

Extended Program: Data Donation Symposium

Day 1 – Thursday, 30 May 2024

12:00-12:45 Walk-in and registrations

12:45 Opening

12:55 Keynote (Chair: Theo Araujo)

Data4Science: Agency, Transparency and Reciprocity in Data Donation

Melinda Mills

University of Oxford

Individuals are increasingly aware of the large amounts of personal data collected from them by when they use services and want more agency in the use and understanding of their own personal data. At the same time, this valuable personal data that could generate fundamental discoveries to benefit the public good is often inaccessible for scientific researchers. They lack agency, transparency and reciprocity. Leveraging recent GDPR regulations of the right to data portability, Data4Science (D4S) is a trusted platform for individuals to take control of their own data, donate, understand it, shape their own behaviour and share it for broader health and behavioural research. Donated data includes data from social media, professional sites, wearables, internet searches and watches, mobility, genetics to microbiome and beyond.

13:55 Coffee Break

14:15 Session: Challenges on Participant Engagement (Chair: Thijs Carriere)

Increasing Data Donation Compliance: Exploring New Strategies and Characteristics

Ernesto De León, Fabio Votta, Theo Araujo, Claes de Vreese, Joris Mulder, Stein Jongerius*

**Amsterdam School of Communication Research, University of Amsterdam*

Addressing reluctance to share data donation packages (DDP) is crucial. Past studies have focused on understanding who refuses to share their DDP with researchers. Generally, there is broad consensus that DDP sharing is hampered by technological efficacy perceptions, lack of familiarity with DDP, and identifying with the political right.

We propose that increasing responsiveness among these groups would greatly help the external reliability of studies using DDP. We do this in two ways. First through an experiment where we hope to reduce non-participation among these specific populations. We will design and test three framings of the study. First, the “Anti Big Tech” frame aims to encourage compliance by emphasizing data donation as a response to intrusive practices by major corporations. Second, the “Social Cues” Frame

draws from social norms literature to normalize the data donation process by showcasing the positive experiences of past donors and emphasizing that many people have previously done the same. Third, the “Information” frame addresses the lack of awareness by employing a step-by-step infographic to demystify the data donation process.

Nevertheless, many of the studies on DDP compliance have been conducted on samples that are limited by their size and representability, as well as the information available on participants. We would also embark on a broader exploration of the characteristics associated with DDP non-compliance. To conduct this study, we are partnering with Centerdata and their LISS Panel of approximately 6,500 participants in the Netherlands. The LISS panel has extensive existing information on these participants, a lot of which have not been used to study DDP compliance. With such a large sample size and the availability of individual-level characteristics, we hope to provide one of the most nuanced and comprehensive portraits of DDP noncompliance to date.

This study would, therefore, both test treatments to existing biases, as well as uncover additional characteristics associated with non-response to DDP.

Why Do People Self-Select Out of Data Donation Studies? Cross-National Insights from Germany and the Netherlands

Hase, V, Struminskaya, B., Araujo, T., Boeschoten, L., Ozornina, N., Lechner, M., & Haim, M.*

**Department of Media and Communication, LMU Munich*

Data donation studies have gained traction in the social sciences. However, we know that errors in representations – i.e., people not being able or willing to participate in data donation studies – may bias related results. Similar to surveys (Keusch, 2015), reasons for drop-out in data donation studies likely not only vary across study contexts, for example, the type of platforms for which researchers request data, but also national contexts (Kmetty et al., 2023). Given the lack of comparative perspectives in hitherto research on digital trace data, our study analyzes cross-national differences in reasons participants mention for their stated non-willingness to donate data (RQ1) as well as predictors of stated and actual willingness to donate their data (RQ2). To describe and explain drop-out, we draw on quantitative insights from two recent data donation studies in Germany (N = 2,039) and the Netherlands (N = 1,163). Moreover, we conducted a manual, quantitative content analysis of open responses participants provided after deciding against donating their data (N = 898, lowest $\alpha = .75$). First results indicate cross-national differences in reasons due to which participants dropped out (RQ1): According to their open responses, German participants were less concerned about their privacy (40% in Germany, 53% in the Netherlands). However, they more often considered their data “irrelevant” due to their infrequent use of platforms for which they were asked to donate data (26% in Germany, 4% in the Netherlands). Similarly, predictors of drop-out varied across national contexts (RQ2). Overall, our study sheds light on the need for cross-national research on errors in representation in data donation studies.

Towards Co-Created and Collaborative Data Donation Projects

Alejandra Gomez Ortega, Jacky Bourgeois, Gerd Kortuem*

**Industrial Design Engineering Faculty, Delft University of Technology*

Data donation is an emerging practice enabling personal digital-trace data collection for research. While it offers opportunities to access new insights into people's behavior and experiences through their digital-trace data, the role of individuals -- as research participants -- is limited in most data donation projects. They primarily contribute with data, limiting the perspectives included and accounted for around critical research-design decisions. Through three empirical studies in designerly contexts, we have explored the challenges and opportunities that come with embedding data donation in research processes that are not only contributory but collaborative and co-created. In this presentation, we provide insights into people's preferences and expectations in participatory data

donation processes and discuss considerations for supporting various degrees of participation in future data donation research.

Recruiting Data Donors through Social Media Advertisements – Opportunities and Caveats

*Nico Pfiffner, Lukas Tribelhorn & Thomas N. Friemel**

**Department of Communication and Media Research, University of Zurich*

Introduction

In this presentation, we report our experiences with recruiting data donors through social media advertisements on Instagram. This is related to a study where we seek young people between the age of 16 and 25 ($n_{\text{target}}=200$) to donate their YouTube and TikTok data. The data collection is currently ongoing but will be concluded by the time of the Symposium. It consists of three parts:

- Part I: Screening survey, YouTube data donation, and survey in relation to YouTube usage.
- Part II: Inviting a friend to also take part (only if they were not invited themselves).
- Part III: TikTok data donation and survey in relation to TikTok usage.

Participation is rewarded with a 10 CHF voucher for each completed part and a personalized YouTube usage report.

Advertisements

To recruit participants, we use online advertisements on Instagram both on mobile and on desktop devices. This approach was chosen because the study focuses on young people – a demographic that is very active on Instagram and hard to reach through other channels (e.g., panel providers).

Because it is still unclear what motivates people most to participate in data donation studies, we employ three ad versions, each highlighting a different motivational aspect: (A) receiving a personal report, (B) (altruistically) contributing to research, or (C) the monetary incentive.

Results and Learnings

Because data collection is still ongoing, we cannot yet discuss the outcome in terms of donation rates. Nevertheless, our experience so far suggests that recruiting data donors through social media ads is a promising approach that comes with some caveats: We experienced issues with ad payments, there are limited opportunities to test the ad before activation, and Instagram opens ad-clicks in their in-app browser which comes with limited capabilities possibly affecting the data collection (e.g., one cannot open multiple tabs and sometimes URL parameters are omitted).

Once in a lifetime - Pros and cons of a full-scale DDP approach

Zoltán Kmetty, Árpád Knap, Bendegúz Váradi*

**CSS-RECENS, HUN-REN Centre for Social Sciences*

Although several data donation-based projects have been launched in recent years, relatively few studies have been completed that have collected DDP-based data from multiple platforms in a large sample. In 2022, the HUN-REN Social Science Research Centre initiated an extensive data collection endeavor to obtain digital and survey data from a representative sample of social media users. The data collection part of the research started in February 2023 and ended in June 2023. Participants were requested to share their Facebook and Google data, encompassing geolocation, searches, and YouTube usage, in exchange for remuneration. Simultaneously, respondents were required to complete a 40-minute survey. Optionally, participants were incentivized to share additional data from Instagram, TikTok, and Twitter for an extra reward. Ultimately, the study incorporated 753 respondents who completed the survey and uploaded their Facebook data and at least one type of Google data.

Subsequently, the collected data underwent a two-step process involving anonymization and normalization. While this method demands a more intricate privacy-focused approach, it offers a

unique opportunity to scrutinize anomalies within data donation-based sources and assess data loss resulting from anonymization and normalization procedures.

In my forthcoming presentation, I will delineate the values and drawbacks of the applied research design, highlighting key insights gleaned from the data pre-processing phase. Additionally, I will explore the potential advantages of a multi-platform data donation project when contrasted with singular-platform research.

15:15 A demonstration of the data donation software *Port*

15:30 Coffee Break

16:00 High-Density Session: Study Designs (Chair: Bella Struminskaya)

A feasibility pilot of data donation among a nationally representative panel of survey respondents

Sherry Emery, Hy Tran*

**Public Health, NORC at the University of Chicago*

Background: Acquiring social media (SM) data is increasingly challenging, expensive, and subject to the changing whims of the platforms and vendors. The goal of our pilot is to determine the feasibility of soliciting a nationally representative panel of survey participants to donate their social media data to NORC.

Methods: We will ask the 1,000+ adult participants from the Wave 2 March 2024 AmeriSpeak Omnibus panel to opt into our pilot study. Every participant will be shown a consent video within the survey environment, which clearly describes that we are asking them to download their own data from Facebook and/or Instagram, and then upload it to AmeriSpeak's secure portal. Those that opt-in will receive more detailed video and written instructions, along with additional incentive compensation. We will ask respondents approximately 10 questions on the Omnibus, including items probing respondent's social media use, attitudes and behaviors related to the influence of social media on society and individuals, and reasons for opting out vs. opting in.

Expected Yield: In our prior work in 2020, 75% of study participants consented to provide their Twitter account handle to NORC researchers. Since more burden is placed on survey respondents to access, request, download, and upload their complete Facebook and/or Instagram data history, our goal for participation is 25% (i.e. 250 individuals opt-in to donate their data and follow through with the upload to NORC).

What We Will Learn: This pilot will allow us to assess the feasibility of collecting survey respondent's social media data to create user-centric measures of social media activity to answer questions about the influence of social media on individual opinions and behaviors. The pilot will also enable us to explore participant characteristics associated with consent to participate, understand reasons for deciding for or against donation, and benchmark participation with these parameters.

Negotiated consent for data donation in clinical research

Zahra Ghasia^{}, Dr. ir. Jacky Bourgeois*

^{}TU Delft*

The use of Electronic Health Records (EHRs) has seen a breakthrough in clinical research for personalized treatments (Hamburg & Collins, 2010.) Despite the potential advantages of vast EHR data available, constraints of privacy and legislation hinder its use (Rieke et al., 2020.) Health data exists in an interconnected healthcare system (Harris et al., 2009), comprising of stakeholders responsible for health data management within the constraints of GDPR. Data sharing platforms, using secure data sharing practices and encrypted technology, can potentially change the landscape of health data donation offered for clinical research (Obermeyer & Emanuel, 2016.)

This project examines the privacy and stakeholder landscape of health data sharing through an evaluation of interaction with patients. It uses a Value-Sensitive design approach (Friedman et al., 2002) to contextualize the privacy values in clinical research for patients, doctors, patient organizations, clinical researchers, pharmaceutical industries, and data sharing platforms. Through this examination, it identifies lack of transparent data use and research practices as a hindrance to the motivation of data donation on a data sharing platform. Transparency is examined through a patient-centric lens for motivation to donate data, wherein information and control over preferences of participation are found to be crucial. This evaluation further identifies roles of researchers and data sharing platforms for a transparent approach.

The project results in a concept termed ‘Negotiated Consent’, which examines the use of data sharing platforms in offering individuals transparency. This is contextualized within patient consent for recruitment in clinical research, wherein data sharing platforms are responsible for data processing using Federated Learning and Natural Language Processing. The results are tested through a study of the prototyped user experience of ‘Negotiated Consent’, wherein participants emphasized ‘feeling more informed and in control’ in comparison to the current informed consent.

The concept of ‘Negotiated Consent’ adds to the literature on attitudes towards data donation for research (Richter et al., 2021), wherein individuals were found to request control of use and protection of health data by independent bodies in case of commercial use. Through the offering of ‘Negotiated Contract’, the data sharing platform can become the independent body, whilst following regulation norms. Moreover, in consideration of the motivation to donate, factors such as concern for others and self-benefit were identified under a study by Skatova and Goulding (2019.) Herein, ‘Negotiated Consent’ is a means of motivating individuals to benefit from participation through their concerns for both-the other and themselves. Whilst the study examines a patient-centric approach towards health data donation for clinical research, it has limitations in addressing the multi-faceted reality of patient’s lives that informs their motivations to enrol in clinical research. The development of patient-centricity in the domain requires further examination of motivations through a lens of disease-specific patient groups, demographics, and personal history with diseases (Hong et al., 2020), which may lead to richer insights.

This project contributes to an understanding of the motivations of individuals in a privacy and stakeholder dynamics informed manner for health data motivation. It illuminates a transparent approach to recruitment for clinical research through a collaborative effort that emphasizes on patient-centric approaches. Using a mutually responsible approach, data sharing platforms can lead the way for data donation in clinical research that simultaneously empowers patients to control their data through a negotiated lens.

Mapping online exposure to gambling marketing communication for vulnerable groups: a data donation study

Lauranna Teunissen, Priska Breves, Merel Walraven, Eva van Reijmersdal, Hilde Voorveld*

**Department of Communication Science, Amsterdam School of Communication Research*

Since online gambling became legalized in the Netherlands in 2021, marketing communication has become extremely ubiquitous. As a result, the government banned untargeted advertising since July 1st, 2023. Given the possibilities for online advertising and sponsorship that remain permitted and the way gambling companies operated on the cutting edge of what was legally possible in the past years, this ZonMW-funded project gives insight into the quantity and content of gambling marketing communication that specific vulnerable groups are exposed to.

We will use social media data donation to gain insights into the exposure to and the type of tactics used in gambling marketing communication on specific social media platforms. With this method, we can circumvent social desirability or attentional biases, which are common in this field of research. A group of 30 participants from several vulnerable groups, varying in age (between 13 and 35 years old), gambling intensity, and SES will be asked to donate their data from a specific social media platform (TikTok, Instagram or YouTube, to be determined based on a co-creation workshop with experience experts). Because we invite participants to the lab, we can help them with the protocol. We will use the infrastructure that is set up within ASCoR. Donated data will be coded by two independent coders. We will code several aspects such as the number of gambling-related social media ads participants have been exposed to, the number of advertisers interacted with, the advertisers that have uploaded contact information, the breadth and depth of profiling, the appeals used (free spins, bonuses), the portrayal of people, interactive elements.

This project is part of a bigger project. The data collection is planned for summer 2024, so at the symposium we are able to present a more detailed study design and protocol and would appreciate feedback from participants.

Measuring the effect of a seven-day Instagram abstinence on users' body image by combining survey and donated data

Daria Szafran, Ruben Bach, Frieder Rodewald, Florian Keusch*

**School of Social Sciences, Professorship for Social Data Science and Methodology, University of Mannheim*

Previous research has found that the use of social media, including Instagram, might lead to adverse body image outcomes such as body dissatisfaction. So far, there are only a few experimental studies dedicated to this topic. Although providing important insights, they have several limitations: on the one hand, they are based on small and selective samples, which hinders the findings' generalizability. On the other hand, they measure Instagram use via self-reports, which might introduce measurement errors due to social desirability or recall bias.

We aim to address the existing research gaps by conducting a two-wave survey using a German probability-based sample of male and female Instagram users to measure how a seven-day abstinence from Instagram use will affect their body image. In addition to providing detailed information on their Instagram use, respondents will be asked to donate their Instagram data. Our study has two main objectives: first, we contribute to existing body image research by experimentally measuring whether abstaining from using Instagram has an effect on users' body image. For this purpose, we use the donated data to control whether respondents actually complied with the treatment. Second, we assess whether and to what extent the respondents' self-reports match their donated data. In the long-term, our findings contribute to the development of robust instruments for measuring digital behavior. Such instruments are crucial, e.g., for the design of interventions aimed at preventing the adverse effects of social media on body image and well-being.

Investigating Rabbit Holes in Social Media Usage: A Novel Perspective Using Data Donation

Qiru Huo, Ine Beyens, & Nadia Bij de Vaate*

**Communication Science, Amsterdam School of Communication Research (ASCoR)*

In recent years, there has been growing concern about whether personalization algorithms on social media lead users into 'rabbit holes', a phenomenon whereby personalization algorithms push users into a never-ending exploration of extremely engrossing and one-sided content. On the one hand, rabbit holes may occur when individuals are trapped in so-called algorithmic 'filter bubbles' of increasingly extreme or problematic content, thereby adversely affecting their ideological perspectives (Tang et al., 2021). On the other hand, rabbit holes may occur when individuals are drawn to social media content so that their usage is prolonged uncontrollably, thereby impacting their digital well-being (Vanden Abeele, 2021). At present, empirical findings concerning the existence of rabbit holes are inconsistent, with some studies supporting and others refuting their existence (Schellingerhout et al., 2023).

To further the debate on the existence of rabbit holes, the presentation will discuss how to detect rabbit holes in users' social media engagement using data donations. Current research on detecting rabbit holes employs various approaches to collect users' browsing history and recommendations, such as employing bots for simulating user interactions and automatically collecting browsing patterns and recommendations, requesting for user self-reports, and leveraging additional data tracking tools. In comparison to the methods above, data donation presents several advantages. It is 'user-centric', which means the study relies on authentic digital traces of users collected in cooperation with the users themselves, ensuring the ecological validity and enabling the implementation of self-reported designs in conjunction with the donated data (Ohme et al., 2023). Moreover, in comparison to other data tracking tools that are confined to a single device or a limited set of data types, the data donation method offers the convenience and broad applicability of monitoring all data associated with an account, across all devices, providing richer dimensions of information (Breuer et al., 2023).

References

- Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2023). User-centric approaches for collecting Facebook data in the 'post-API age': Experiences from two studies and recommendations for future research. *Information, Communication & Society*, 26(14), 2649–2668. <https://doi.org/10.1080/1369118X.2022.2097015>
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2023). Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *Communication Methods and Measures*, 1–18. <https://doi.org/10.1080/19312458.2023.2181319>

Schellingerhout, R., Beraldo, D., & Marx, M. (2023). Accounting for Personalization in Personalization Algorithms: YouTube's Treatment of Conspiracy Content. *Digital Journalism*, 1–29. <https://doi.org/10.1080/21670811.2023.2209153>

Tang, L., Fujimoto, K., Amith, M. (Tuan), Cunningham, R., Costantini, R. A., York, F., Xiong, G., Boom, J. A., & Tao, C. (2021). “Down the Rabbit Hole” of Vaccine Misinformation on YouTube: Network Exposure Study. *Journal of Medical Internet Research*, 23(1), e23262. <https://doi.org/10.2196/23262>

Vanden Abeele, M. M. P. (2021). Digital Wellbeing as a Dynamic Construct. *Communication Theory*, 31(4), 932–955. <https://doi.org/10.1093/ct/qtaa024>

Two become one: A collaborative data donation project using Netflix viewing data

Joël Hendrix, Gijsje Maas*, Rebecca de Leeuw, Anita Eerland, Gijs Bijlstra, & Marieke Fransen*

**Behavioural Science Institute, Radboud University*

This data donation project is a combination of two separate, research projects in communication science. Although these projects focus on somewhat related research themes, their main similarity lies in the methods. The latter was, therefore, the primary reason for us to collaborate. Both projects study viewing behaviour, a variable which has so far mostly been measured using self-report measures in communication science. However, as self-reported exposure measures are subject to scrutiny due to their unreliable nature, we hope to find a more ecologically valid measure in data donation. Specifically, we will ask participants to donate their Netflix streaming data to study their viewing behaviour. This data will be combined with cross-sectional data (Project 1 and 2) and subjected to quantitative content analysis (Project 2) to assess our individual research questions.

The first project (Hendrix et al.) aims to identify the relationships between viewing behaviour and openness towards the LGBTQ+ community and underrepresented groups in general, and mechanisms potentially underlying these relationships. The second project (Maas et al.) will investigate the relationship between individuals' viewing behaviour and their perceived gender norms about sex, specifically the sexual double standard (SDS). Subsequently, a quantitative content analysis will be conducted to ascertain the occurrence of portrayals of SDS in each participant's individual media diet. In turn, this will be linked to perceived norms to establish what role media might play in the representation and continuation of norms like SDS.

Ethical Considerations in Exploring Student Interactions with ChatGPT: A Pilot Study in the Netherlands

Karin van Es, Dennis Nguyen*

**Utrecht University*

The advent of Generative AI (GenAI) technologies like ChatGPT has stirred considerable debate within educational settings, challenging traditional pedagogies and raising ethical concerns. What is missing in the cycles of hype and fear about GenAI in education is empirical research into the situation on the ground. How students use it in different contexts has only been tentatively explored, mostly through surveys. Limited research is available on how they perceive and harness GenAI services such as ChatGPT for study-related purposes. Researching student interactions through their actual user data collected by ChatGPT/OpenAI opens new paths for better understanding how GenAI entered higher education from a student perspective.

However, while preparing this pilot study, we found that soliciting students to donate their ChatGPT data presents unique ethical challenges. For example, there is lingering potential for identification of individual students through their data even when direct identifiers are removed.

Furthermore, the possibility of detecting academic dishonesty/misconduct in the donated data poses another ethical dilemma regarding our responsibility in such discoveries. This presentation will

briefly reflect on these challenges, the discussions with privacy officers and the ethics review board, and steps we intend to take to mitigate against potential harm.

Unlocking Insights into Adolescent Peer Relationships Through Social Media Data Donations

Loes Pouwels

Behavioural Science Institute, Radboud University

In the digital age, adolescents engage extensively with peers through social media platforms. A pressing societal and scientific question is, therefore, how these interactions transform their peer relationships (Nesi et al., 2018). Traditional research has predominantly revolved around self-reported quantitative metrics, like time spent on social media or interaction frequency. While self-reported methods could provide valuable insights into social media effects on outcomes like social displacement, they are not detailed enough to fully understand adolescents' online peer relationships. For example, to find out whether and how social media use makes adolescents popular among their offline peers, we should be able to investigate what specific social media adolescents encounter, share, or create, and with whom they interact online.

This presentation delves into the potential of social media data donations to measure adolescents' online peer interactions. To do so, I will showcase the design of project SOCIAL, which addresses how the effects of social media use on social connectedness vary across social contexts, adolescents, and timescales. This project is innovative because it will combine data donations with experience sampling methodology and peer nominations. Social media use and social connectedness will be measured intensively during three weeks in adolescents' daily lives using experience sampling methodology. These short-term dynamics will be linked to adolescents' long-term development of social connectedness measured in three consecutive annual classroom assessments using peer nominations. To get objective insights into the types and companies of adolescents' online activities, the questionnaires will be complemented with voluntary data donations of adolescents' social media archives. Methodological challenges and best practices will be discussed based on focus groups.

Exploring Well-being in Gender and Sexual Minorities: Understanding the Buffering Role of Online Community

*Prof. Dr. Enny Das, Dr. Sarah Burke, PhD Charlie Loopuijt**

**Department of Language and Communication, Radboud University*

In recent years, societal shifts in perceptions of gender and sexuality have been accompanied by a widespread adoption of more inclusive communication practices. Yet research consistently demonstrates that gender and sexually minoritized people experience lower well-being compared to cisgender or heterosexual individuals, influenced by societal and cultural factors. While progressive voices advocate for inclusivity, more conservative factions voice concerns about cultural values and norms. Amidst this, the rise of social media has offered online spaces to build community and find a sense of belonging, a known protective factor for poor well-being among this population. The present study aims to, from a minority stress theory (MST) perspective, better understand the role of social media and online community engagement in relation to well-being among gender and sexually minoritized people. This study collaborates with the D3I consortium and with the Lifelines cohort study. D3I offers a novel approach to data collection by enabling, in legally and ethically sound ways, participants to donate specific data download packages (DDP). Lifelines cohort study is a longitudinal population cohort study in the Netherlands that assesses health and well-being broadly. We will utilize measures from the Lifelines cohort to assess well-being as well as add existing measures of MST (e.g., Everyday Discrimination Scale) and online group membership in an additional survey. We assess the online membership of minority groups via data donation and employ other minority groups as controls. Participants from the Lifelines study are invited to donate their Facebook data, where the

D3I consortium's software will facilitate the extraction of relevant variables. We hypothesize, in line with MST, that membership in online minority communities mitigates the impact of stress on well-being. Our study advances our broader understanding of factors of risk and resilience that impact the well-being of gender and sexually diverse people.

17:00-18:00 Drinks

Day 2 – Friday, 31 May 2024

9:00 **Session: Data Donation Infrastructures (Chair: Laura Boeschoten)**

The RegretsReporter Program: evolving methods to meet research challenges with crowdsourced data

Jesse McCrosky, Rebecca Ricks*

**Open Source Research and Investigations, Mozilla Foundation*

Mozilla's RegretsReporter program is now preparing its third data-donation study of social media platforms. Each study has explored pressing research questions in support of Mozilla's policy and advocacy objectives. Findings have included the ineffectiveness of YouTube user control features, measurement of rates of harmful content served by YouTube, and characterization of the audiences and surfaces disproportionately responsible for this harmful content.

Through this history, the research design has evolved, from a mixed-methods observational study, to a randomized controlled trial, and now has jumped from desktop to mobile, supporting research into the popular TikTok video sharing platform. The TikTok Report app will support data donation by mobile users on both iOS and Android.

Throughout the studies, various tools and infrastructure to support this research have been developed and open-sourced, including machine learning pipelines to calculate semantic similarity of videos and to analyze and annotate screen recordings of social media sessions. Mozilla's best practices for privacy and consent have informed the entire program of work.

Bridging the Gap: Using Data Donations to Enhance Adolescent Cohort studies

Sebastian Kurten, Amanda Ferguson, Amelia Leyland-Craggs, Amira Skeggs, Valerie Yap, Amy Orben*

**School of Clinical Medicine, University of Cambridge*

There exists an urgent need for novel methods to be developed and deployed that can both capture the complexity of the digital world and be included in existing cohort and household panel investigations, which are currently falling short in helping researchers understand the nuanced impacts of online environments on adolescent health. This challenge is getting ever more prominent as adolescents are not only spending many hours per day online, but the content they are exposed to is getting more diverse and tailored. New ways of digital data collection are needed to address this.

Existing cohort studies, such as the Millennium Cohort Study and Understanding Society, rely on self-reported screen time measures that fail to capture the diverse and intricate nature of online activities. The inadequacy of such measurement methods hinders researcher's ability to understand

how online environments impact adolescent health. Furthermore, inaccurate self-reporting of digital device usage and activities introduces unreliability into the data.

Our study tries to advance existing methods for adolescent cohort studies by investigating the feasibility of using data donation in adolescent samples. The project has three principal aims: firstly, testing the feasibility of data donation as a method for collecting digital data in future cohort studies; secondly, exploring the development of interactive visualizations to engage adolescents and provide valuable feedback on their digital data; thirdly, integrating data donation and visual interactions into school curricula to enhance the attractiveness for teachers and schools.

This innovative project seeks to lay the foundation for a more comprehensive understanding of the digital world's impact on adolescent health. By addressing the limitations of current research methods, we hope to in future introduce data donation into a cohort study landscape. This would ultimately contribute valuable insights that can inform public health recommendations and technology regulations, thereby facilitating more effective and nuanced policymaking in response to the evolving digital landscape.

Augmenting data donations – integrating TikTok DDPs, video metadata, and the multimodal nature of audio-visual content

Lion Wedel, Dr. Jakob Ohme, Prof. dr. T.B. Theo Araujo*

**Weizenbaum Institute*

This methodological contribution explores the potential of augmented Data Download Packages (aDDPs) as a novel approach to analyze digital trace data, using TikTok as a use case to demonstrate the broader applicability of the method. We demonstrate how these data packages can be used in social science research to understand user behavior, content consumption patterns, and the relationship between self-reported preferences and actual digital behavior.

We introduce the concept of aDDPs, which extend the conventional Data Donation Packages (DDPs) by augmenting the collected data with survey data, metadata, and multimodal content embeddings, among other possibilities. Specifically, this study uses the combination of data components in aDDPs to facilitate research on user endorsement behavior and content classification. This work provides an overview and, consequently, guidance for collecting and augmenting DDPs on TikTok and beyond.

In a pilot study on 18 aDDPs, we showcase the potential of the information breadth that can be collected and integrated. We train and compare uni- and multimodal classifiers, classify the 18 aDDPs, and investigate the extent to which endorsement behavior impacts future content suggestions. Finally, we contrast the empirical findings with the users' self-reported content consumption.

This study shows that aDDPs open up new spheres of research. With such a procedure, researchers are not merely bound to the information the donations carry but can investigate a plethora of questions that rely on classifications that the platforms do not provide. Questions on the distribution of anti-vax (Kim et al., 2023) or sexualized content (George & Surdeanu, 2023) can be expanded with an user centered perspective: what do users actually see, and how to they react to it? Besides supervised machine learning, an aDDP also provides data for unsupervised machine learning tasks to exploratively investigate content from digital platforms.

Port data donation software service on the Next platform

Adrienne Mendrik

Eyra

The Next platform is an open source web platform (<https://github.com/eyra/mono>) developed by Eyra in collaboration with researchers from various universities, that serves as an integration hub for sustainable software-as-a-service (SaaS) solutions, empowering research. The Port data donation software service is integrated into the Next platform and supports privacy preserving donation of digital trace data. Researchers can use the content management system on the platform to create a task list for participants, with a Qualtrics questionnaire task, pdf manuals with instructions to request and download the digital trace data, and a data donation task that enables researchers to extract relevant data from a data donation package and present it to participants for data donation consent. During the extraction process the data does not leave the device of the participant, but runs in their browser using Pyodide. If no privacy sensitive information is extracted, the privacy of the participant is preserved. Researchers or research engineers can write a custom python script (<https://github.com/eyra/feldspar>) to customize the flow to fit their research question and data source (i.e. Instagram, Netflix, Google, and so on). This flow can then be uploaded on the Next platform and presented to participants as a data donation task. The data donation software service on the Next platform is hosted and supported by Eyra. It can be used by sharing a link with participants to access the task list and is compatible with various panels, including the LISS panel, and with various data storage solutions at research institutions. A self-hosted version is available as well. The data donation software is governed by the Port program under the Nova Software Foundation, with members from Utrecht University, the University of Amsterdam and Eyra.

Hate Speech and Misinformation on WhatsApp: Insights from a Large Data Donation Program in India and Brazil

*Simon Chauchard**, *Kiran Garimella (Rutgers)*, *Fernando Mello (UC3M)*

**Social Sciences, University Carlos 3 of Madrid*

Over the past few years, alarming press reports assigning blame to WhatsApp usage for a variety of events have proliferated. In countries like Brazil and India, analyses have repeatedly suggested that group-based interactions on WhatsApp distort beliefs among the electorate (see, for instance Perrigo 2019; Bengani 2019; Benevenuto et al 2018), and beyond, that they impact various outcomes, including (but not limited to) individuals' propensity to engage in hostile, radical or violent behaviors (Chopra 2019, Magenta et al 2018; Ozawa et al 2023). While academic research has recently started to examine these dramatic narratives, much research remains to be done to evaluate their validity, and to disentangle the mechanisms through which WhatsApp may or may not be associated to these outcomes.

In this paper, we rely on the world's largest-to-date (WhatsApp) data donation program in India and Brazil to assemble a unique dataset of private WhatsApp groups in both countries. Relying on a new data extraction tool called WhatApp Explorer (Chauchard and Garimella 2024) and on an extensive, GDPR-compliant anonymization and privacy protocol developed as part of the ERC POLARCHATS project, we field (the data collection is ongoing at the time of this submission) two large, face-to-face data collection campaigns in a large number of locations in each of the two countries. During these campaigns, diverse sample of users in both countries are able to donate past and future data about their groups/threads, with the assistance of local implementing partners, in exchange for small incentives. Altogether around 3000 users in each country each donate up to 4 months of content about their groups (donating an average of around 5 groups each, as per recent pre-tests), allowing us to analyze the prevalence of specific types of content over vast amounts of data, their virality, to explore the type of groups or communities on which problematic content tend to proliferate, as well as document the profiles of users that tend to forward and/or receive such content.

10:00

Coffee Break

10:30

**Session: Applications in Health and Physical Activity
(Chair: Kasper Welbers)**

Measurement of Physical Activity in Older Adults through Data Donation

Bella Struminskaya, Florian Keusch, Joris Mulder, Stein Jongerius, & Adrienne Mendrik*

**Utrecht University*

Physical activity (PA) is a key predictor of many health outcomes, especially for aging populations. The accurate measurement of PA is key to identifying determinants of health and developing appropriate interventions. To measure PA, most population studies use self-report. However, self-reports are usually limited to global measures of PA (e.g., average daily hours of moderate/vigorous activity) and suffer from misclassification (e.g., walking the dog not considered PA). More fine-grained day-reconstruction methods are burdensome for respondents and prone to recall error. As an alternative researchers are providing study participants with wearable devices that passively track PA, which reduces reactivity and recall error. However, participants' non-compliance and high device costs are problematic. Many older adults now have smartphones that track physical activity and individuals can share these passively collected physical activity data with researchers. In this study, we test a data donation approach among older adults. Based on legal requirements such as the General Data Protection Regulation (GDPR), all data collecting companies and digital platforms need to provide users access to all of their data in a machine-readable format. We use a privacy-preserving data donation tool (Boeschoten et al. 2023) integrated in a probability-based online panel of the Dutch general population to collect PA data from health apps and Google Semantic Location History. We investigate determinants of consent and selection bias in PA data donation among ca. 2,000 individuals aged 50 and older in the Netherlands. Panel members who own a smartphone are asked to locate the in-built health apps on their smartphone (iPhone or Samsung) and perform the data request. Smartphone owners of other Android devices are asked to request Google Semantic Location History data from Google. Using the rich data available about panelists (from the last 15 years), we assess the quality of the donated PA data and evaluate how well multi-source physical activity data can predict health outcomes. Our study contributes to the development of future-proof methods for collecting high-quality PA data and innovations in surveys. Having self-report measures on technological skills (e.g., smartphone functions used) as well as measures on digital literacy already present in the LISS Panel and objective measures (i.e., donation outcome) will allow us to provide recommendations about using technologies such as apps and sensors for data collection in older populations.

Urban Hearing: Investigating individual noise, stress, sleep, and smartphone use via Apple Watch and iPhone Data Donation

Paulien Decorte, Karolien Poels, Gert-Jan de Bruijn, Cedric Vuye, Jonas Lembrechts, Karolien Couscheir, Sanne de Rooij, Ablenya Barros*

**Communication Sciences, University of Antwerp*

Urban noise pollution is a growing concern, with 40% of Europeans exposed to considerable noise levels every day (Hurtley, 2009). It affects our health, as it is associated with sleep disruption and increased stress (Münzel et al., 2014). Smartphones can be used to relieve stress or during sleep disturbances (Nabi et al., 2017; Tavernier & Willoughby, 2014), which may demonstrate how people cope with urban noise and its correlates. To discern these interrelationships in more detail on a day-to-day basis, the current study will investigate individual-level noise exposure, sleep quality and

duration, and stress (heart rate variability; HRV) through an Apple Watch data donation design. Encapsulated with a larger citizen science project, this data will be collected over a period of six weeks using the native Health Data from their Apple Watches. For two of these weeks, we will also examine their smartphone use through the Screen Time app, particularly when they engage in specific app use. During these six weeks, participants will also receive weekly surveys assessing their self-reported sleep, stress, and urban noise experience. Around N=330 participants have registered, with data collection planned from March-April 2024. After providing informed and GDPR consent, the first survey will instruct participants on the settings to allow for data collection. After two weeks, they will be asked to upload screenshots of their Screen Time. After six weeks, they will securely upload an export of their relevant Health Data from their Apple Watch for analysis of their data related to individual noise exposure, sleep, and stress (HRV). All data will immediately be anonymized through automatic scripts (Apple Watch data) or quantitative content analysis (Screen Time data). Results will be presented during the symposium, with attention for the exact procedure and implications for health and urban planning.

Mapping digital food environment on social media: a data donation approach

Kaiyang Qin^{}, Sixu Cai, Saar Mollen, Wilma Waterlander, Eline Smit*

^{}University of Amsterdam*

Objective: Social media represents an essential and emerging dimension of the digital food environment. Growing evidence indicates that exposure to the food environment on social media may be relevant to individuals' food choices and physical health. Yet less is known about what contributes to food content exposure on social media. In the present study, we aim to map the digital food environment on social media by exploring the two sources of (unhealthy) food content: food Ads and food posts.

Methods: A combined survey and data donation approach was utilized. Participants contributed their YouTube data, including video viewing history, likes, searches, subscriptions, and a company list that targeted users as potential customers. Additionally, they completed questionnaires addressing media algorithm awareness, perception of food norms, experiences with personalized ads, and demographics.

Analysis Plan: The study will categorize videos into healthy and unhealthy food-relevant content, along with likes, searches, channels, and company lists. Initial analyses will focus on descriptive statistics of food-related ads and posts, followed by correlational analyses linking survey results with views of food-relevant videos. Mediation analysis will explore the role of social media interactions (like food video likes and searches) as independent variables, with the number of food companies as mediators and the frequency of food video viewings as the dependent variable. Sub-group and latent class analyses will identify groups more likely targeted by food companies.

Implications: The findings are expected to have two major implications. Firstly, they may highlight the need for regulatory measures from legal and public health perspectives regarding the use of user data by food companies and social media platforms. Secondly, the study could offer insights into utilizing social media data through data donation methods for researching food-related topics.

Tesco Data Portability Tool: Empowering customers to donate shopping data for chronic pain research

Neo Poon, Claire Haworth, Elizabeth Dolan, & Anya Skatova*

**Bristol Medical School, University of Bristol*

Pain condition is considered a priority in healthcare and a threat to well-being across the globe, it is thus crucial to accurately measure the national levels of pain conditions and their impacts on workplace productivity and well-being.

Chronic pain has traditionally been studied in isolation with either self-reported survey data or standalone shopping records. The former are limited in scale, while the latter lack 'ground truths': what research teams can measure are usually the purchase patterns of pain relief products, but neither the severity nor types of pain conditions. Data donation tools offer a novel approach to study chronic pain by linking the two aspects and establish statistical relationships between medicine consumptions and the multiple facets of pain experience. In a survey, we asked participants (N = 1,000) to share their loyalty card data with us, which is made possible with the data portability tool provided by Tesco (i.e., the largest supermarket chain in the United Kingdom) as part of the GDPR. We also asked participants to report the details of their pain conditions, hours of employment, and mental health states, therefore associating chronic pain with its economic and personal consequences.

One key value of this project is to pioneer a measure of chronic pain that can be applied to transactional records that are much bigger in scales in future analytic works. Our research team has access to an array of different digital footprints data, including longitudinal transactional data provided by a major pharmacy chain (~20 million customers). In order to utilise these data to associate them with workplace productivity measures from the Office for National Statistics, a metric must be defined to extract the prevalence of chronic pain from shopping data, which is informed by the patterns found by the data donation project.

Data donation of personal physical activity trackers

Maaïke Kompier, Anne Elevelt, Annemieke Luiten, Joris Mulder, Barry Schouten, Vera Toepoel*

**Methodology, Statistics Netherlands*

Physical inactivity is a growing worldwide concern. Population monitoring of physical activity (PA) is generally done using questionnaires, yet recently there has been a strong interest in more objective forms of measurement using wearable activity trackers. Depending on the prevalence of personal activity trackers, respondents could be invited to share the data from their own devices. PA could then be determined with the help of the donated data. In this study, we explored two different methods of data donation to measure PA: uploading of spreadsheets and manual copying of data into questionnaires. Next to the response and representativity of those willing to donate, we compared the substantial outcomes of the two methods to assess PA. The results showed that prevalence of personal trackers was lower than anticipated, partially due to the focus in this study on the four mayor brands. Additionally, ownership was biased with age and education level. The majority of respondents were willing to copy their data in a questionnaire, whereas uploading with a spreadsheet proved to be too difficult. Future research should focus on assessing differences between brands and finding alternatives to measure PA amongst the population without a personal tracker.

The data in this study was gathered by Centerdata in collaboration with Statistics Netherlands. The research was conducted as part of a bigger, ongoing, PA project of Statistics Netherlands with the National Institute for Public Health and the Environment and the Amsterdam UMC in which objective alternatives to conduct population monitoring of PA are studied.

11:30

Coffee Break

11:45

**Session: Evaluation of Data Donation Approaches
(Chair: Heleen Janssen)**

A Systematic Review of Data Donation Literature: Defining Key Concepts, Best Practices, and Directions for Future Research

Yucan Xiong, Amber van der Wal & Teun Siebers*

**Communication Science, Amsterdam School of Communication Research (ASCoR)*

In response to the General Data Protection Regulation (GDPR), major digital platforms are obligated to offer users the ability to access their personal data through downloadable data packages (DDPs) (Breuer et al., 2023). The DDPs also introduce a great opportunity for researchers: by asking users to donate their DDPs researchers can collect digital trace data and gain more insights into users' behaviors (Boeschoten et al., 2020; Ohme et al., 2021). As a promising research tool, data donation has drawn increasing attention and is becoming more common in academic practices (Boeschoten et al., 2023). Yet, data donation studies are scattered across disciplines and vary widely in terms of the topics investigated, target group, and analytical approach employed (Boeschoten et al., 2020). As such, there is a current lack of a comprehensive overview of data donation practices, creating challenges in grasping the method's full potential and its practical applications. To address this issue, the present study has systematically reviewed 33 articles to provide a structured overview of the existing data donation literature. We begin by introducing key concepts of data donation. Secondly, we address the advantages and challenges associated with the application of this method, offering insights into how data donation can be effectively implemented in research settings. This includes various participant engagement strategies, exploring tactics to enhance participants' willingness to be involved in data donation studies. Thirdly, the focus is shifted to the tools and frameworks for data donation developed in previous studies. Suitable workflows and analytical methods are recommended for future research. Through this overview, we aim to establish a foundation for understanding data donation's suitability in various research contexts and pave the way for more standardized and efficient use of this method in academic research.

References

- Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). Digital trace data collection through data donation (arXiv:2011.09851). arXiv. <http://arxiv.org/abs/2011.09851>
- 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), 5596. <https://doi.org/10.21105/joss.05596>
- Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2023). User-centric approaches for collecting Facebook data in the 'post-API age': Experiences from two studies and recommendations for future research. *Information, Communication & Society*, 26(14), 2649–2668. <https://doi.org/10.1080/1369118X.2022.2097015>
- Ohme, J., Araujo, T., De Vreese, C. H., & Piotrowski, J. T. (2021). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, 9(2), 293–313. <https://doi.org/10.1177/2050157920959106>

Private messaging not so private: Is donating WhatsApp chats a good idea?

Felicia Loecherbach (the tool in question has been written by Kiran Garimella and Simon Chauchard, however they were not involved in the presentation and the tool is open source)

ASCoR, Universiteit van Amsterdam

Data donation allows researchers to access otherwise inaccessible data sources, for example in the case of end-to-end encrypted messaging services such as WhatsApp. They potentially play a role for e.g. forming opinions and sharing political information – however, this is difficult to tell especially in European contexts since data access to far has not been very tedious for researchers. While WhatsApp data access can be obtained manually, asking respondents to export the data chat by chat and upload it into a system (limiting the amount of chats and leading to high respondent burden), there is also the option of using an API-like system. After the respondent scans a QR code, the researcher almost instantaneously gains access to the full WhatsApp history (including media data and metadata) of the user. This offers new opportunities for studying long-term processes on private messaging with very little participant effort (as opposed to manual exporting) but also poses manifold legal and ethical challenges. In this presentation an overview will be given of what is technically possible while also discussing the manifold pitfalls especially in the context of GDPR and with different data sources that might be of interest for different research purposes (e.g. links, images, messages). In addition, this also involves how to properly communicate to respondents which data will and will not be accessed, as it can be assumed that donation of private messaging is even more subject to privacy concerns than other forms of social media data donation. The tool in question does already exist and has been used in different countries (India, Brazil, U.S.), but it remains to be seen how useful it can be in a European context and whether the legal and ethical implications make it worth using.

Data Donations in the Media: The public portrayal of a promising method

Schmidbauer, Elisabeth; Haßler, Jörg*

**Department for Media and Communication, Ludwig-Maximilians-Universität Munich*

Data donations are new to communication science. They offer advantages over conventional methods, but studies show that the willingness to donate is a decisive hurdle for data quality. One problem could be that the public has relatively little knowledge about the risks and benefits of participating in data donations studies. E.g., only 10 percent of survey participants in a sample representative for the Swiss population have heard the term before (Pfiffner & Friemel, 2023).

Media reporting is generally an instance of public knowledge transfer. As such, media coverage could be a first touching point with the term data donations for recipients and could potentially influence awareness, offer orientation, and shape opinions.

Against this background, our study analyses (1) the extent to which media report on data donations, (2) which key actors (e.g., private sector or public universities) are mentioned to be stakeholders in the context of data donations, (3) whether data donations are displayed positively in general, and (4) whether and to which extent risks and benefits for potential participants and the public are mentioned. Our study—which is work in progress at the moment—aims to map the media coverage of eleven major German news outlets and German Google News results, from 2014 to 2024.

We find 330 articles and 48 Google results mentioning data donations, most of them touching upon the RKI's COVID data donation. The comparatively low extent exacerbates the question of public's awareness and the necessity to identify key actors in the field. Moreover, it is important to understand whether the little information that is publicly transmitted additionally emphasizes risks (e.g., data protection violations) or benefits. These and further results will be available in early spring and discussed at the symposium.

Understanding the Potential of Data Donations: Evaluating Platform Compliance for Impactful Insights

Leonie Manzke (Tenure-Track Professorship for Digital Transformation, Friedrich-Alexander University Erlangen-Nürnberg); Philipp Hartl (Chair of Machine Learning and Uncertainty Quantification, University of Regensburg)

An increasing number of projects make use of Data Download Packages (DDPs). Applications range from generating insights in social and life sciences to empowering data-based solutions furthering the common good. Whether a research question can be addressed with a data donation study partly depends on the relevant platforms' adherence and implementation of the legally required data access to consumers. To mobilize individuals to donate their data, substantial efforts are needed. To justify these efforts, it is necessary to evaluate the DDPs provided by data-controlling platforms, to consider the feasibility, scalability, and impact potential of data donations in the given context.

The present work pursues two goals: 1.) Creating a structured framework of practical considerations relevant to evaluating DDPs. 2.) Sharing the results of evaluations investigating several platforms' (non-)compliance with designated laws and their adherence to the standards of conciseness, accessibility, completeness, and transparency. Other researchers engaged in data donation studies have collected such insights related to DDPs from social media platforms (e.g. Instagram: Van Driel et al., 2022). This presentation supplies evaluations for two more use cases: user data from the language learning platform Duolingo, and loyalty card data from different supermarkets in Germany (e.g., EDEKA). While the content and structure of DDPs may change and evolve over time, these evaluations offer a snapshot of the current situation and highlight potentially worthwhile characteristics. By compiling these insights, this presentation endeavors to offer orientation and enable impactful data donation initiatives in the future.

12:35

Lunch

13:35

Session: Challenges Related to Legal, Ethical Considerations or Platform Restrictions (Chair: Felicia Loecherbach)

How digital platforms may narrow down data access rights: data download packages reveal platform interpretations of GDPR obligations

H.L. Janssen, J. Ausloos, T.C. Carrière, B. Struminskaya, L. Boeschoten, T. Araujo & N. De Schipper*
**Institute for Information Law, Faculty of Law, University of Amsterdam*

Information rights and access rights in the General Data Protection Regulation (GDPR) aim to empower individuals. Yet the right of access also holds potential for lawful purposes beyond individual data protection: researchers may partner with individuals and request them to donate data obtained from access requests for academic research.

Platforms (data controllers) increasingly respond to these requests by sending individuals a 'data download package' ('DDP'). DDPs theoretically provide individuals with transparency regarding the processing of their data and with other information the individual is entitled to under the GDPR. A platform's self-engineered DDP can offer tools to appropriately inform individuals but DDPs also permit controllers to 'shape' the interpretation of their GDPR obligations. We compared eight platform controller DDPs with their own privacy policy. We found that none of the platform DDPs, nor their privacy policies contain complete and transparent information as required by the GDPR.

In the gap between regulatory intention of individual empowerment through controller information obligations and access rights on one hand, and controller interests in lowering cost on the other, “interpretive entrepreneurship” has occurred. Interpretive entrepreneurship refers to businesses developing methods and strategies to shape interpretation of existing laws to their advantage. The risk-based nature of the GDPR, meaning that the regulator has essentially left execution of and compliance with legal obligations to controllers, and vagueness of the legal requirements, including transparency and completeness, permit powerful platform controllers significant leeway to interpret and operationalise legal obligations as they deem appropriate.

Such practices may lead to wrong interpretations and faulty compliance, effectively undermine the right of data access, and, related to scientific research, to incomplete data sets and incorrect research results. We aim to demonstrate and explain this entrepreneurship, and consider ways to improve the effectiveness of information rights and access rights more systematically.

Yet Another Privacy Paradox? Stated intention and actual willingness for data donations

*Philipp Hartl**, Chair of Machine Learning and Uncertainty Quantification, University of Regensburg, 93040 Regensburg, Germany

Daniel Schnurr, Chair of Machine Learning and Uncertainty Quantification, University of Regensburg, 93040 Regensburg, Germany

Many people readily disclose personal information if such disclosure is associated with personal benefits. This behavior is often in stark contrast to their stated concerns about data protection and use of their data [Gerber et al., 2018], giving rise to the renowned privacy paradox [Norberg et al., 2007]. However, much less is known about individuals’ stated intentions and actual behavior when data is disclosed for the common good. Therefore, this study investigates whether the privacy paradox extends to data donations. From a theoretical perspective, we hypothesize that the intention-behavior gap may even be more relevant for data donations, as prosocial behavior itself is often characterized by inconsistencies between individuals’ stated intentions and actual action. In particular, people often express a higher willingness to engage in prosocial behavior than they are actually willing to do due to social desirability [Crowne and Marlowe, 1960]. Based on a between-subjects online experiment, we therefore measure both the stated intentions and actual data disclosure of participants when data is disclosed either for personal or social benefits. By doing so, we can test for potential intention-behavior gaps [Webb and Sheeran, 2006] and whether they can be explained by well-accepted theories on individuals’ privacy decision-making [Culnan and Bies, 2003] and pro-social behavior. To minimize priming effects, the experimental design separates measurement of intentions from actual data disclosure by a temporal delay. Finally, we elicit subjects’ personality traits and privacy attitudes to evaluate their influence on stated intentions and actual willingness to donate data to gain a better understanding of individuals’ incentives to engage in pro-social data disclosure. Altogether, this study contributes to an improved understanding of intention-behavior gaps in the context of data donation. These findings bear important insights for data regulation and the design of data donation platforms and campaigns.

Development of a TikTok Data Donation Study Among Adolescents: A Promising Method, A Challenging Process

Nadia Bij de Vaate, Ine Beyens, Amber van der Wal, Tim Verbeij, Teun Siebers, and Patti Valkenburg*

**Communication Science, Amsterdam School of Communication Research (ASCoR)*

Ever since the emergence of social media, scholars have exerted great efforts to accurately operationalize social media use (e.g., boyd & Ellison, 2008; Verbeij et al., 2021). Most studies have investigated the time participants spend on social media on a typical day or the frequency of engagement in specific activities, like posting or browsing (Gerson et al., 2017). A main drawback of these frequency-based measures is their lack of information about the content users encounter,

hindering a comprehensive understanding of social media use. The overwhelmingly diverse content encountered during social media use impedes accurate measurement through self-reports. To address the limitations of frequency-based operationalizations, scholars generally advocate for more objective content-based social media use measures.

A promising method to gain more insights into the content that users post and see on social media is data donation. In our project, we specifically focus on the donation of TikTok data by adolescents, given that they are among the most avid users of social media, especially TikTok. Our aim is twofold: 1) to inductively explore what content adolescents share and browse on TikTok, and 2) to combine self-reported social media experience data and data donation data to unravel how and when TikTok elicits positive and/or negative experiences. Our goal is to acquire data donation packages (DDPs) from approximately 400 adolescents, using the data donation software tool Port.

While data donations hold promise, the process of developing such a study poses challenges. At the Data Donation Symposium, we aim to share our latest insights, challenges, and obstacles encountered in the development of our TikTok Data Donation project. We will discuss challenges related to obtaining and exploring TikTok DDPs. More specifically, we will address issues with requesting and downloading the DDPs, the evolving structure of the DDPs, and the deployment of Port.

References:

- boyd, d. m., & Ellison, N. B. (2008). Social networking sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Gerson, J., Plagnol, A. C., & Corr, P. J. (2017). Passive and active Facebook use measure (PAUM): Validation and relationship to the reinforcement sensitivity theory. *Personality and Individual Differences*, 117, 81-90. <https://doi.org/10.1016/j.paid.2017.05.034>
- Verbeij, T., Pouwels, J. L., Beyens, I., & Valkenburg, P. M. (2021). The accuracy and validity of self-reported social media use measures among adolescents. *Computers in Human Behavior Reports*, 3, 100090. <https://doi.org/10.1016/j.chbr.2021.100090>

Measuring Privacy Behavior with Donated Data

Frieder Rodewald, Florian Keusch, Daria Szafran, Ruben L. Bach*

**School of Social Sciences, University of Mannheim*

Privacy is a ubiquitous issue in our increasingly digitalized society. This study investigates the impact of users' privacy decisions on what data social network platforms collect about them. We build on a data donation approach to observe people's privacy behavior on Instagram directly. Our findings contextualize and advance previous studies that almost exclusively relied on users' self-reported behavior. Data is collected from a German probability-based online panel in two stages: Participants first complete a survey about their privacy behavior and well-studied predictors of that behavior, such as privacy concerns and privacy literacy. One week after the survey, participants are asked to download and donate their Instagram usage data. A one-week gap between the survey and donation request minimizes the mere-measurement effect, ensuring participants are less likely to hesitate in donating personal data due to prior privacy-related questions. The paper investigates 1) how participants' privacy behavior observed in the donated data compares to their self-reported privacy behavior and 2) how using observed privacy behavior enhances current models studying the relationship between privacy concerns and privacy behavior. Ultimately, our research shows that data donation is a valuable data collection method to understand online privacy and how that is a vital next step to explaining people's online privacy behavior.

The Feasibility of WhatsApp Chat Log Donations: Willingness vs. actual Donations in an opt-in Sample.

Julian Kohne ; Christian Montag*

**Computational Social Science, GESIS - Leibniz Institute for the Social Sciences*

Our study explored the feasibility of WhatsApp data donations using ChatDashboard, our open-source tool comprising the WhatsR R-package, ChatDashboard Shiny app, and DashboardTester script. We conducted an opt-in study with a survey and optional data donation, offering interactive feedback on participants' chat behavior post-donation. The survey (N = 320) gathered data on demographics, personality (BFI-10; ANPS), attachment style (AAS), social media usage, and privacy concerns (CFIP). Participants indicating willingness to donate a WhatsApp chat log provided additional details about their chat partner, including demographics, closeness, online/offline contact frequency, and relationship type and duration. They then uploaded the chat log, could remove unwanted information, and received interactive feedback. Specifically, we addressed the following research questions:

1. Who is participating in opt-in studies with the option for data donations?
2. What differentiates people who indicate they are willing to donate data from those who are not?
3. What differentiates willing donors from unwilling ones?
4. How much, and what kind of information, are willing donors manually excluding from their donations?

First results indicate that willingness (N = 183; 52.18%) and actual donation rates (N = 125; 39.06%) were remarkably high in our design, but still showed a large intention-behavior gap, casting doubt on the usability of willingness as a proxy for measuring actual donation behaviors. Furthermore, we observe a quite homogenous self-selected sample consisting mostly of younger, female university students. Interestingly, initial analyses do not show any significant effects of demographics, personality, attachment style, or concerns for privacy on data donation willingness or actual data donations. In our data, 14 participants manually deleted a complete variable from their donation (mostly domains of sent links or type of media attachments), and 3 participants removed all information about a specific message.

14:35

Coffee Break

15:05

Session: Data Donation Applications (Chair: Theo Araujo)

Investigating the influence of friendships on YouTube usage history similarity using data donations

Nico Pfiffner, Lukas Tribelhorn, Thomas N. Friemel*

**Department of Communication and Media Research, University of Zurich*

This study investigates the impact of friendships on the similarity of YouTube usage histories among youth, utilizing data donated by participants. For many young people, YouTube is an important media platform when it comes to information access, entertainment, and community building. Beside relying on YouTube's algorithmic recommendations, users' media choices are also influenced by their social (offline) contacts. Drawing on co-orientation and social network theory, the study posits that individuals with close relationships exhibit similar media preferences due to frequent communication

and a desire to connect with similar others. This kind of interpersonal influence can occur through explicit content recommendations or implicit perceptions of others' preferences.

Previous research showed that interpersonal offline relationships can influence young people's choices of overarching content categories such as TV programs, YouTube channels, or music genres. However, it remains unclear to what extent such similarities can be observed on the micro level of individual media usage histories in a highly personalized media environment such as YouTube. Therefore, this study seeks to exploratively answer the question: How does the (non-)existence of friendship ties between individuals relate to the similarity of their YouTube usage?

Employing a network analytical approach, the study combines survey data with digital trace data from 50 to 100 friend dyads. Participants complete a questionnaire and donate their YouTube usage histories, encompassing watch history, search history, and subscribed channels (using the Google Takeout function and a dedicated data donation application). Participants are aged between 16 and 25.

Ethics approval has been secured, and data collection is currently ongoing. Presently, 85 "seed" participants have completed the study, with 24 successfully inviting friends who also contributed their YouTube data. Upon conclusion of data collection in February, the study will analyze similarities in various metrics, including watched videos, channels, topics, and temporal patterns.

Subjective and objective perceptions of hate speech in Switzerland

Dominique Wirz & Sina Blassnig (University of Zurich)*

**Amsterdam School of Communication Research, University of Amsterdam*

This study uses a representative survey in Switzerland (N = 2000) and a subsequent online diary study with data donations (N = 150) to investigate how often Swiss media users perceive hate speech, which statements they classify as "hate speech" and what role social identities plays in this. 69% of Swiss people have perceived hate speech at least once, 24.2% are confronted with hate speech (several times) a day or several times a week. In general, statements are more likely to be classified as "hate speech" if they concern minorities. Insulting and rude statements are more likely to be classified as "hate speech" if they concern a group with which the respondents identify. This shows that the perception and assessment of hate speech is partly subjective. 150 participants who indicated that they see hate speech at least several times a week were invited to an online diary study and asked to screenshot each evidence of hate speech they see. We then conducted a content analysis with these data donations to identify if the statements contain hate speech, and if yes what type of hate speech. This study provides indications of the objective confrontation with hate speech: Most participants are indeed confronted with hate speech several times a week. At the same time, some overestimated in the survey how often they encounter hate speech online. The diary study confirms the findings of the survey that hate speech occurs particularly frequently on social media and that insults are perceived most often, followed by defamation, while threats are less common. The examples uploaded in the diary study contain rudeness more frequently than hate speech. This indicates that the participants also perceive some "impolite" statements as hate speech. Finally, members of marginalized groups uploaded examples relating to their own social group membership more frequently.

Enhancing Social Media Data Donation for Research: A Researcher-Assisted Approach

Yibei Yu, Xiaohua He*, Kaiyang Qin, Tianlai Ye, Annemiek Linn*

**University of Amsterdam*

The data that people leave on social media platforms offer significant opportunities for academic research. When donating their data, participants normally follow the instructions independently to complete the data donation without immediate assistance. However, such practices have been hampered by several difficulties such as non-response and dropouts. Anticipated challenges for participants include potential technical issues, privacy concerns, and uncertainties during the data

retrieval process. Consequently, the opportunities presented by data donation for research are not fully optimized. As of now, standardized methods for data collection have not been fully established, and there is no comprehensive overview of the perceived difficulties participants may encounter when donating their data. The aim of this study is to test a structured and detailed approach to facilitate voluntary social media data donation. Utilizing the PORT framework, Facebook data is donated with participants receiving assistance through one-on-one online meetings. During these meetings, participants are guided by a researcher to retrieve and upload their data. Difficulties encountered during the process are systematically documented and recorded in a study log. This approach proves to be an effective alternative compared to those without immediate assistance, resulting in a higher response rate. This approach is helpful in increasing the response rate and decreasing the drop-out rate when organizing and implementing the data donation studies with various social media platforms.

Data donations: A mixed-methods approach to explore how users engage with news

Tim Groot Kormelink, Fiore Houwing*

**Language, Literature & Communication, Vrije Universiteit Amsterdam*

In a digital media landscape that is characterized by audience fragmentation and algorithmically curated feeds, it has become increasingly challenging to study how users engage with news on digital platforms. A new method of digital data collection that shows great promise in this area is data donation. Data donation refers to the tools and processes through which individuals make available for academic research the data that digital platforms (e.g. Google, Facebook, TikTok, LinkedIn) have collected about them (Araujo et al. 2020). This study uses an infrastructure that was developed to guide users willing to donate their data for academic research through the process of requesting their data from platforms and donating these to research.

Combining data donations with think-aloud protocols and in-depth interviews, this mixed-methods study has two key aims. First, methodologically, it explores how research participants experience and understand the data donation process. Second, substantively, this study explores how people engage with news and information on different online platforms, and what (relative) role these platforms play in their overall information repertoire.

The data collection will take place in the Netherlands between January and February 2024. 20 participants will be selected via snowball sampling. While studies using data donation have so far typically focused on one platform, this pilot study includes seven platforms: Google, YouTube, Facebook, Twitter (X), Instagram, LinkedIn, TikTok. The project is user-centered, in the sense that the data donation will focus on those platforms that each participant in question predominantly uses. For instance, if participant A uses Google, Twitter, and TikTok, they will be assisted to request and donate the data from these specific platforms. Overall, this study hopes to generate a more detailed, nuanced understanding of the role digital platforms play in people's news use.

Analyzing User Engagement with TikTok's Short Format Video Recommendations using Data Donations

Savvas Zannettou

Delft University of Technology

Short-format videos have exploded on platforms like TikTok, Instagram, and YouTube.

Despite this, the research community lacks large-scale empirical studies into how people engage with short-format videos and the role of recommendation systems that offer endless streams of such content. In this work, we analyze user engagement on TikTok using data we collect via a data donation system that allows TikTok users to donate their data. We recruited 347 TikTok users and collected 9.2M TikTok video recommendations they received. By analyzing user engagement, we find that the average daily usage time increases over the users' lifetime while the user attention remains stable at

around 45%. We also find that users like more videos uploaded by people they follow than those recommended by people they do not follow.

Our study offers valuable insights into how users engage with short-format videos on TikTok and lessons learned from designing a data donation system.

Behind the Screens: Exploring Netflix via Data Donations

Dennis Nguyen^{}, Karin van Es*

^{}Media and Culture Studies, Utrecht University*

Netflix is often hailed as a disruptor of the traditional media industry, primarily due to its extensive data collection and analysis capabilities, as well as the effectiveness of its recommendation system. As a subscription-based service, Netflix has historically been reticent to share viewing data with external parties. However, under increasing public pressure and with the introduction of an ad-supported tier, Netflix has begun to show greater transparency. Nonetheless, as Lotz (2023) observes, “Netflix is only sharing information it wants to share.” This selective data sharing results in a reliance on Netflix's own narratives about aspects like binge-watching, content popularity and diversity, and the intricacies of their recommendation algorithms. Such narratives contribute to perpetuating myths, limiting our understanding of streaming platforms and obscuring their true disruptive impact (van Es 2023).

In the context of the European Union's General Data Protection Regulation (GDPR), Netflix subscribers can request their viewership data, which is provided in CSV files containing detailed information such as viewing activity, device usage, profile interactions, and playback events. Our Data Donations project engaged Netflix subscribers, inviting them to share this data for research purposes, along with completing a survey about their streaming preferences and perceptions of Netflix's recommender system. By analyzing data and survey responses from approximately 200 Netflix subscribers, this study conducts a critical examination of Netflix's assertions. It also sheds light on the potential limitations and inaccuracies in the data and metrics released by Netflix, underscoring the relevance of data donations.

References

Lotz, A. D. (2023). ""Netflix Data Dump: Cautions Regarding Gifts of Data."" Retrieved from <https://www.linkedin.com/pulse/netflix-data-dump-cautions-regarding-gifts-amanda-lotz-sa7pf%3FtrackingId=dzRGj195SdOBwqZA%252BS62eA%253D%253D/?trackingId=dzRGj195SdOBwqZA%2BS62eA%3D%3D>

Van Es, K. (2023). ""Netflix, (Re)Claiming Television: Myths and Horseless Carriages."" Keynote at ECREA TV Studies Section Conference, Potsdam, Germany, 26 October 2023.

16:20-16:30

Closing