All	Sensor	Capture / scraping	Data donation
Data is processed at multiple stages, carefully consider who has access at each stage	Participants might forget they are being observed	Participants might forget they are being observed (screenomics)	Participants can inspect their data before sharing and give “true” consent. → Can participants always oversee the large amounts of data?
		Information of contacts might be revealed	Information of contacts might be revealed → local processing can prevent this.



Questions?

Lab meeting:

- Learn how to work with Zeeschuimer and understand how data collected using Zeeschuimer looks like.
- Explore your own digital traces using the “Digital Footprints Explorer”.
- Questions about literature.

Next week:
Platform-centric approaches