UNIVERSITY OF MANNHEIM
School of Social Sciences

Collecting Individual-level Digital Behavioral Data via Data Donation

Florian Keusch
University of Mannheim

✉ f.keusch@uni-mannheim.de
floriankeusch.weebly.com
@floriankeusch.bsky.social

Mannheim Center for Data Science
Data Science in Action Lecture Series
October 17, 2024

Outline

- What are digital behavioral data?
- Ways to collect individual-level digital behavioral data
- The idea of data donation
- Data donation case studies
- Challenges of data donation studies
- Summary & outlook

What Are Digital Behavioral Data?

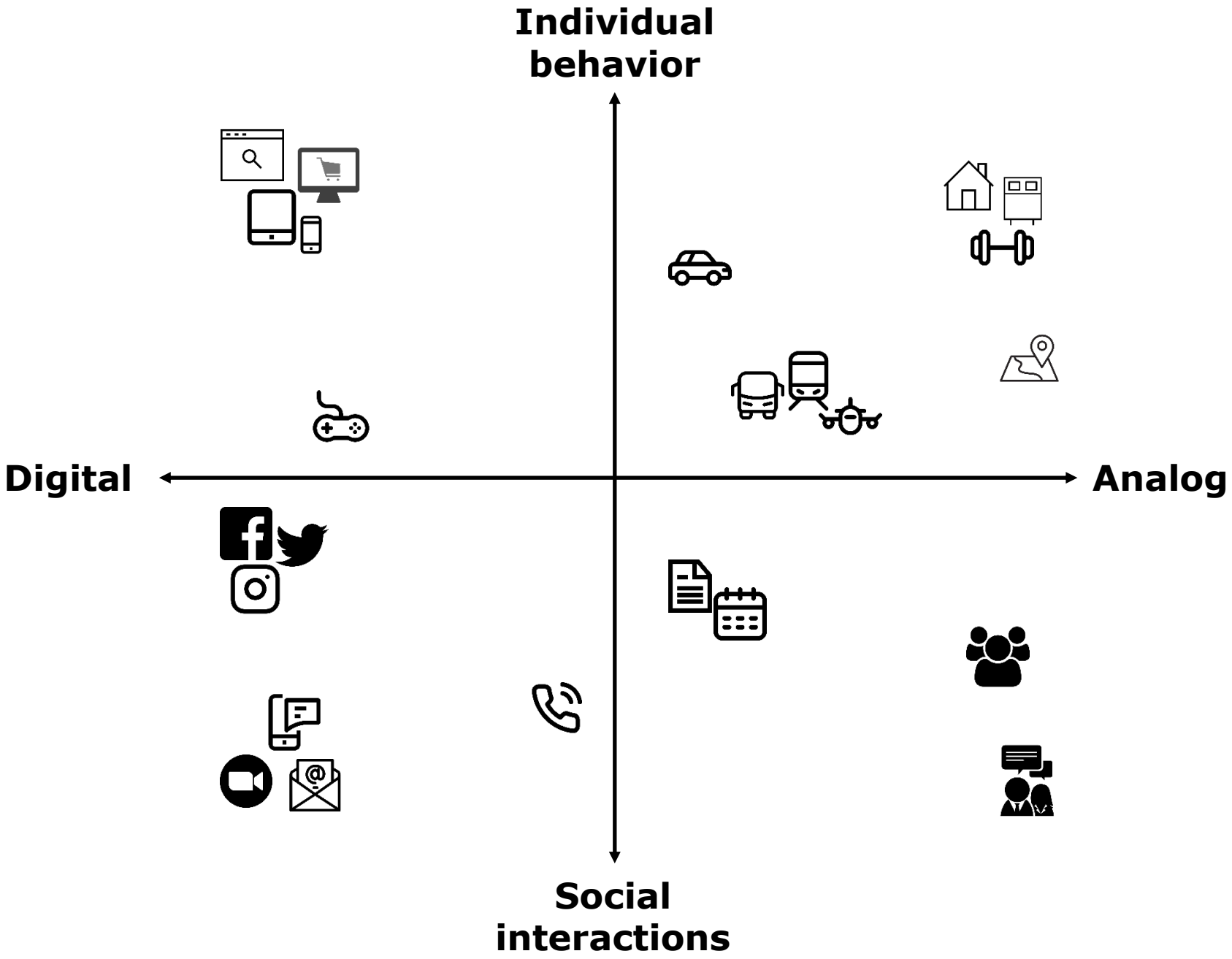
What Are Digital Behavioral Data?

"Records of activity (trace data) undertaken through an online information system (thus, digital)"

(Howison et al. 2011:769)

"Behavioral residue [individuals leave] when they interact online"

(Hinds & Joinson 2018:2)



Keusch & Kreuter (2022)

Benefits of Digital Behavioral Data

- Allow measurement of behavior (in-the-moment) at high frequency
  - Evaluation of moment-to-moment changes
  - Without increasing burden on participants
  - Scalable

- Measurement is nonreactive, i.e., without direct solicitation of subject studied
  - Digital behavioral data should be unaffected by measurement itself
  - Reduced measurement error when measuring…
    - smartphone use (Kobayashi & Boase 2012; Boase & Ling 2013; Andrews, et al. 2015; De Reuver & Bouwman 2015; Revilla et al. 2017; Jones-Jang et al. 2020)
    - online media consumption (Araujo et al. 2017; Haenschen 2020; Revilla et al. 2017; Scharkow 2016)
    - mobility (Stopher et al. 2007; Scherpenzeel 2017)

How to Collect Individual-Level Digital Traces

How to Collect Individual-Level Digital Traces

- Use APIs

Meta

An Update on Our Plans to Restrict Data Access on Facebook

April 4, 2018

Chesnot / Getty Images

SCIENCE / TWITTER / TECH

Twitter just closed the book on academic research / Twitter was once an indispensable resource for academic research. That's changed under Elon Musk.

By Justine Calma, a science reporter covering the environment, climate, and energy with a decade of experience. She is also the host of the Hell or High Water podcast.

May 31, 2023, 3:19 PM GMT+2 | 23 Comments / 23 New

https://www.theatlantic.com/technology/archive/2023/02/elon-musk-twitter-ethics-algorithm-biases/673110/

FEATURE

Reddit pricing: API charge explained

Providing a free API was becoming costly for Reddit. Enterprise-scale developers now have to pay for access to Reddit's data.

By Ben Lutkevich, Technical Features Writer Published: 11 Jul 2023

Shutterstock

How to Collect Individual-Level Digital Traces

- Use APIs
- Collaborate with industry

How to Collect Individual-Level Digital Traces

- Use APIs
- Collaborate with industry
- Have users install meters and apps that continuously…
  - …track information on web browsing
  - …logs usage behavior, (native) mobile browsing, and sensor readings

- Allows tracking of individual behavior over longer periods of time
- Various commercial and non-commercial tools available
  - e.g., Wakoopa, Movisense, Mumuras, Ethica, Aware, Beiwe, ResearchStack, ResearchKit, …
- Some (market) research organizations maintain "metered panels"
  - e.g., Netquest, Gapfish, YouGov, GESIS Panel.dbd, …

Browsing and App Use Data

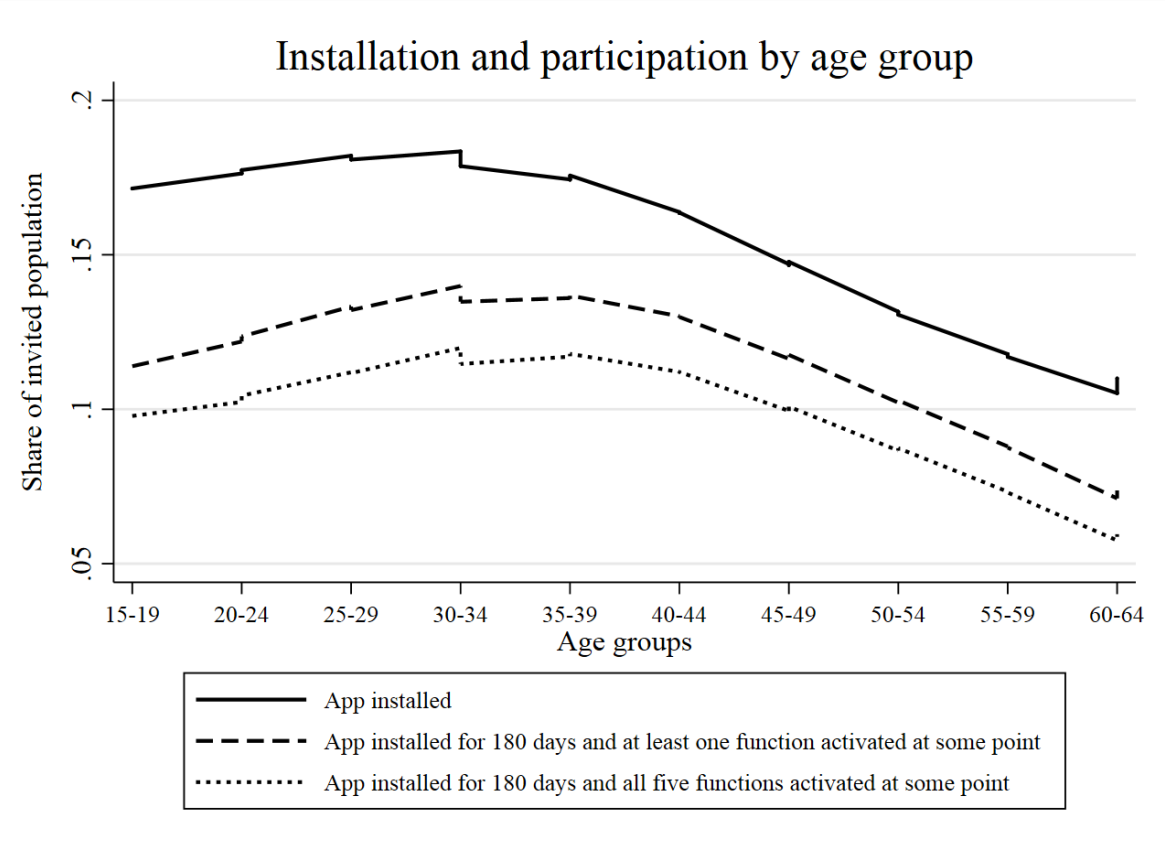
ID	URL	used_at	active_seconds
1	tipico.de/de/online-sportwetten	2017-09-12 10:45:15	83
1	rib-software.com/de/loesungen/architektur-und-bauplanung/arriba-ava.html	2017-09-12 10:45:19	40
1	wetteronline.de/wettertrend/cala-d-or	2017-09-12 10:45:23	182
1	facebook.com/messages/t	2017-09-12 10:45:36	2
1	facebook.com/messages/t/Peter.Mustermann	2017-09-12 10:45:44	74
1	severin-produkttest.de	2017-09-12 10:45:55	8
1	severin-produkttest.de/bewerbung	2017-09-12 10:46:05	46
1	sparda-b.de	2017-09-12 10:46:37	6
2	rp-online.de/nrw/staedte/duesseldorf/airport-duesseldorf-flughafen-nicht-verantwortlich-fuer-das-was-dort-passiert-aid-1.7075717	2017-09-12 11:19:48	9
2	autoscout24.de/ergebnisse?mmvmk0=29&mmvmd0=15537&mmvco=1&body=5&pricefrom=0&priceto=8000&cy=D&zipr=50&fuel=B&powertype=kw&offer=U&offer=J&offer=O&offer=D&lat=49.15148&lon=9.17345&zip=Heilbronn&atype=C&ustate=N%2CU&sort=standard&desc=0&page=1&size=20	2017-09-12 12:01:10	28
2	trauer.weser-kurier.de	2017-09-12 12:06:45	5
2	amazon.de/Kee-Hearts/dp/B072Z8DMRY/ref=sr_1_1?ie=UTF8&qid=1505211277&sr=8-1&keywords=kee+of+hearts	2017-09-12 12:14:44	10
2	sat1gold.de/tv/zwei-bei-kallwass/video/117-brunos-geheimnis-ganze-folge	2017-09-12 12:15:10	49
2	bibeltv.de/livestreams	2017-09-12 12:28:40	1
2	arbeitsagentur.de/meine-eservices	2017-09-12 12:36:46	11
3	facebook.com	2017-09-12 12:55:35	6
3	bild.de/news/ausland/hurrikan-irma/irma-live-ticker-karibik-puerto-rico-domrep-haiti-florida-usa-53130320.bild.html	2017-09-12 12:55:41	2
3	spiegel.de	2017-09-27 09:56:46	14
3	https://www.spiegel.de/politik/deutschland/wolfgang-schaeuble-ueber-grosse-koalitionen-kann-auf-dauer-nur-schiefgehen-a-1299654.html	2017-09-27 09:57:00	195

ID	app_n	used_at	active_seconds	app_cat
1	To-Do Calendar Planner	2017-09-12 10:45:15	83	Email, Messaging & Telephone
1	WhatsApp Messenger	2017-09-12 10:45:19	40	Email, Messaging & Telephone
1	Kleiderkreisel	2017-09-12 10:45:23	182	Lifestyle, food & nightlife
1	Camera	2017-09-12 10:45:36	2	Audio, video & entertainment
1	eBay Kleinanzeigen for Germany	2017-09-12 10:45:44	74	Lifestyle, food & nightlife
1	dm glÃ¼ckskind	2017-09-12 10:45:55	8	Lifestyle, food & nightlife
1	GMX Mail	2017-09-12 10:46:05	46	Email, Messaging & Telephone
1	WhatsApp Messenger	2017-09-12 10:46:37	6	Email, Messaging & Telephone
2	WhatsApp Messenger	2017-09-12 11:19:48	9	Email, Messaging & Telephone
2	WhatsApp Messenger	2017-09-12 12:01:10	28	Email, Messaging & Telephone
2	WhatsApp Messenger	2017-09-12 12:06:45	5	Email, Messaging & Telephone
2	WhatsApp Messenger	2017-09-12 12:14:44	10	Email, Messaging & Telephone
2	WhatsApp Messenger	2017-09-12 12:15:10	49	Email, Messaging & Telephone
2	WhatsApp Messenger	2017-09-12 12:28:40	11	Email, Messaging & Telephone
2	GMX Mail	2017-09-12 12:36:46	11	Email, Messaging & Telephone
3	GMX Mail	2017-09-12 12:55:35	243	Email, Messaging & Telephone
3	GMX Mail	2017-09-12 12:55:41	24	Email, Messaging & Telephone
3	Period tracker	2017-09-27 09:56:46	14	Health & fitness
3	Pedometer	2017-09-27 09:57:00	195	Health & fitness

See Clemm von Hohenberg et al. (2024) on how to analyse web browsing data

Challenges When Using Meters and Apps

- Coverage & nonparticipation
  - Who uses the technology and is willing to share their data?



Keusch et al. (2022)

Reasons not to participate	
Privacy, data security concerns	44%
No incentive, incentive too low	17%
Not enough information provided	12%
Do not download any apps	8%
Not interested	6%
Not enough time/Study too long	5%
Don't use smartphone enough	5%
Not enough storage	1%
Other reasons	6%
n=1,154	

Keusch et al. (2019)

Challenges When Using Meters and Apps

- Coverage & nonparticipation
- Ethics, privacy, & data security
  - Passively collected data might be considered very sensitive
  - Users might not (fully) understand what data are collected and how they are being used
  - To comply with GDPR…
    - …multiple consent steps required (trade-off between informed consent and burden)
    - …data needs to be stored in Germany/EU
  - ➢ "Whitelisting" of URLs, pre-processing data on device before transfer to server, option to temporarily turn-off tracking, privacy nudging,…

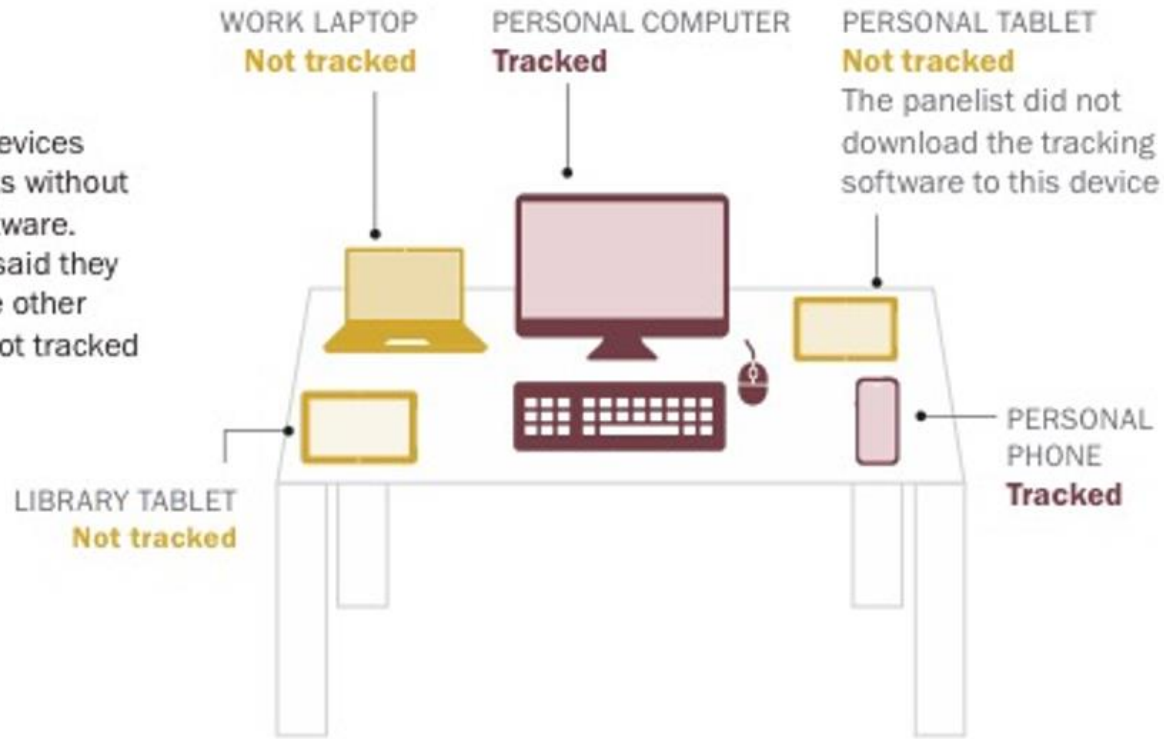
Challenges When Using Meters and Apps

- Coverage & nonparticipation
- Ethics, privacy, & data security
- Missing data

DEVICE TRACKING

Tracked:
Devices tracked in this study included computers, tablets and mobile phones used by panelists that they downloaded tracking software onto

Not tracked:
Any additional devices used by panelists without the tracking software. Many panelists said they had one or more other digital devices not tracked



Source: https://www.pewresearch.org/journalism/2020/12/08/measuring-news-consumption-in-a-digital-era/

See also Bosch et al. (2024)

Challenges When Using Meters and Apps

- Coverage & nonparticipation
- Ethics, privacy, & data security
- Missing data

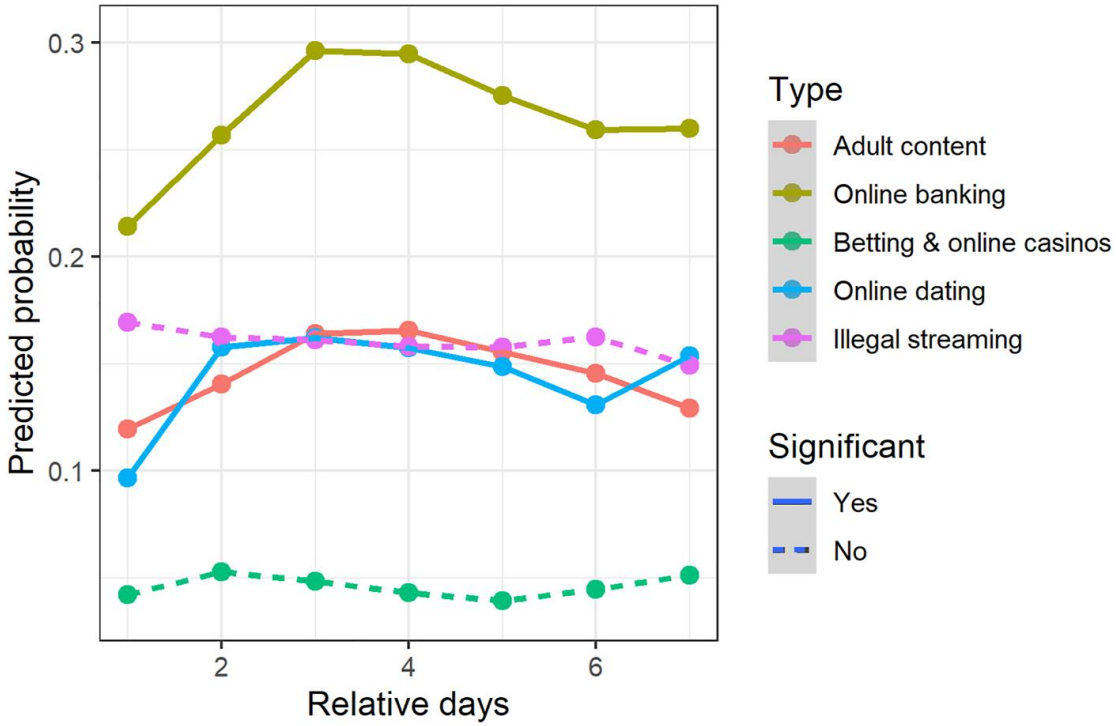
- Comparison of traced FB use to self-report (Pankowska et al. under review)
  - Frequent (C2) and infrequent users (C3) relatively equal
  - One class (C1) with large inconsistencies, especially if smartphone data missing

	Class		
	1	2	3
Size	0.38	0.33	0.29
FB use survey			
Once a month or less	0.00	0.00	0.30
Several times a month	0.01	0.02	0.31
Several times a week	0.10	0.13	0.37
Daily	0.46	0.35	0.02
Several times a day	0.44	0.50	0.00
FB use tracking app			
Once a month or less	0.52	0.01	0.61
Several times a month	0.13	0.02	0.14
Several times a week	0.23	0.17	0.20
Daily	0.10	0.00	0.05
Several times a day	0.01	0.80	0.01

See also Bosch et al. (2024)

Challenges When Using Meters and Apps

- Coverage & nonparticipation
- Ethics, privacy, & data security
- Missing data
- Reactivity



Keusch et al. (2023)

Source: https://twitter.com/mbrennanchina/status/1128201958962032641

How to Collect Individual-Level Digital Traces

- Use APIs
- Collaborate with industry
- Have users install meters and apps
- Ask users to “donate” data
  - User-centered approach for retrospective digital behavioral data collection

The Idea of Data Donation

Data Donation

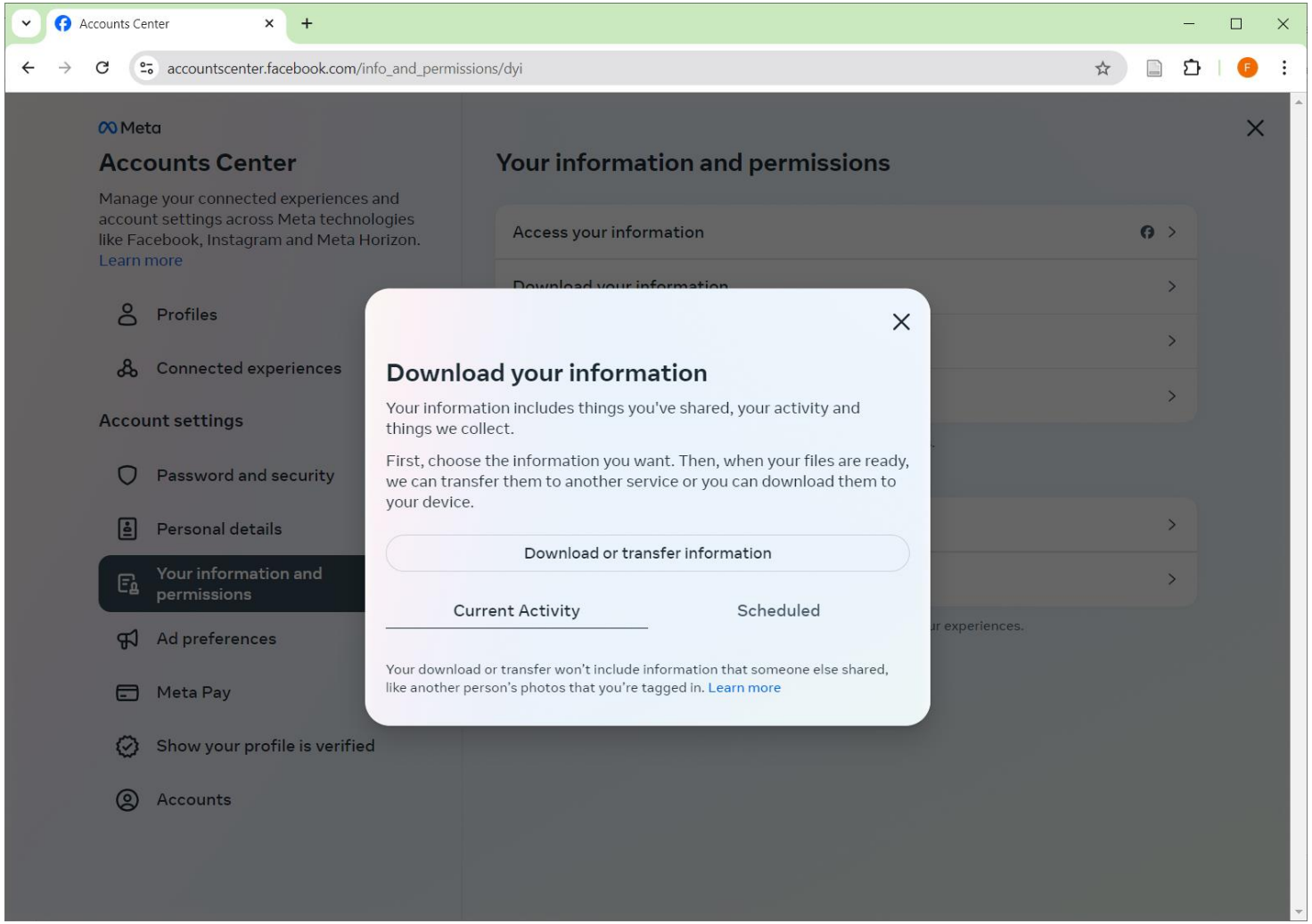
- Takes advantage of legislation such as *General Data Protection Regulations* (GDPR), *California Consumer Privacy Act* (CCPA), and *Lei Geral de Proteção de Dados* (LGPD) in Brazil
  - GDPR Art. 15: Right of access by the data subject
  - GDPR Art. 20: Right to data portability

- Now often implemented in online surveys (see Haim et al. 2023)
  - Respondents asked for willingness to donate data at end of survey
  - Respondent requests data download package (DDP) from data controller
  - Respondent donates DDP to researcher
    - Via direct sharing (e.g., ftp server, email)
    - Privacy-preserving data donation platforms

Data Donation

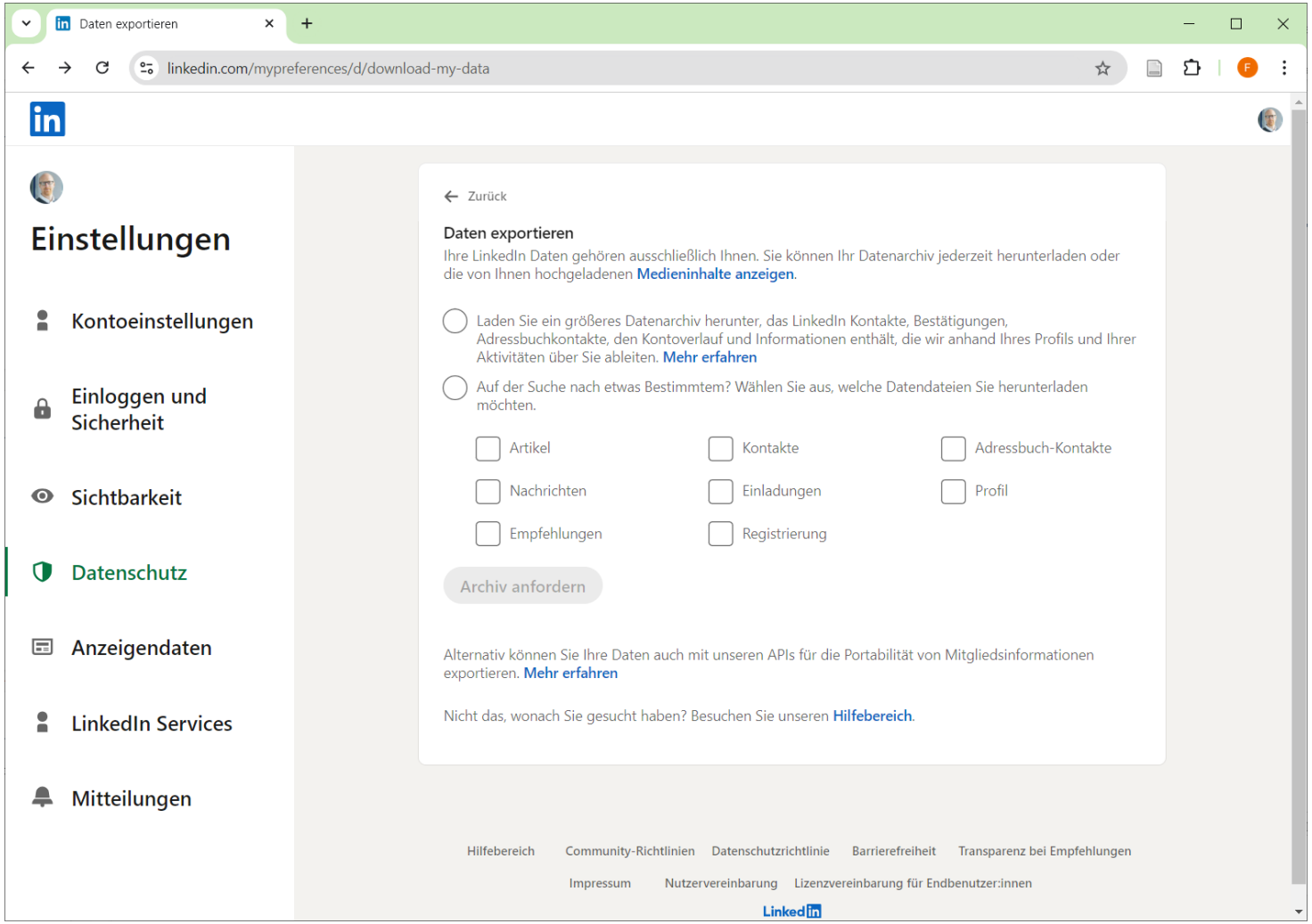
https://datadonation.eu/

Users Request DDP – Meta



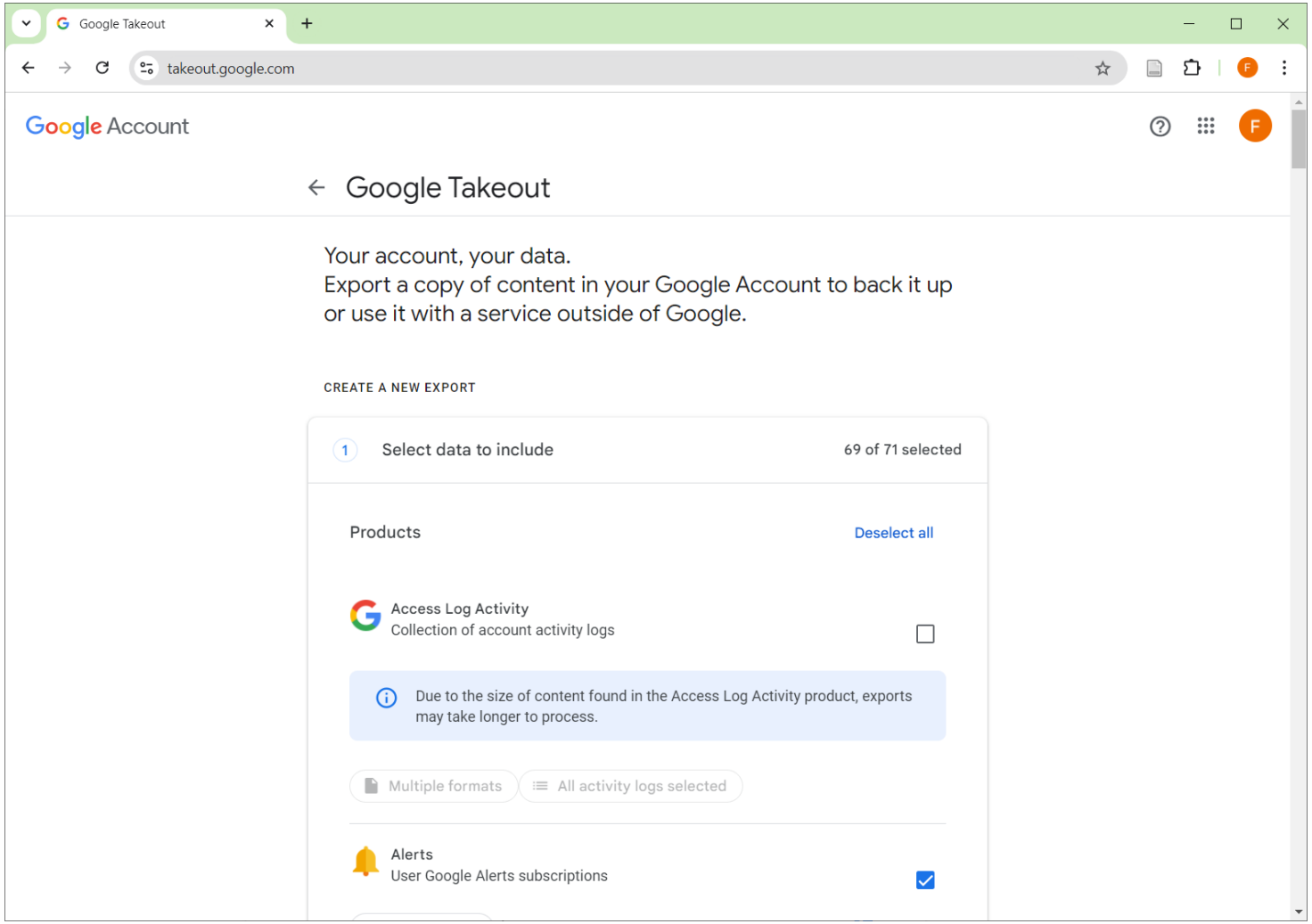
https://accountscenter.facebook.com/info_and_permissions/dyi

Users Request DDP – LinkedIn



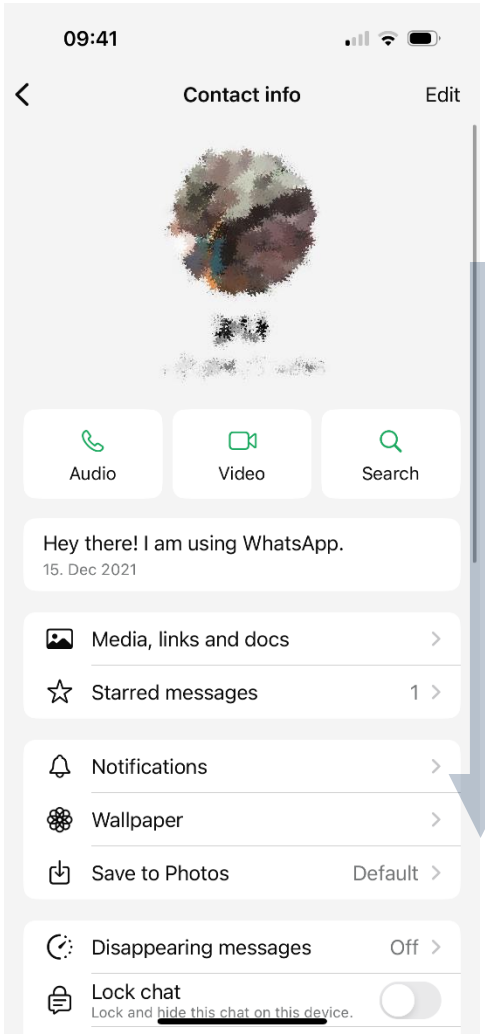
https://www.linkedin.com/mypreferences/d/categories/privacy

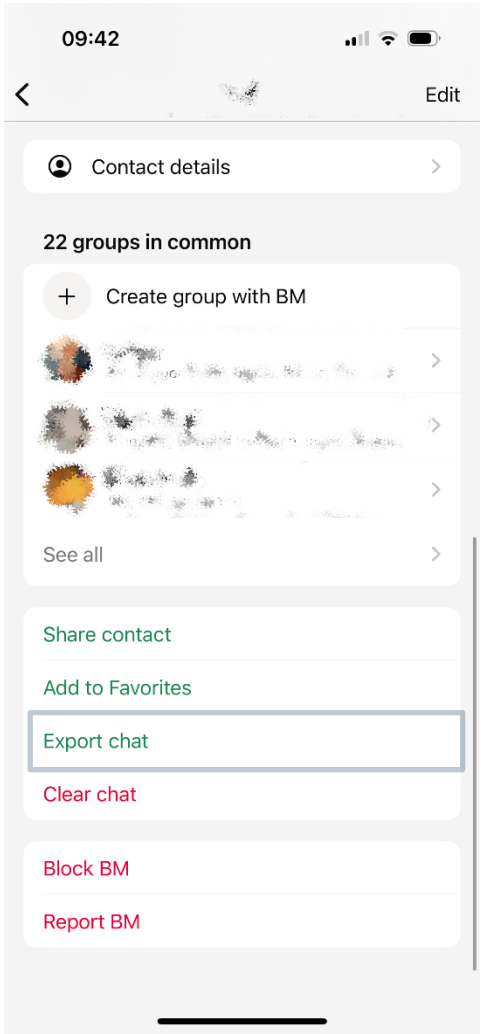
Users Request DDP – Google



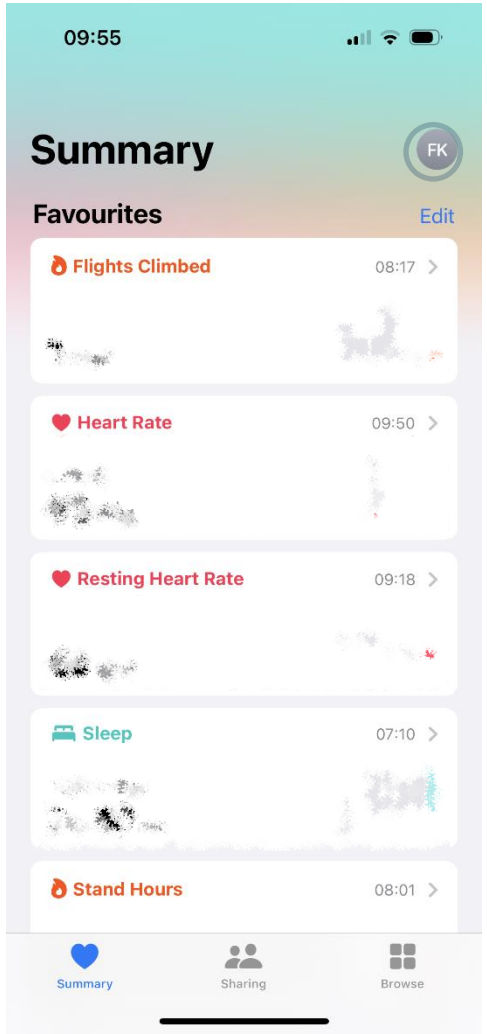
https://takeout.google.com/

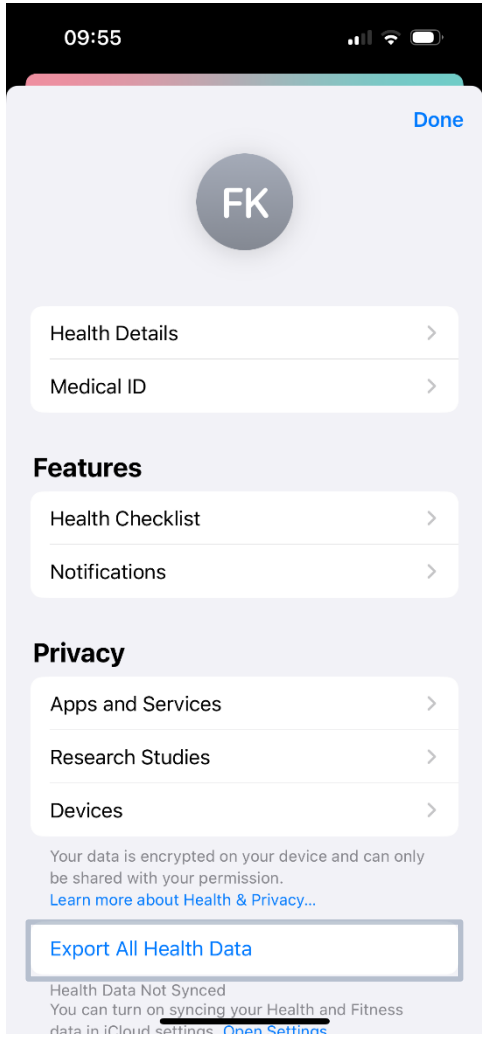
Users Request DDP – WhatsApp





Users Request DDP – Apple Health





What Data Are Available?

- In theory, everthing data controller collects about user needs to be provided back to them and can be donated
- Data available depends on type of data controller (i.e., social media platform, health app, etc.)
- For typical social media platforms (e.g., Facebook, Instagram, Youtube), data includes…
  - …digital user profiles (e.g., sociodemographics, privacy settings)
  - …digital activity (e.g., friends, likes, searches, privat messages, passive usage)
  - …digital targeting (e.g., ads, algorithmically inferred interrests)
  - …analoge activities (e.g., attandance at events, life events)

Job-related Data in DDPs

		Facebook	Instagram	TikTok	Google/ YouTube	LinkedIn	XING
DDP Request	File format of DDP	HTML or JSON	HTML or JSON	TXT or JSON	depends on service	CSV	HTML
	Time until participants can download DDP	< 24h	< 24h	< 96h	< 48h	< 48h	< 72h
	DPP arrives as single file	✓	✓	✓	✓*[1]	×*[2]	×*[3]
Data in DDPs	Exemplary variables: *Ties and social interactions*						
	Friends & followers	✓	✓	✓	✓	✓	✓
	Published posts & likes	✓	✓	✓	(✓)	✓	✓
	Exemplary variables: *Self-presentation*						
	CV-like information	×	×	×	×	✓	✓
	General profile information	✓	✓	✓	✓	✓	✓
	Exemplary variables: *Job-related activities*						
	Search history	✓	✓	✓	✓	✓	✓
	(Submitted) applications	×	×	×	×	✓	✓
	Exemplary variables: *Job-related microtargeting*						
	Targeted ads/algorithmically inferred interests	✓	✓	✓	×	✓	✓
	Consumed advertisement	✓	✓	(✓)	✓	✓	✓

*[1] only split if large, *[2] basic and advanced, *[3] internal and external

Source: own research

Hase et al. (2023)

Local Data Processing

- User receives DDP(s) (e.g., .csv, .json, .html) from data controller and temporarily stores them on their device

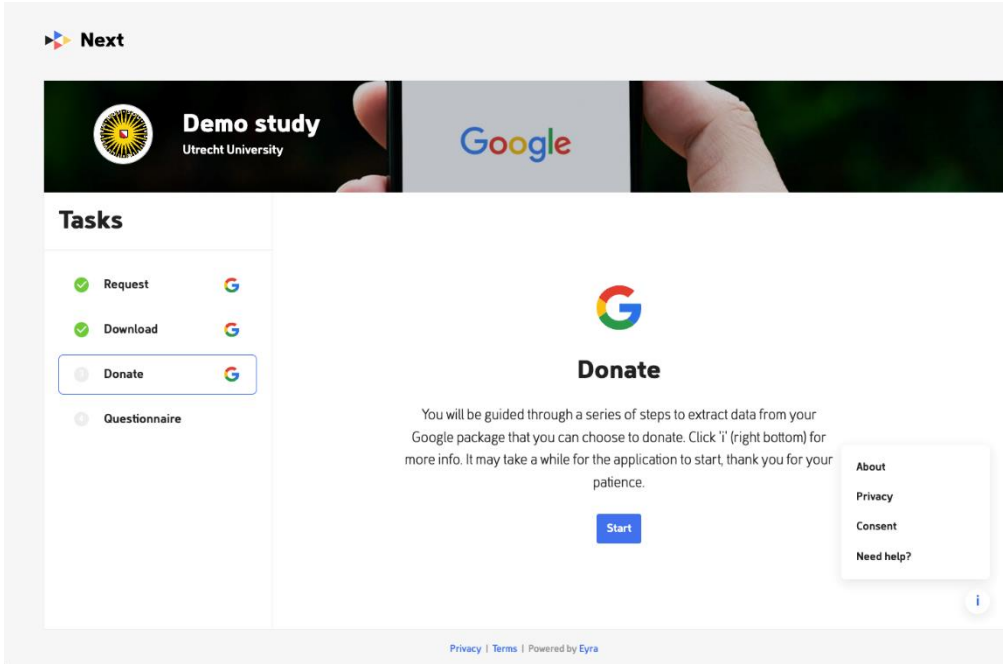
- Using data donation platform, DDPs can be locally processed
  - Script reads in DDP and selects relevant files to process
  - Data of interest extracted and potentially processed further (e.g., aggregating step count to day-level, whitelisting, striping any PII)
  - Interactive script elements might be used for respondent to engage with data (e.g., annotation, provide information about chat partner)
  - Processed data presented back to respondent as table or graph for review and respondent might delete certain data points

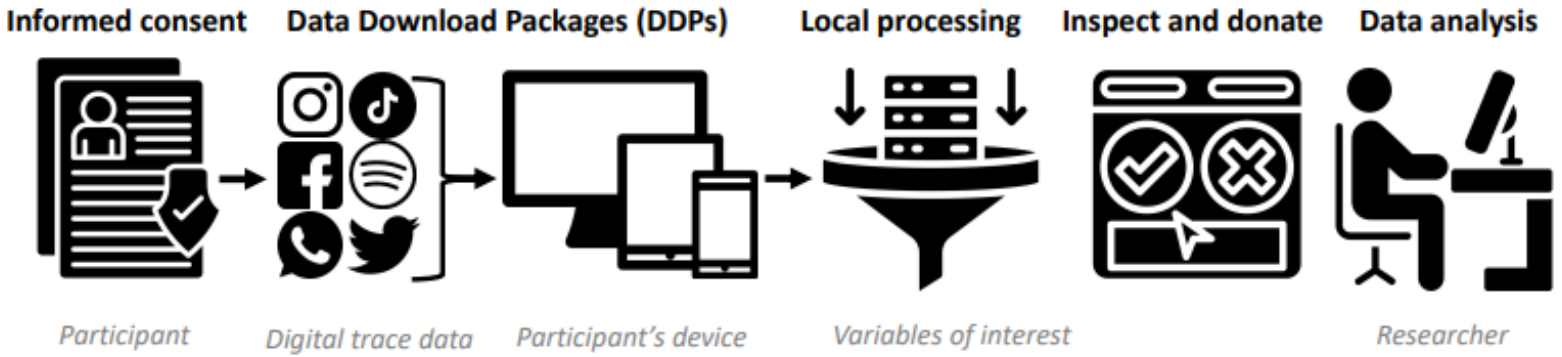
➢ Only final pre-processed and reviewed data set transmitted

Carrière et al. (2024)

Local Data Processing

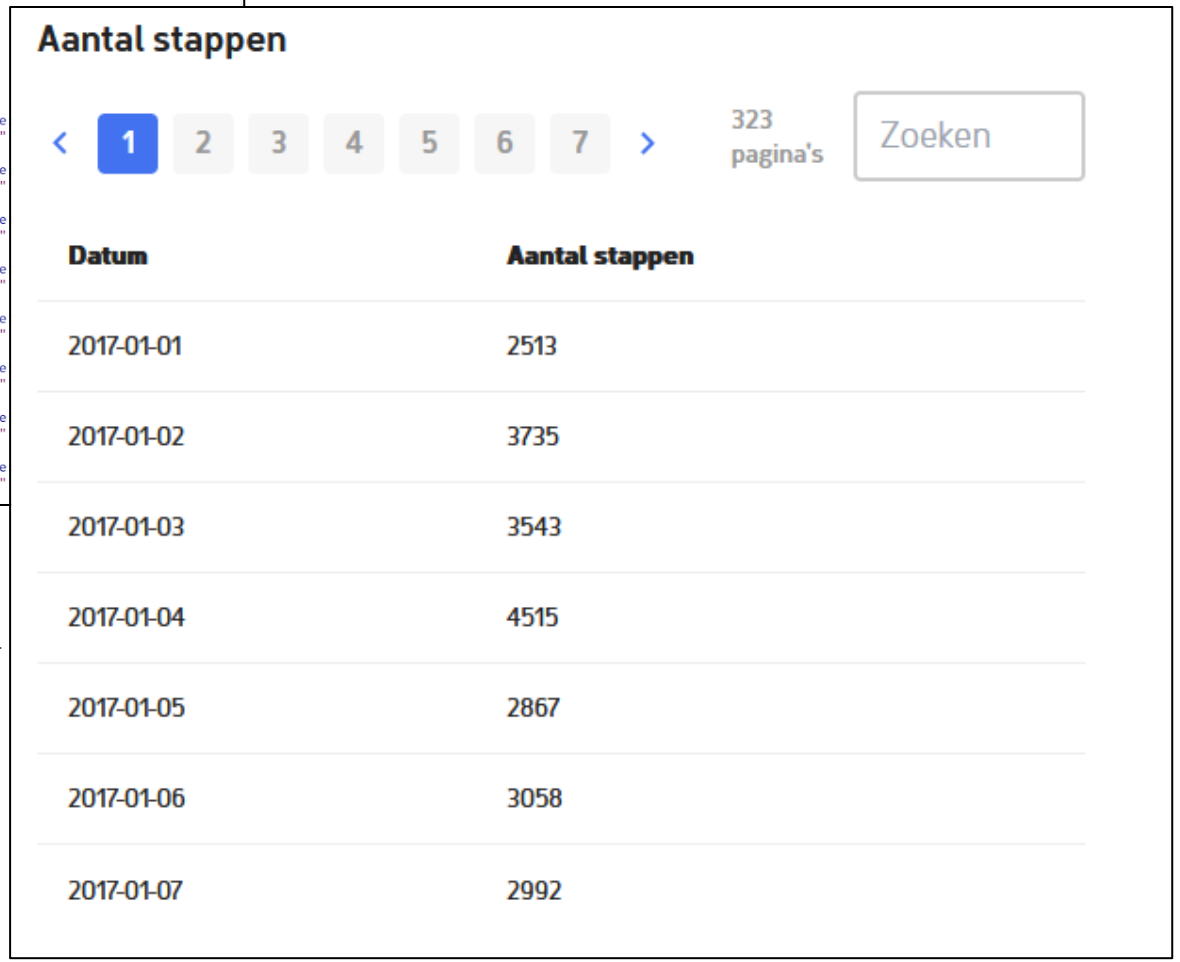
[Data Donation Task/Port](#) & [Next](#)





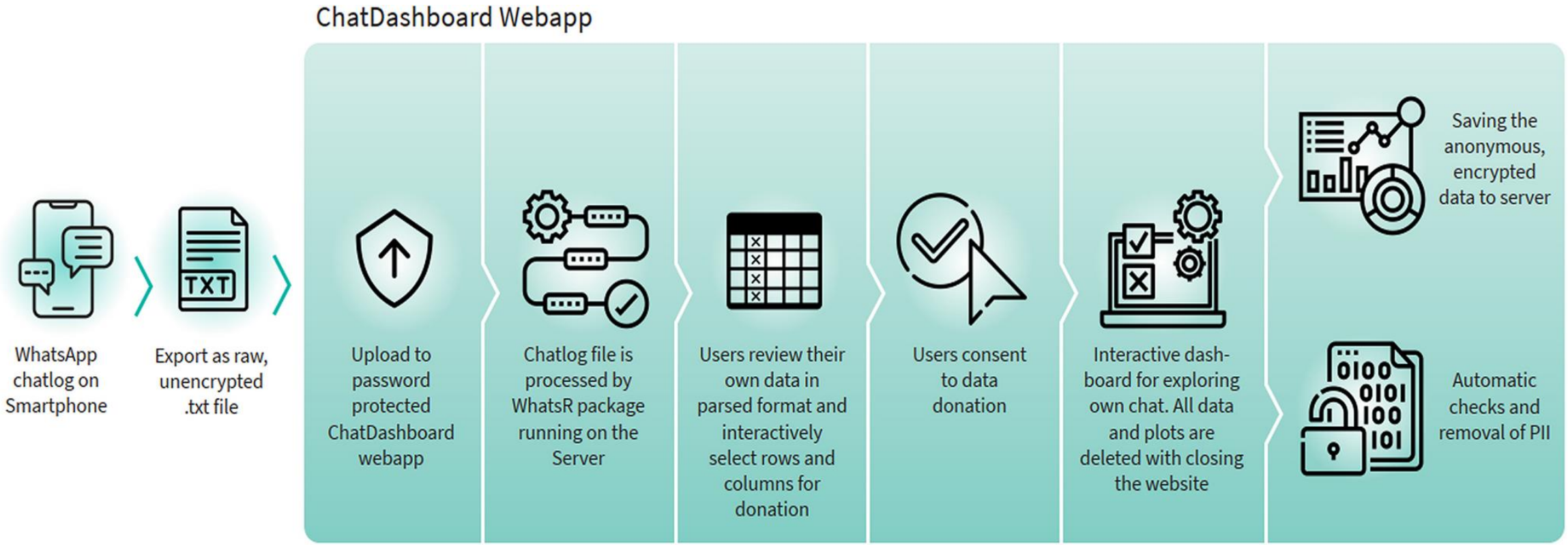
Boeschoten et al. (2023)

Local Data Processing – Apple Health



Local Data Processing

ChatDashboard & WhatsR



Data Donation

Advantages

+ Allows access to longitudinal digital behavioral data that cannot be collected otherwise

+ Works for large range of platforms
  - WhatsApp, Facebook, Instagram, Google, LinkedIn, YouTube, Netflix, Apple Health, Fitbit,…

+ Fully GDPR-compliant
  - User retains full control over what data are donated

Challenges

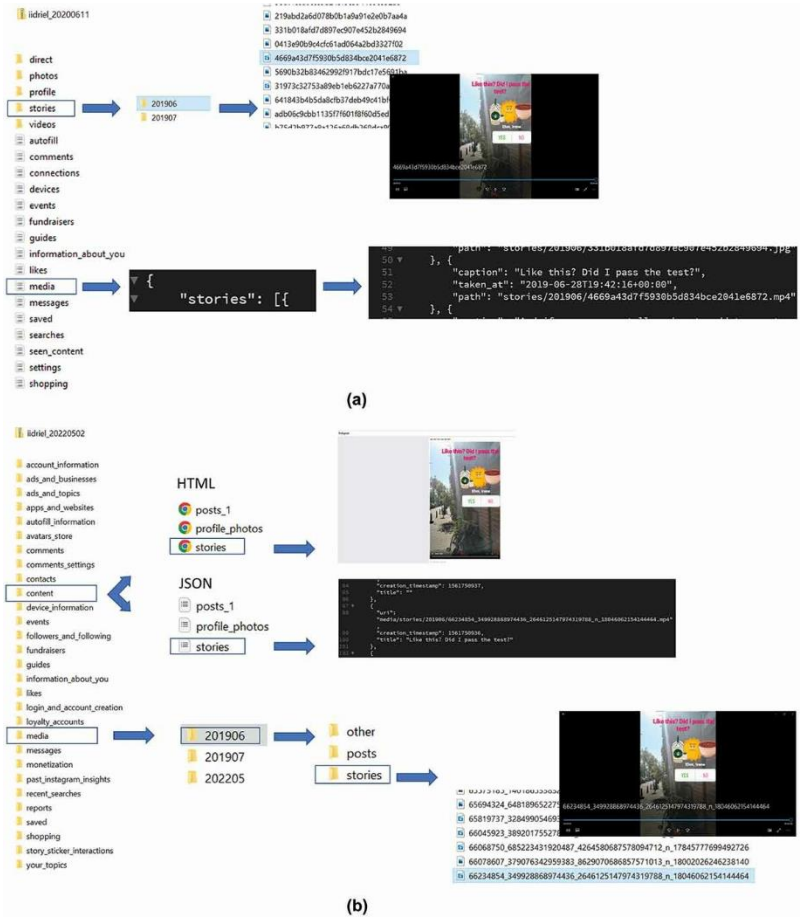
- Data donation process can be rather cumbersome for users (consent, data download, data storage, data upload)

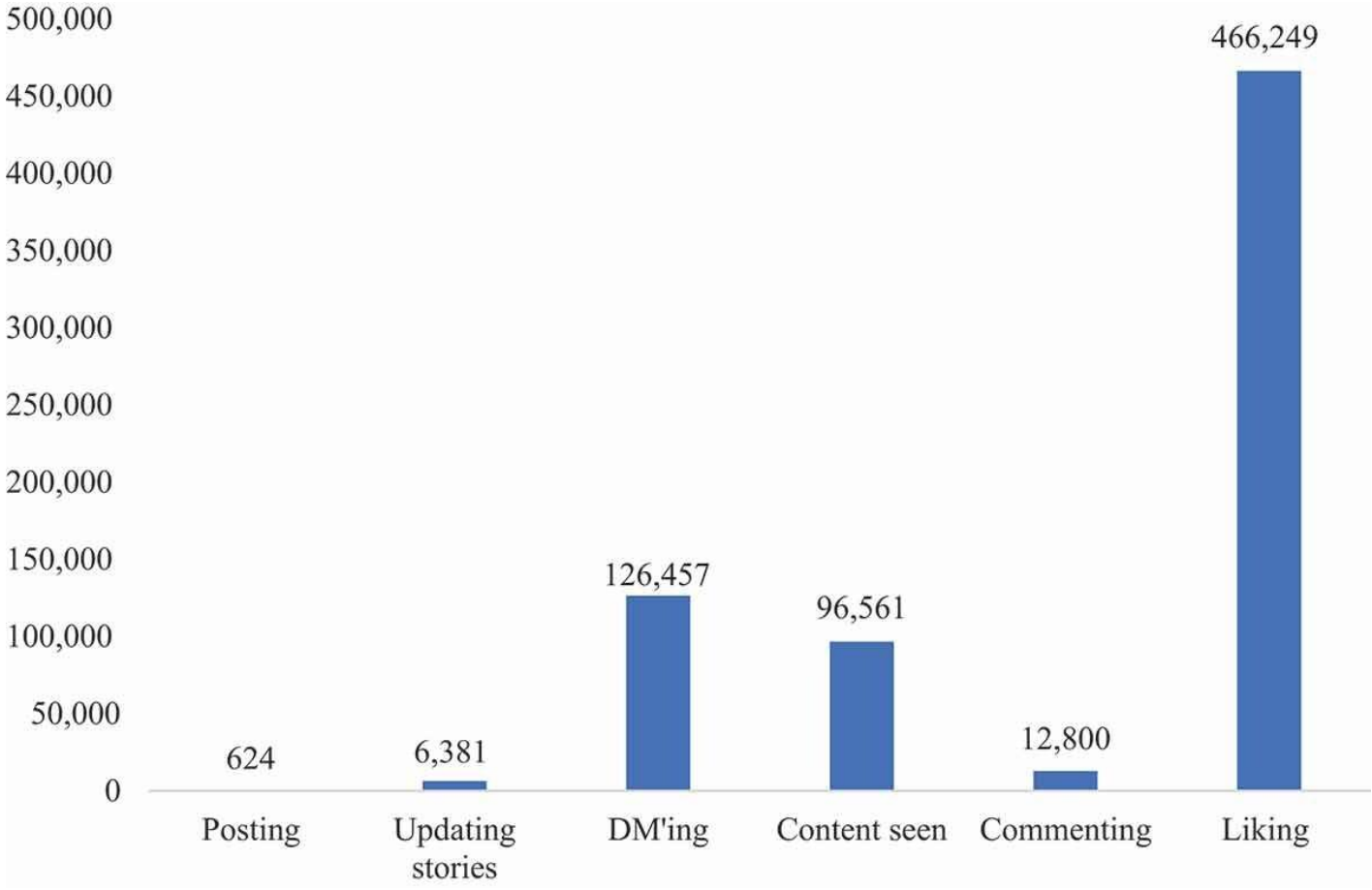
- (Basic) technical skills needed to participate

Data Donation Case Studies

Example Studies

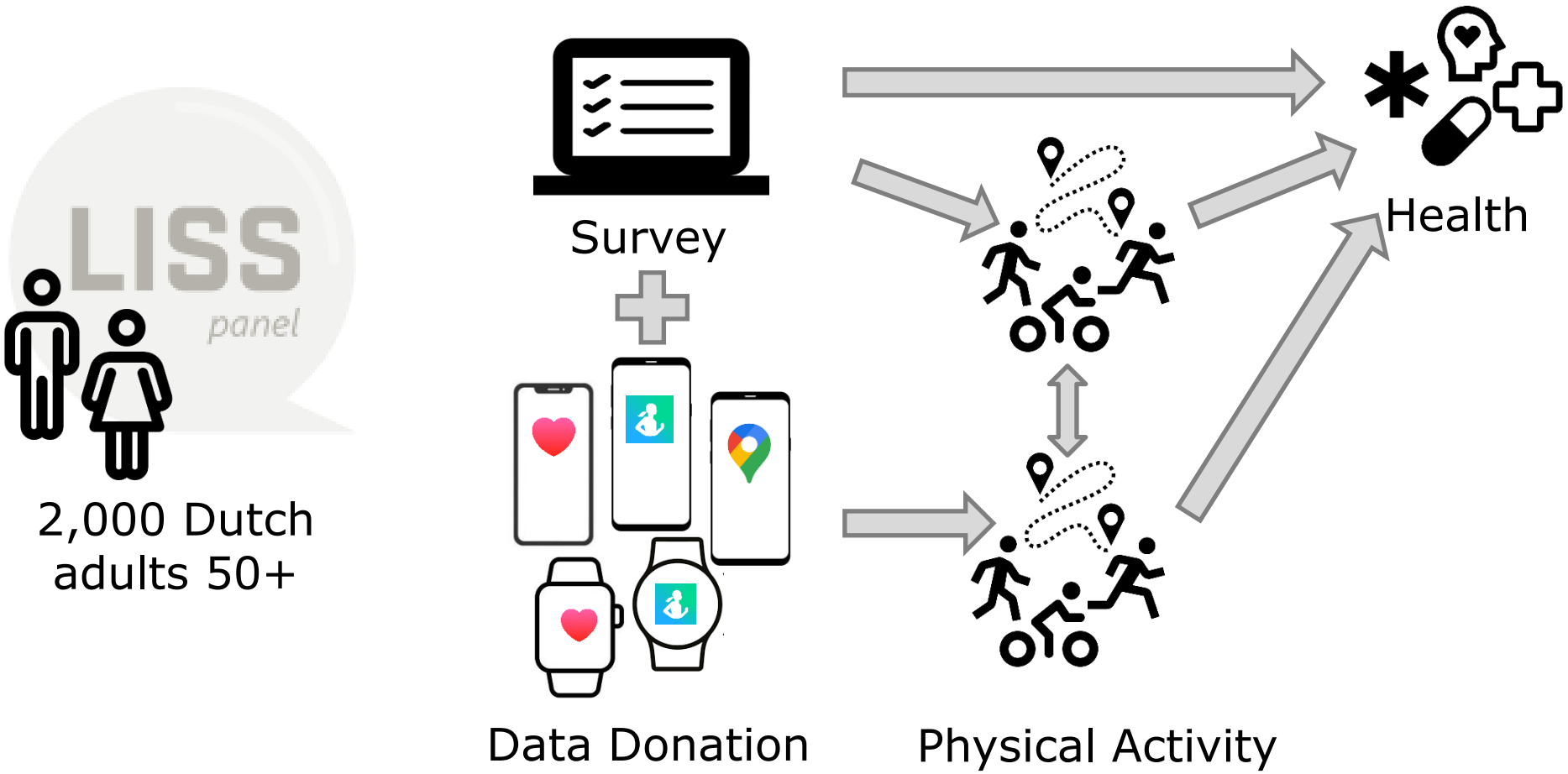
- Instagram activities among 8th and 9th graders (van Driel et al. 2022)





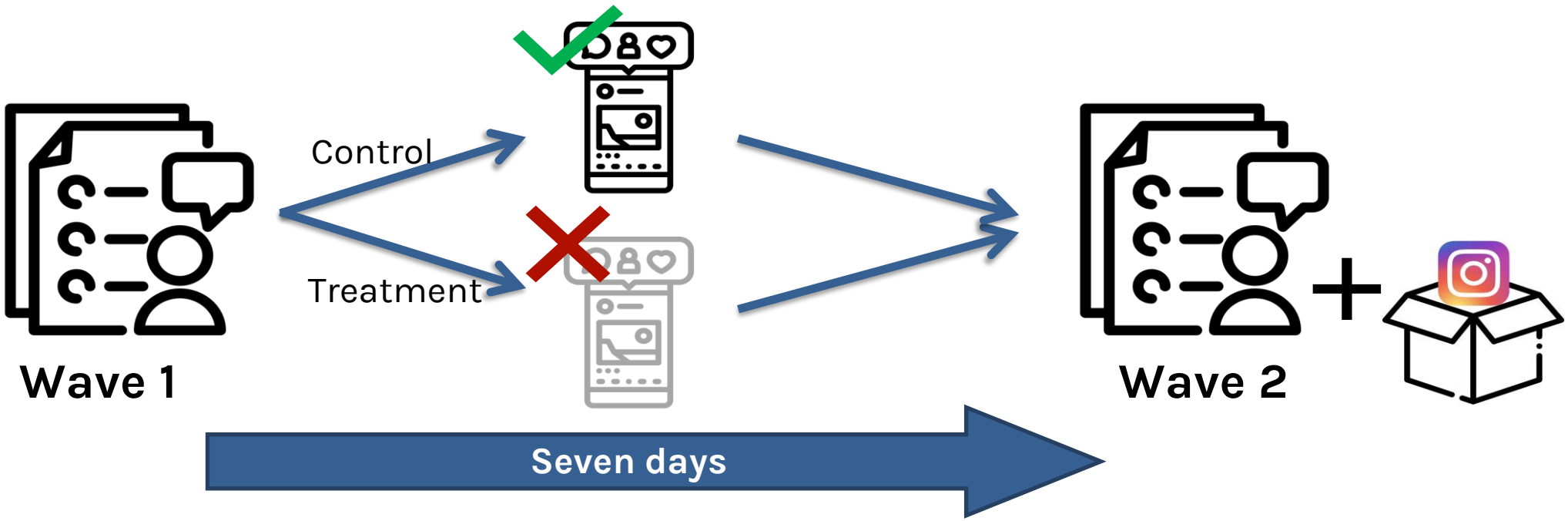
Example Studies

- Physical activity of older adults (Keusch & Struminskaya in preparation)

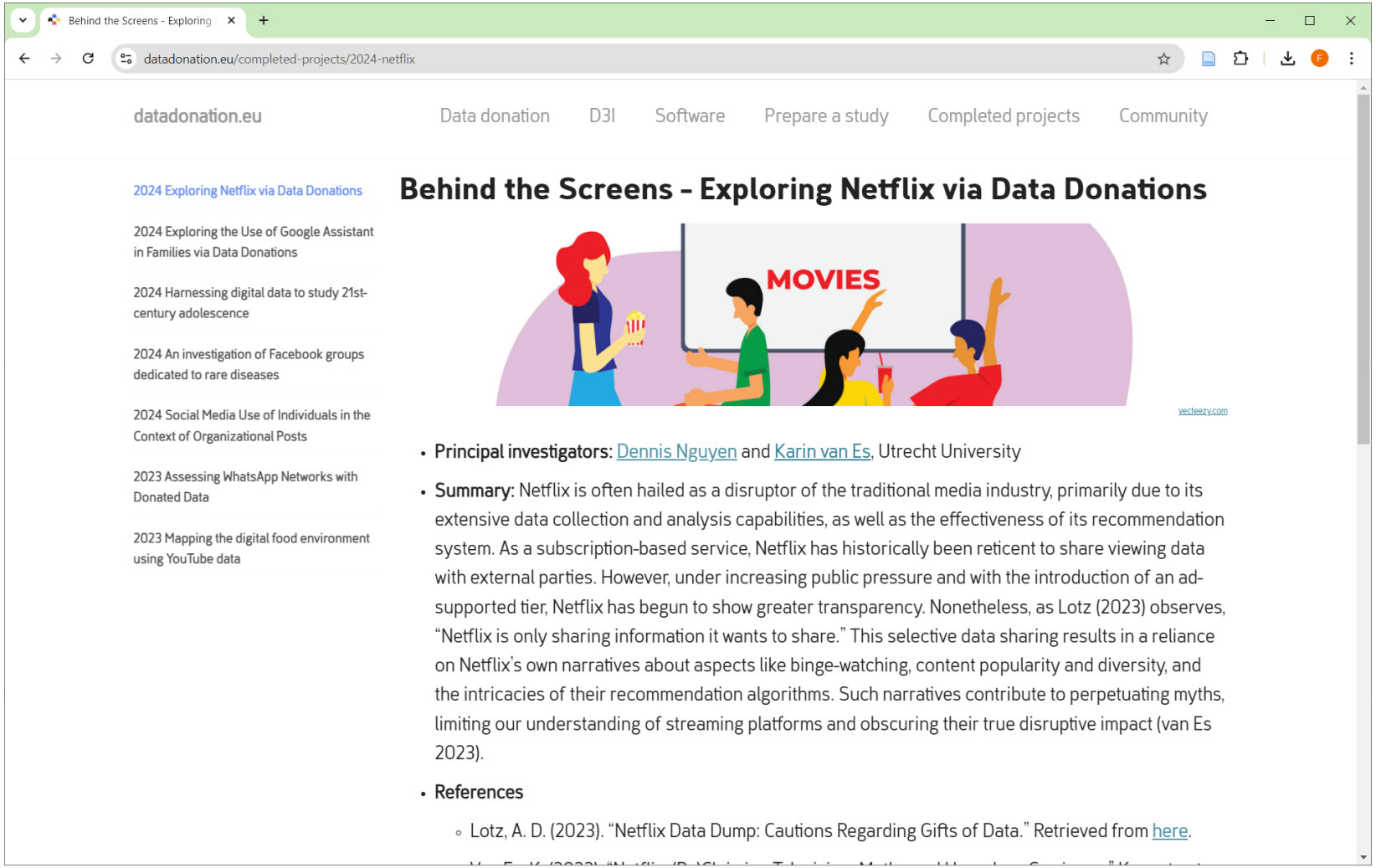


Example Studies

- Effect of Instagram abstinence on body image (Szafran et al. in preparation)



Example Studies

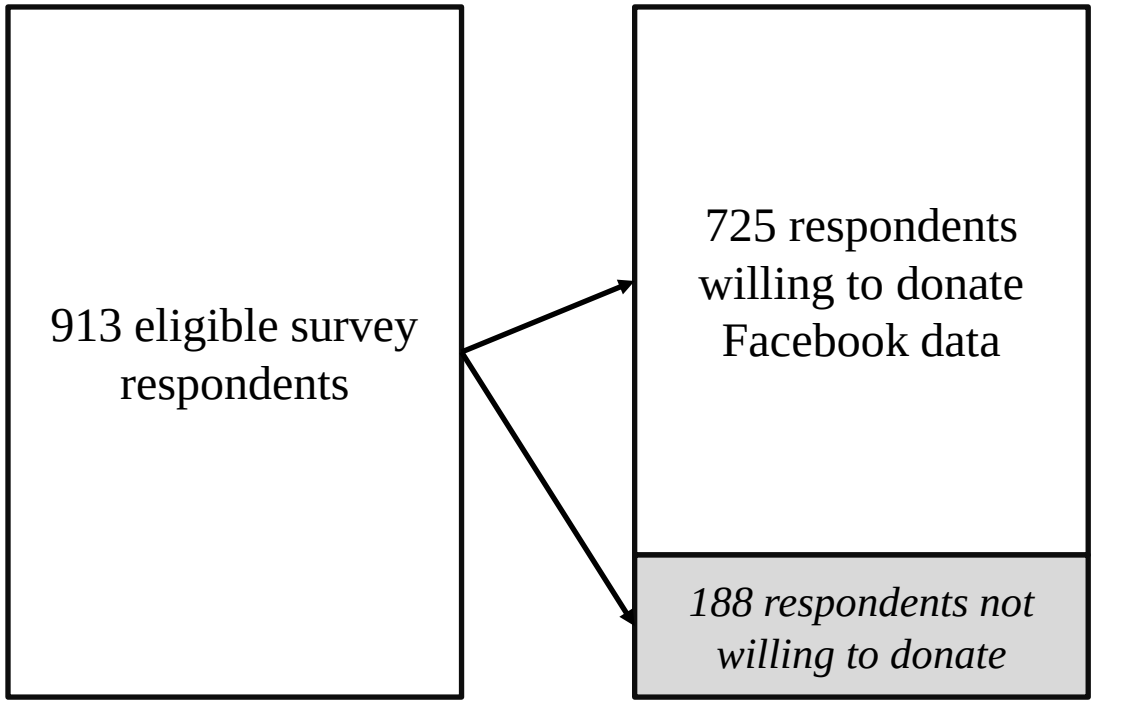


https://datadonation.eu/

Challenges of Data Donation Studies

Willingness to Donate Data

- Compared to request to download meters and apps, users report relativly high willingness to donate data (30-80%)
  (Hase & Haim 2024; Keusch et al. 2024; Ohme et al. 2021; Pfiffner and Friemel 2023; Silber et al. 2022; van Driel et al. 2022)



Willigness to Donate: 79%

Keusch et al. (2024)

Privacy Concerns

- Compared to request to download meters and apps, users report relativly high willingness to donate data (30-80%)
  (Hase & Haim 2024; Keusch et al. 2024; Ohme et al. 2021; Pfiffner and Friemel 2023; Silber et al. 2022; van Driel et al. 2022)
  - Main reason for not being willing: <u>privacy concerns</u> (Keusch et al. 2024)

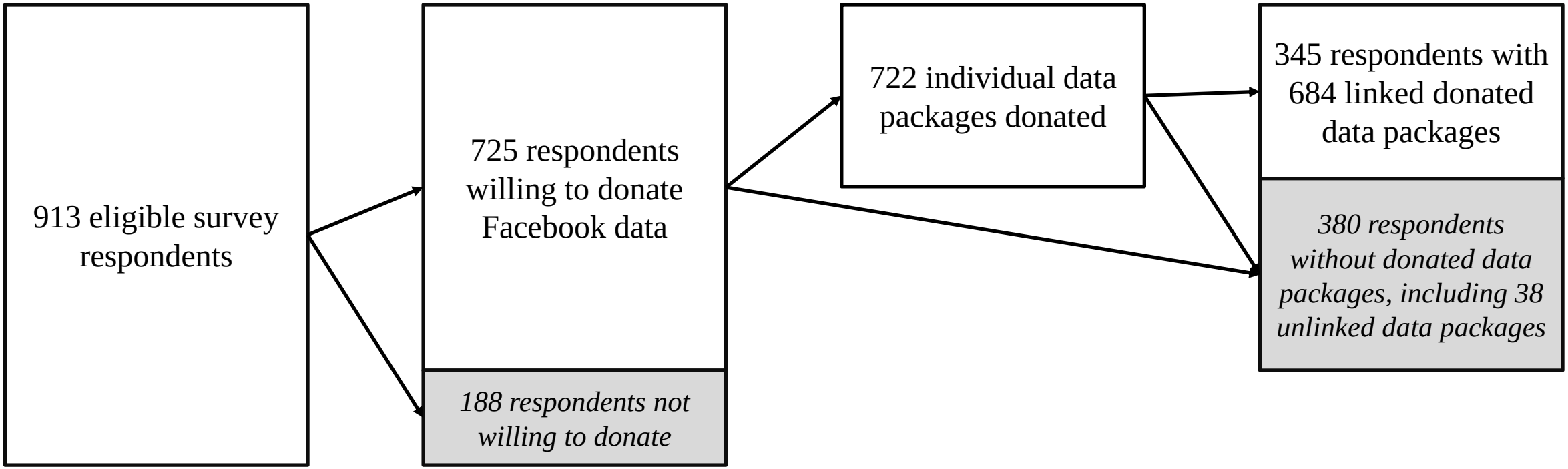
Table D1. Coded answers to open-ended question on reasons for not being willing to donate Facebook data.

Code	Reason	n	%[1]
1	*Need to protect privacy*	*34*	24.2
1.1	Generally not willing to provide data	(32)	
1.2	Fear that personal data are never protected anyways	(2)	
2	*Concern about data misuse/data breach*	*29*	20.7
2.1	General data protection concerns	(17)	
2.2	Missing information on how data will be protected	(9)	
2.3	Worries about third party access to data	(3)	
3	*Anticipating difficulties with data donation*	*19*	13.6
3.1	Anticipated technical problems	(1)	
3.2	Too complicated	(4)	
3.3	Too cumbersome	(8)	
3.4	Other practical issues	(6)	
4	*Perceived usefulness and potential personal consequences of data donation*	*13*	9.3
4.1	Do not use Facebook often enough	(10)	
4.2	Study goal seems not useful	(2)	
4.3	Concern about some personal disadvantage	(1)	
5	*No trust in researchers/university*	*5*	3.6
6	*Not enough information*	*2*	1.4
7	*No reasons provided*	17	12.1
7.1	No reason/DK	(12)	
7.2	No useful/sensical response	(5)	
8	*Incentive too small*	5	3.6
9	*Concerns related to Facebook*	*9*	6.4
10	*Unconcrete refusal*	*23*	16.4
10.1	Don't want to	(14)	
10.2	General concern	(5)	
10.3	Uncertainty	(4)	

Notes: n=140
[1]Each answer could be coded in multiple categories, thus percentages can add up to more than 100.

Drop-out During the Donation

- Large share (up to 50%) drop out during data donation process
  (Hase & Haim 2024; Keusch et al. 2024; Ohme et al. 2021; Silber et al. 2022; van Driel et al. 2022)

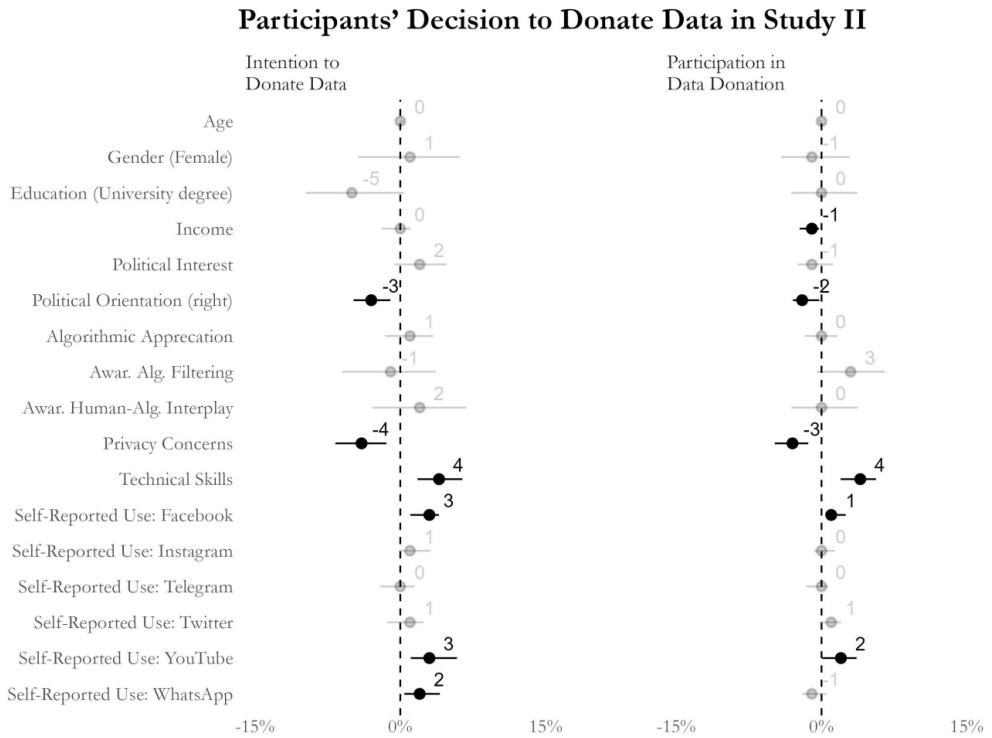


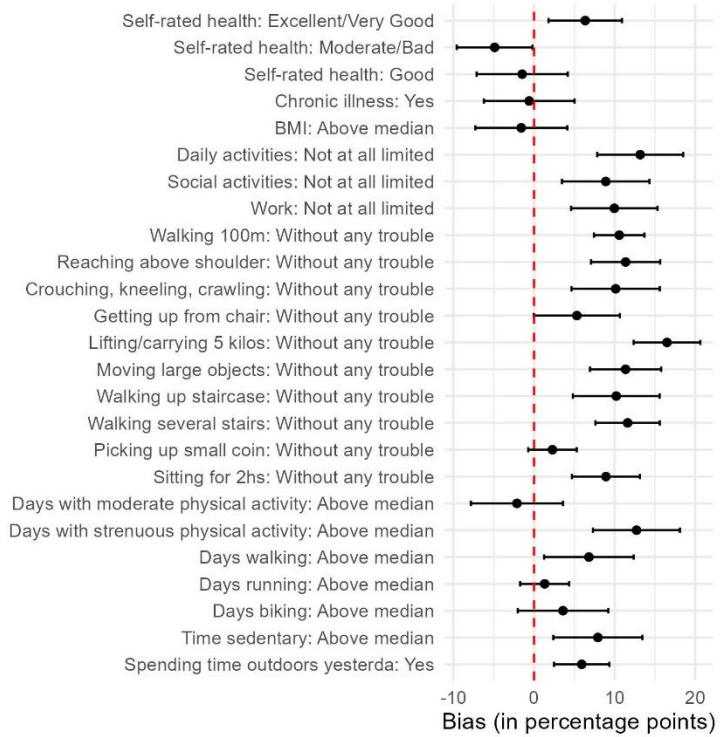
Keusch et al. (2024)

➢ Rather cumbersome process that requires some technical skills

Bias

- So far, few studies have assessed nonparticipation bias
  - Some evidence for higher social media use (Hase & Haim 2024; but not Keusch et al. 2024), political interest (Hase & Haim 2024), and "healthy donor" bias (Keusch & Struminskaya in preparation)



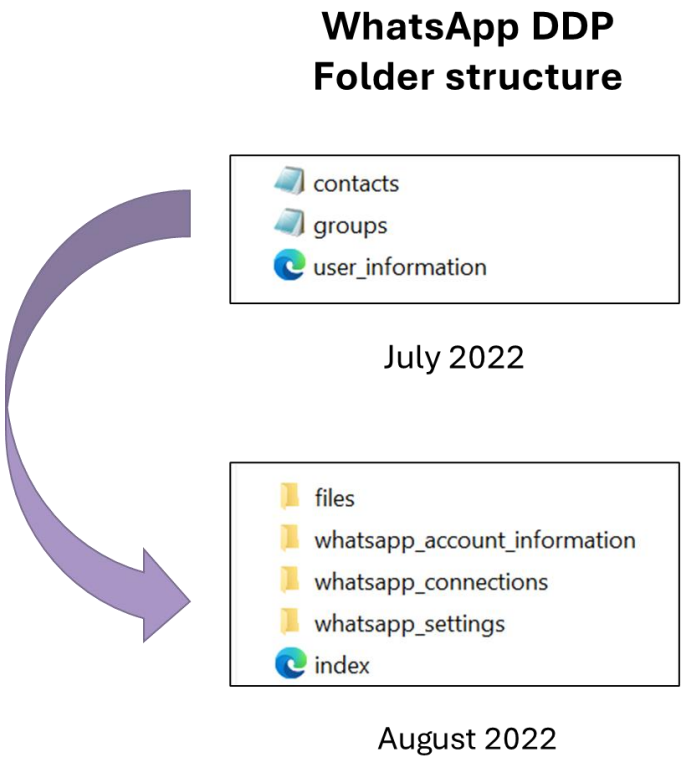


Hase & Haim (2024)

Keusch & Struminskaya (in preparation)

Volatility of DDPs

- Platforms/data controllers keep changing structure of and access to DDPs



Example provided by Thijs Carriere

➢ Constant monitoring (by researchers) and/or standardization (by legislators) (Hase et al. 2024)

Open Methodological Questions

- How to best communicate idea of data donation to participants
  - e.g., stressing data privacy advantages for participants, showing examples of data
- How to effectively integrate data donation into survey infrastructure
  - e.g., when to ask for donation, how to frame, how to incentivize
- How to design efficient data donation process for participants
  - Reduce friction of downloading, processing, and uploading data
- How to unpacking full potential of digital behavioral data
  - In combination with self-reports, e.g., measurement error models (Cernat et al. 2024)

Summary & Outlook

Summary & Outlook

- In post-API age, data donation is viable alternative for collecting individual-level digital behavioral data
  - Provides access to data not available otherwise
  - Gives users agency in deciding exactly what data to be shared

- Limitations (still) exist around willingness and privacy concerns, UX barriers, and potential bias

➢ Opens avenues to address further research questions
  - ➢ Methodological: e.g., reducing nonparticipation bias, measurement properties
  - ➢ Substantive: e.g., recommendation algorithms, close social relationships

Thank You!

Florian Keusch
University of Mannheim
School of Social Sciences
Social Data Science & Methodology
f.keusch@uni-mannheim.de
floriankeusch.weebly.com
@floriankeusch.bsky.social

References

Andrews, Sally, David A. Ellis, Heather Shaw, and Lukasz Piwek. 2015. "Beyond self-report: Tools to compare estimated and real-world smartphone use." *PLOS ONE* 10(10):e0139004. https://doi.org/10.1371/journal.pone.0139004

Araujo, Theo, Anke Wonneberger, Peter Neijens, and Claes de Vreese. 2017. "How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use." *Communication Methods and Measures*, 11:173-90. https://doi.org/10.1080/19312458.2017.1317337

Boase, Jeffrey, and Rich Ling. 2013. "Measuring mobile phone use: Self-report versus log data." *Journal of Computer-Mediated Communication* 18:508-19. https://doi.org/10.1111/jcc4.12021

Boeschoten, Laura, Niek C. de Schipper, Adriënne M. Mendrik, Emiel van der Veen, Bella Struminskaya, Heleen Janssen, and Theo Araujo. 2023. "Port: A software tool for digital data donation." *Journal of Open Source Software* 8(90):5596. https://doi.org/10.21105/joss.05596

Bosch, Oriol J., Patrick Sturgis, Jouni Kuha, and Melanie Revilla. 2024. "Uncovering digital trace data biases: Tracking undercoverage in web tracking data." *Communication Methods and Measures*. https://doi.org/10.1080/19312458.2024.2393165

Carrière, Thijs C., Laura Boeschoten, Bella Struminskaya, Heleen L. Janssen, Niek C. de Schipper, and Theo Araujo. 2024. "Best practices for studies using digital data donation." *Quality & Quantity*. https://doi.org/10.1007/s11135-024-01983-x

Cernat, Alexandru, Florian Keusch, Ruben L. Bach, and Paulina K. Pankowska. 2024. "Estimating measurement quality in digital trace data and surveys using the MultiTrait MultiMethod model." *Social Science Computer Review*. https://doi.org/10.1177/08944393241254464

Clemm von Hohenberg, Bernhard, Sebastian Stier, Ana S. Cardenal, Andrew Guess, Ericka Menchen-Trevino, and Magdalena Wojcieszak. 2024. "Analysis of web browsing data: A guide." *Social Science Computer Review*. https://doi.org/10.1177/08944393241227868

de Reuver, Mark, and Harry Bouwman. 2015. "Dealing with self-report bias in mobile Internet acceptance and usage studies." *Information & Management* 52:287-94. https://doi.org/10.1016/j.im.2014.12.002

Haenschen, Katherine. 2020. "Self-reported versus digitally recorded: Measuring political activity on Facebook." *Social Science Computer Review* 38:567-83. https://doi.org/10.1177/0894439318813586

Haim, Mario, Dominik Leiner, and Valerie Hase. 2023. "Integrating data donations into online surveys." *Medien & Kommunikationswissenschaft* 71(1-2):130-7. https://doi.org/10.5771/1615-634X-2023-1-2-130

Hase, Valerie, Florian Keusch, Frauke Kreuter, and Mark Trappmann. 2023. "Integrating data donation in survey infrastructure: Quantifying, explaining, and addressing errors in representation and measurement." DFG Proposal.

Hase, Valerie and Mario Haim. 2024. "Can we get rid of bias? Mitigating systematic error in data donation studies through survey design strategies." *Computational Communication Research* 6(2). https://doi.org/10.5117/CCR2024.2.2.HASE

Hase, Valerie, Jef Ausloos, Laura Boeschoten, Nico Pfiffner, Heleen Janssen, Theo Araujo, Thijs Carrière, Claes de Vreese, Jörg Haßler, Felicia Loecherbach, Zoltán Kmetty, Judith Möller, Jakob Ohme, Elisabeth Schmidbauer, Bella Struminskaya, Damian Trilling, Kasper Welbers, and Mario Haim. 2024. "Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies?" *Internet Policy Review* 13(3). https://doi.org/10.14763/2024.3.1793

Hinds, Joanne, and Adam N. Joinson. 2018. "What demographic attributes do our digital footprints reveal? A systematic review." *PLOS ONE* 13(11):e0207112. https://doi.org/10.1371/journal.pone.0207112

Howison, James, Andrea Wiggins, and Kevin Crowston. 2011. "Validity issues in the use of social network analysis with digital trace data." *Journal of the Association for Information Systems* 12:767-97. https://doi.org/10.17705/1jais.00282

Jones-Jang, S Mo, Yu-Jin Heo, Robert McKeever, Jung-Hyun Kim, Leigh Moscowitz, and David Moscowitz. 2020. "Good news! Communication findings may be underestimated: Comparing effect sizes with self-reported and logged smartphone use data." *Journal of Computer-Mediated Communication* 25:346-63. https://doi.org/10.1093/jcmc/zmaa009

Keusch, Florian, Ruben L. Bach, and Alexandru Cernat. 2023. "Reactivity in measuring sensitive online behavior." *Internet Research* 33(3):1031-52. https://doi.org/10.1108/INTR-01-2021-0053

Keusch, Florian, Sebastian Bähr, Georg-Christoph Haas, Frauke Kreuter, Mark Trappmann, and Stephanie Eckman. 2022. "Non-participation in smartphone data collection using research apps." *Journal of the Royal Statistical Society. Series A* 185:S225-45. https://doi.org/10.1111/rssa.12827

Keusch, Florian, and Frauke Kreuter. 2022. "Digital trace data. Modes of data collection, applications, and errors at a glance." In *Handbook of Computational Social Science. Volume 1: Theory, Case Studies and Ethics*, edited by Uwe Engel et al., 100-118. Milton Park: Routledge. https://doi.org/10.4324/9781003024583-8

Keusch, Florian, Paulina Pankowska, Alexandru Cernat, and Ruben Luksa Bach. 2024. "Do you have two minutes to talk about your data? Willingness to participate and nonparticipation bias in Facebook data donation. *Field Methods* 36(4):279-93. https://doi.org/10.1177/1525822X231225907

Keusch, Florian, & Bella Struminskaya. in preparation. "Measuring physical activity in older adults through data donation: Consent rates, donation success, & bias."

Keusch, Florian, Bella Struminskaya, Christopher Antoun, Mick P. Couper, and Frauke Kreuter. 2019. "Willingness to participate in passive mobile data collection." *Public Opinion Quarterly* 83:210-35. https://doi.org/10.1093/poq/nfz007

Kobayashi, Tetsuro, Jeffrey Boase, Tsutomu Suzuki, and Takahisa Suzuki. 2015. "Emerging from the cocoon? Revisiting the tele-cocooning hypothesis in the smartphone era." *Journal of Computer-Mediated Communication* 20:330-45. https://doi.org/10.1111/jcc4.12116

Kohne, Julian, and Christian Montag. 2024. "ChatDashboard: A Framework to collect, link, and process donated WhatsApp chat log data." *Behavioral Research Methods* 56:3658-84. https://doi.org/10.3758/s13428-023-02276-1

Ohme, Jakob, Theo Araujo, Claas de Vreese, and Jessica T. Piotrowski. 2021. "Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function." *Mobile Media & Communication*, 9:293-313. https://doi.org/10.1177/2050157920959106

Pankowska, P., Alexandru Cernat, Florian Keusch, and Ruben L. Bach. (under review). "Using hidden Markov models to assess and correct for measurement error in digital trace data."

Pfiffner, Nico, and Thomas N. Friemel. 2023. "Leveraging data donations for communication research: Exploring drivers behind the willingness to donate." *Communication Methods and Measures* 17:227-49. https://doi.org/10.1080/19312458.2023.2176474

Revilla, Melanie, Carlos Ochoa, and Germán Loewe. 2017. "Using passive data from a meter to complement survey data in order to study online behavior." *Social Science Computer Review* 35:521-36. https://doi.org/10.1177/0894439316638457

Scharkow, Michael. 2016. "The accuracy of self-reported internet use—a validation study using client log data." *Communication Methods and Measures* 10:13-27. https://doi.org/10.1080/19312458.2015.1118446

Scherpenzeel, Annette. 2017. "Mixing online panel data collection with innovative methods." In *Methodische Probleme von Mixed-Mode-Ansätzen in Der Umfrageforschung*, edited by Stefanie Eifler and Frank Faulbaum, 27-49. Schriftenreihe Der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15834-7_2

Silber, Henning, Johannes Breuer, Christoph Beuthner, Tobias Gummer, Florian Keusch, Pascal Siegers, Sebastian Stier, and Bernd Weiß. 2022. "Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behaviour." *Journal of the Royal Statistical Society Series A: Statistics in Society* 185:S387-S407. https://doi.org/10.1111/rssa.12954

Stopher, Peter, Camden FitzGerald, and Min Xu. 2007. "Assessing the accuracy of the Sydney Household Travel Survey with GPS." *Transportation* 34:723-41. https://doi.org/10.1007/s11116-007-9126-8

van Driel, Irene I., Anastasia Giachanou, J. Loes Pouwels, Laura Boeschoten, Ine Beyens, and Patti M. Valkenburg. 2022. "Promises and pitfalls of social media data donations." *Communication Methods and Measures* 16:266-82. https://doi.org/10.1080/19312458.2022.2109608