

Mobility Data Mining: from Technical to Ethical

Bettina Berendt^{*1}, Stan Matwin^{*2}, Chiara Renso^{*3}, Fran Meissner⁴,
Francesca Pratesi⁵, Alessandra Raffaetà⁶, and Geoffrey Rockwell⁷

- 1 TU Berlin, Germany & Weizenbaum Institute – Berlin, Germany & KU Leuven, Belgium. berendt@tu-berlin.de
- 2 Dalhousie University – Halifax, Canada. stan@dal.ca
- 3 Istituto di Scienza e Tecnologie dell'Informazione, National Research Council of Italy – Pisa, Italy. chiara.renso@isti.cnr.it
- 4 University of Twente, The Netherlands. f.meissner@utwente.nl
- 5 Istituto di Scienza e Tecnologie dell'Informazione, National Research Council of Italy – Pisa, Italy. francesca.pratesi@isti.cnr.it
- 6 University Cà Foscari of Venice, Italy. raffaeta@unive.it
- 7 University of Alberta – Edmonton, Canada. grockwel@ualberta.ca

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 22022 “Mobility Data Analysis: from Technical to Ethical” that took place fully remote and hosted by Schloss Dagstuhl from 10–12 January 2022. An interdisciplinary team of 23 researchers from Europe, the Americas and Asia in the fields of computer science, ethics and mobility analysis discussed interactions between their topics and fields to bridge the gap between the more technical aspects to the ethics with the objective of laying the foundations of a new Mobility Data Ethics research field.

Seminar January 10–12, 2022 – <http://www.dagstuhl.de/22022>

2012 ACM Subject Classification Computing methodologies → Machine learning; Computing methodologies → Artificial intelligence; Security and privacy → Human and societal aspects of security and privacy; Human-centered computing → Interaction design; Social and professional topics → Computing / technology policy; Security and privacy → Database and storage security; Security and privacy → Software and application security

Keywords and phrases Dagstuhl Report, Mobility Data Mining: from Technical to Ethical

Digital Object Identifier 10.4230/DagRep.12.1.35

1 Executive Summary

Bettina Berendt

Stan Matwin

Chiara Renso

Fran Meissner

Francesca Pratesi

Alessandra Raffaetà

Geoffrey Rockwell

License © Creative Commons BY 4.0 International license

© Bettina Berendt, Stan Matwin, Chiara Renso, Fran Meissner, Francesca Pratesi, Alessandra Raffaetà, and Geoffrey Rockwell

Mobility data is one of the fastest growing types of data, thanks to the increasing number of mobile devices approaching the population of the globe. The collection, storage and analysis of spatio-temporal data representing trajectories of moving objects is one of the topics that

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Mobility Data Analysis: from Technical to Ethical, *Dagstuhl Reports*, Vol. 12, Issue 1, pp. 35–66

Editors: Bettina Berendt, Stan Matwin, Chiara Renso, Fran Meissner, Francesca Pratesi, Alessandra Raffaetà, and Geoffrey Rockwell



Dagstuhl Reports

DAGSTUHL
REPORTS

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

received major attention in the field of data analytics. The more semantic information is collected from various sources, the richer is movement data. This enriched mobility data is typically referred to as semantic trajectories. The analysis of such trajectories can produce powerful results in domains such as transportation, security, tourism, health, environment and even policy design. The recent COVID-19 outbreak shed a light on the importance of collecting mobility data for public health. However, at the same time, the more mobility data is enriched with semantics, the larger the risks of violating the privacy of users and of possible unethical uses of these data analysis results. Aspects of Computational Ethics include privacy, but they go beyond this, towards a more general vision of ethical gathering, processing, uses of data and the results of data analyses. How ethics interrelates with mobility data analysis is an emerging issue.

The objective of this Dagstuhl Seminar was therefore to start a deep interacting discussion between Mobility Data Analysis researchers and Ethics experts to link these two fields with the objective of creating the foundations of a new Mobility Data Ethics research field.

This Dagstuhl Seminar, organised by Chiara Renso, Bettina Berendt and Stan Matwin as an activity from and beyond the MASTER project [1], aimed at bringing together researchers from different disciplines from Computer Science, Mobility Analysis and Ethics to trace the path from a technical vision of mobility Analysis to an also ethics-based approach to the field.

The three-day seminar was structured into three main modules: (1) round-table presentations in which each participant presented him/her self with a question about Mobility and Ethics that represents his/her interest and an object to visualise this interest or serve as a starting point for further discussion; (2) three tutorial on “technical”, “ethical” and “legal” aspects of mobility data; (3) the working groups to discuss the main topics of interest that emerged during phases (1) and (2).

As a result of the group discussions on participants’ interests and the issues raised in the tutorials, we formed five main working groups:

- What is/are the trade-off(s) between data privacy and data utility?
- Mobility Data Anonymity (Can location data be really anonymous?)
- Ethics of Mobility Data: What is unique? Which guidelines?
- Mobility Data Analysis Ethics beyond the data
- Mobility Data Analysis Ethics beyond humans only: Tracking animals and moral agency

The tutorials and each of the working groups are described in a chapter of this report. Like other Dagstuhl Seminar reports, these chapters aim at making the scientific results re-usable and extendable by others. In addition, we also want to help others profit from our experiences with the videoconferencing and other media technologies that we employed and the interaction-design choices that we made. This last chapter is a reflection also on ethical aspects of the precluded and the newly added forms of mobility of scientists (and others) in meetings during and after COVID-19.

References

- 1 Chiara Renso, Vania Bogorny, Konstantinos Tserpes, Stan Matwin, and José Antônio Fernandes de Macêdo. Multiple Aspect Analysis of semantic trajectories (MASTER). *Int. J. Geogr. Inf. Sci.*, 35(4):763–766, 2021.

2 Table of Contents

Executive Summary

<i>Bettina Berendt, Stan Matwin, Chiara Renso, Fran Meissner, Francesca Pratesi, Alessandra Raffaetà, and Geoffrey Rockwell</i>	35
---	----

Tutorials

Location privacy: an overview <i>Sébastien Gambs</i>	38
Mobility data analysis: ethical issues <i>Geoffrey Rockwell</i>	40
Connected vehicles and mobility data – work done by the EDPB <i>Peter Kraus</i>	42

Topic-based Working Groups

What is/are the trade-off(s) between data privacy and data utility? <i>WG scribe and other members: Francesca Pratesi, Bettina Berendt, Thierry Chevallier, Josep Domingo-Ferrer, Ioannis Kontopoulos, Jeanna Matthews, Anna Monreale</i>	43
Mobility data anonymity <i>WG scribe and other members: Francesca Pratesi, Jeanna Matthews, Anna Monreale, Florence Chee, Ioannis Kontopoulos, Karine Zeitouni</i>	47
Ethics on mobility data: what is unique? Which guidelines? <i>WG scribe and other members: Geoffrey Rockwell, Christine Ahrend, Florence Chee, Thierry Chevallier, Maria Luisa Damiani, Peter Kraus, Fen Lin, Fran Meissner, Alessandra Raffaetà, Chiara Renso, Paula Reyero Lobo, Yannis Theodoridis, Karine Zeitouni</i>	48
Mobility Data Analysis Ethics beyond the data <i>WG scribe and other members: Fran Meissner, Fen Lin, Florence Chee, Peter Kraus, Chiara Renso, Paula Reyero Lobo, Yannis Theodoridis</i>	53
Mobility Data Analysis Ethics beyond humans only: Tracking animals and moral agency <i>Chiara Renso, Bettina Berendt, Stan Matwin</i>	55

Online Interactions in a COVID-19-era Dagstuhl Seminar: Design, Experiences, and Reflections

What do we consider “success”?	57
What did we do, and what did we learn from it?	57
A sense of place in a hybrid world – and other parallels between the medium and the message	63

Conclusions 64

Participants 66

3 Tutorials

3.1 Location privacy: an overview

Sébastien Gambs (Université du Québec à Montréal (UQAM), Canada)

License  Creative Commons BY 4.0 International license
© Sébastien Gambs

In the introduction of his tutorial, Sébastien highlighted the fact that location is personal data that is collected and used in many different contexts such as location-based services, geolocated advertising, augmented reality, mobile game, collaborative traffic monitoring, call details records, physical analytics, smart cities as well as electronic payments, just to name a few. Moreover in situation of crisis, such as the COVID pandemic, there was a huge pressure in many countries by public health agencies for the access to mobility data (*e.g.*, Call Details Records collected by telecom operators) to use it to understand how the movements of persons affect the spread of the disease as well as whether the population was respecting the confinement rules. While location privacy can be preserved from an external attacker through classical security mechanisms (*e.g.*, the use of TLS (Transport Layer Security) to secure communications between the client and service provider), one of the inherent risks is its possible abuse by this service provider due to rich inference potential of location data.

In the second part of his talk, Sébastien Gambs reviewed inference attacks against mobility data whose main objective is to quantify the risks in terms of privacy related to the collection and disclosure of mobility data. After defining the goal of location privacy as preventing an undesired entity from learning the past, present and future location of an individual [1], he discussed why pseudonymization offers a very low level of privacy protection for mobility data due to the possibility of using the pair home-work as quasi-identifiers if the spatial granularity is too much fine-grained [2] or the observations that the combination of 4 four random locations visited by a user is usually unique in the population as previous works have shown.

Afterwards, Sébastien Gambs presented some of his own work on inference attacks against mobility data such as the identification of points of interests based on a clustering algorithm [3], the prediction of mobility patterns using a mobility Markov chain [4] or a de-anonymization attack against anonymized mobility traces in which the mobility model is used as auxiliary knowledge by the adversary to re-identify mobility traces [5]. Other inference attacks briefly reviewed include profiling, exploiting the co-location of users to predict their location [6], as well as performing membership inference [7] or reconstructing trajectories from aggregated data [8].

The third part of the tutorial was dedicated to privacy-preserving methods for mobility data publishing. The objective of these methods is to sanitize the data to preserve location privacy, in particular by preventing some of the attacks described in the previous part. When sanitizing data, there is an inherent trade-off between the desired level of privacy and the utility of the sanitized data. Here, utility can be defined with respect to global properties of the data or be dependent on the application considered. First, simple sanitization mechanisms were introduced such as the geographical masks, which protect privacy through aggregation or perturbation, or approaches based on sampling or removing records that are too atypical. Afterwards, the spatio-temporal version of the k -anonymity privacy model, called spatial cloaking, was presented [9]. Methods that address privacy by limiting the possibility for the adversary to link together the mobility traces belonging to the same identity, such as mix-zones and Swapmob [10], were also discussed. Finally, the differential privacy model [11] as well as its applications for privacy-preserving mobility analytics [12] as well as privacy-preserving trajectory synthesis [13] were also briefly mentioned.

In his conclusion, the speaker mentioned that while there has been a huge scientific literature in the last 20 years on location privacy, the fundamental question of defining and quantifying location privacy in a formal yet actionable manner is still partially open. More precisely, what does it mean to have “good” location privacy?

- To be hidden inside a crowd gathered in a small area?
- To be alone in a desert?
- To have a behavior indistinguishable from those of a non-negligible number of other individuals?
- To be unlinkable between different positions?

Sébastien Gambs emphasized that whatever the approach chosen, it should prevent the inference of sensitive information from the location data revealed, rather than focusing on hiding only the location itself. Finally, he listed several open challenges for research on location privacy: (1) the lack of a set of (preferably few) privacy and utility metrics on which there is consensus in the community, (2) the absence of large-scale reference datasets that could be used to benchmark algorithms (mainly due to confidentiality and privacy reasons) and (3) the fact that often the source code of algorithms is not published. All these three challenges limit the possibility of comparing different sanitization algorithms and hinder reproducible science.

References

- 1 Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Comput.*, 2(1):46–55, 2003.
- 2 Philippe Golle and Kurt Partridge. On the anonymity of home/work location pairs. In Hideyuki Tokuda, Michael Beigl, Adrian Friday, A. J. Bernheim Brush, and Yoshito Tobe, editors, *Pervasive Computing, 7th International Conference, Pervasive 2009, Nara, Japan, May 11-14, 2009. Proceedings*, volume 5538 of *Lecture Notes in Computer Science*, pages 390–397. Springer, 2009.
- 3 Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Show me how you move and I will tell you who you are. *Trans. Data Priv.*, 4(2):103–126, 2011.
- 4 Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In Hamed Haddadi and Eiko Yoneki, editors, *Proceedings of the First Workshop on Measurement, Privacy, and Mobility, MPM '12, Bern, Switzerland, April 10, 2012*, pages 3:1–3:6. ACM, 2012.
- 5 Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. De-anonymization attack on geolocated data. *J. Comput. Syst. Sci.*, 80(8):1597–1614, 2014.
- 6 Alexandra-Mihaela Olteanu, Kévin Huguenin, Reza Shokri, Mathias Humbert, and Jean-Pierre Hubaux. Quantifying interdependent privacy risks with location data. *IEEE Trans. Mob. Comput.*, 16(3):829–842, 2017.
- 7 Apostolos Pyrgelis, Carmela Troncoso, and Emiliano De Cristofaro. Measuring membership privacy on aggregate location time-series. In Edmund Yeh, Athina Markopoulou, and Y. C. Tay, editors, *Abstracts of the 2020 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems, Boston, MA, USA, June, 8-12, 2020*, pages 73–74. ACM, 2020.
- 8 Fengli Xu, Zhen Tu, Yong Li, Pengyu Zhang, Xiaoming Fu, and Depeng Jin. Trajectory recovery from ash: User privacy is NOT preserved in aggregated mobility data. In Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich, editors, *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 1241–1250. ACM, 2017.

- 9 Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In Daniel P. Siewiorek, Mary Baker, and Robert T. Morris, editors, *Proceedings of the First International Conference on Mobile Systems, Applications, and Services, MobiSys 2003, San Francisco, CA, USA, May 5-8, 2003*, pages 31–42. USENIX, 2003.
- 10 Julián Salas, David Megías, and Vicenç Torra. Swapmob: Swapping trajectories for mobility anonymization. In Josep Domingo-Ferrer and Francisco Montes, editors, *Privacy in Statistical Databases – UNESCO Chair in Data Privacy, International Conference, PSD 2018, Valencia, Spain, September 26-28, 2018, Proceedings*, volume 11126 of *Lecture Notes in Computer Science*, pages 331–346. Springer, 2018.
- 11 Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 2006.
- 12 Mohammad Alaggan, Mathieu Cunche, and Sébastien Gambs. Privacy-preserving wi-fi analytics. *Proc. Priv. Enhancing Technol.*, 2018(2):4–26, 2018.
- 13 Takao Murakami, Koki Hamada, Yusuke Kawamoto, and Takuma Hatano. Privacy-preserving multiple tensor factorization for synthesizing large-scale location traces with cluster-specific features. *Proc. Priv. Enhancing Technol.*, 2021(2):5–26, 2021.

3.2 Mobility data analysis: ethical issues

Geoffrey Rockwell (*University of Alberta, Canada*)

License  Creative Commons BY 4.0 International license
© Geoffrey Rockwell

Geoffrey Rockwell’s presentation started with an example from the early days of the pandemic when Tectonix, a geospatial analysis company, tweeted a video of a visualization showing where people active during the March 2020 spring break on a beach in Ft. Lauderdale returned to after their break. Tectonix was promoting how their analysis and visualization technology combined with mobility data from X-Mode could help efforts to control the pandemic by tracking the irresponsible youth who come to Florida to party on the beach instead of social distancing.

This technology demonstration on the one hand appalled some on Twitter who were surprised by how their privacy wasn’t respected, but it also coincided with stories that the Trump administration was in conversation with various companies to see if location data could be used in anonymized form to track the novel coronavirus. An official quoted in The Washington Post article that broke the story said that mobility data could, “help public health officials, researchers, and scientists improve their understanding of the spread of COVID-19 and transmission of the disease” [1]. The point of this introductory case study was that there are significant ethical issues around how mobility data is gathered, used, aggregated, and shared. The variety of ways mobility data can be gathered, the inferences possible, and the number of players who are commercializing such data raise difficult problems around principles, privacy, the application of ethical frameworks and how to develop a culture of ethical use. In the rest of the talk, Rockwell surveyed these issues in order to provide a common ground for discussion.

Why bother with ethics? Ethics is often seen as a drag on innovation. Rockwell reviewed some reasons why addressing ethics is becoming important in both commercial and academic research and development [2].

What ethical principles apply? A common starting point in data ethics is to identify common principles that should guide researchers, developers and users. Rockwell surveyed some applicable sets of big data principles like the 2017 Big Data Guidelines from the Information and Privacy Commissioner of Ontario and principles from the Harvard Business School [3]. Researchers, however, often come to different conclusions, even if they start from similar principles. For example, on the issue of using mobility data for public health we found positions like: Ultimately, if residual risk is acceptable, analysis of mobility data can be justified if it can yield actionable insights that benefit public health [4].

From these discussions spurred by the current pandemic, *three major location data issues* emerged: (1) anonymized data sometimes are not anonymous, (2) location data are often not representative and can exacerbate inequality, and (3) location data are a key part of the extension of the surveillance state.

Rockwell concluded the case study by asking whether this moment of visibility for location data collection could provide an opportunity to push for new media literacies [5].

Privacy. Rockwell reviewed some of the definitions of privacy and how they might apply to mobility data [6, 7, 8].

Ethical Approaches. Finally, Rockwell reviewed some of the common ethical approaches that can be used to think through situations:

- Duty-Based Ethics where rules and guidelines are used to establish the morality of actions rather than their consequences;
- Consequentialist Ethics such as Utilitarianism which judge the morality of actions based on their consequences;
- Ethics of Care which emphasizes the relationships in actions and care for others.

Rockwell concluded by asking how we can develop a culture of ethics in data disciplines. Some general resources mentioned in the talk include also [9].¹

References

- 1 Tony Romm, Elizabeth Dwoskin, and Craig Timberg. U.S. government, tech industry discussing ways to use smartphone location data to combat coronavirus. *The Washington Post*, 2020. <https://www.washingtonpost.com/technology/2020/03/17/white-house-location-data-coronavirus/>.
- 2 Anne-Laure Thieulent et al. AI and the ethical conundrum: How organizations can build ethically robust ai systems and gain trust. Technical report, Capgemini Research Institute, 2020. <https://www.capgemini.com/research/ai-and-the-ethical-conundrum/>.
- 3 Catherine Cote. 5 Principles of Data Ethics for Business, 2021. Harvard Business School online, <https://online.hbs.edu/blog/post/data-ethics>.
- 4 Bouke C. de Jong et al. Ethical considerations for movement mapping to identify disease transmission hotspots. *Emerging Infectious Diseases*, 25(7):e181421, 2019. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6590736/>.
- 5 Jordan Frith and Michael Saker. It is all about location: Smartphones and tracking the spread of COVID-19. *Social Media + Society*, 6(3), 2020. <https://journals.sagepub.com/doi/10.1177/2056305120948257>.
- 6 James Moor. The ethics of privacy protection. *Library Trends*, 39, 1990.
- 7 Kate Raynes-Goldie. Aliases, creeping, and wall cleaning: Understanding privacy in the age of facebook. *First Monday*, 15(1-4), 2010. <https://doi.org/10.5210/fm.v15i1.2775>.

¹ The slides of this tutorial are available at:
<http://www.master-project-h2020.eu/dagstuhl-materials/>.

- 8 Danah Boyd. Debating privacy in a networked world for the WSJ, 2011. <http://www.zephoria.org/thoughts/archives/2011/11/20/debating-privacy-in-a-networked-world-for-the-wsj.html>.
- 9 Josh Lauer. Sarah E. Igo. The Known Citizen: A History of Privacy in Modern America. *The American Historical Review*, 124(3):1019–1021, 2019.

3.3 Connected vehicles and mobility data – work done by the EDPB

Peter Kraus (European Data Protection Board, Brussels, Belgium)

License  Creative Commons BY 4.0 International license
© Peter Kraus

Peter Kraus gave a short presentation of how the European Data Protection Board (EDPB) works as a European Body, composed of Members from the different Member States of the European Economic Area and the European Data Protection Supervisor. The EDPB, being tasked with ensuring the consistent application of Europe’s General Data Protection Regulation (GDPR), has issued Opinions, Binding Decisions, Guidelines, Best Practices and Recommendations. He also gave a short overview of the EDPB Strategy and the transnational enforcement done by the Member State authorities.

Concerning the topic of mobility, Mr Kraus highlighted the work done by the EDPB regarding contact tracing during the COVID-19 pandemic and presented the EDPB’s Guidelines on connected vehicles [1]. These Guidelines analyse the different legal instruments that are relevant from a data protection perspective and their interaction: the GDPR and the e-Privacy Directive. The e-Privacy Directive, as a so called *lex specialis*, provides additional protection to electronic communications, whereas the GDPR sets the general framework.

Following the presentation by Mr Rockwell and the mentioning of possibly applicable ethical principles, Mr Kraus highlighted that the GDPR is known as a principles-based legislation, which in its Article 5 contains the legal principles of data protection, which in many respects are similar to the ethical principles mentioned for ethical use of mobility data. This includes, in particular, the principles of fairness, transparency and data minimisation. These principles are highlighted and operationalised by Article 25 of the GDPR – Data Protection by Design and by Default.

One of the possible legal bases under the GDPR to process and to use personal data, as well as one of the requirements of the e-Privacy directive to access personal data on an end device, is consent. However, consent has a list of properties for it to be considered valid consent in the legal context. It needs to be informed, unambiguous, specific and freely given. This relates back to the principles of data protection. Further, the GDPR requires that consent can be withdrawn and that this withdrawing must be as easy to do as it is to give consent. Therefore, when using consent as a legal basis, consideration should be given to what happens when the consent is withdrawn.

One way to avoid being in scope of data protection law altogether is to effectively anonymise personal data. However, as already presented by Mr Gambs, anonymisation is hard to do in practice. Encryption and hashing are frequently mistaken for anonymisation processes, whereas from a legal data-protection perspective, they merely amount to pseudonymisation. One has to be aware of re-identification attacks that may turn a dataset previously deemed to be anonymised into one that is not anonymised at all. Lastly, it was highlighted, that anonymisation is particularly difficult for location data due to the additional context that can be derived from a specific location.

Finally, a few examples that are provided in the Guidelines on connected vehicles were given, which showcase how the legal framework can be applied to use cases such as contact

tracing or usage based insurance services.

References

- 1 European Data Protection Board. Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications, 2020. https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-12020-processing-personal-data_en.

4 Topic-based Working Groups

Inspired by the round-table presentations of attendees and the tutorials and discussions, we identified the following main topics as of interest of the audience to build the new mobility data ethics “from the technical to the ethical”: (1) What is/are the trade-off(s) between data privacy and data utility?, (2) Mobility data anonymity – asking whether such data can ever be really anonymous, and (3) Ethics on mobility data: what is unique? Which guidelines? While these three groups retained the focus on data of the seminar title and the focus on humans often implicit in the term “ethics”, two further groups explored the need to go beyond this: (4) Mobility Data Analysis Ethics beyond the data and (5) Mobility data analysis beyond humans only: Tracking animals and moral agency.

4.1 What is/are the trade-off(s) between data privacy and data utility?

WG scribe and other members: Francesca Pratesi, Bettina Berendt, Thierry Chevallier, Josep Domingo-Ferrer, Ioannis Kontopoulos, Jeanna Matthews, Anna Monreale

License © Creative Commons BY 4.0 International license
 © WG scribe and other members: Francesca Pratesi, Bettina Berendt, Thierry Chevallier, Josep Domingo-Ferrer, Ioannis Kontopoulos, Jeanna Matthews, Anna Monreale

To understand the bond between data privacy and data utility, in a first step one needs to consider or develop definitions and measures for both these dimensions. What is *utility*? How to measure it? How to do a utility vs. privacy (but also fairness and other ethical dimensions) analysis? How do we actually and effectively define and measure privacy?

As starting points, from the background and the past experience of participants regarding the processing of mobility data, the best way to act is to reduce granularity. Unfortunately, even if many papers in the literature prove that this approach can really work [1, 3, 2], there are two main problems: (1) sometimes it is hard to apply data minimization (indeed, it is hard to define what we need before knowing what you want to do); (2) there is a huge difference between one-shot collection vs. the continuous anonymization of trajectories (i.e., data streams). Regarding the first problem, purpose specification can help to overcome this, especially for researchers; moreover, the General Data Protection Regulation (GDPR) defines the “public interest” purpose and allows derogation for research, cf. [4].

However, connected to this latter point, there is the *secondary-use problem*. Although it is generally considered good practice for administration or companies to *reuse* data (even for potentially good purposes) without the need to perform potentially expensive and time consuming new data collection (e.g., establish new surveys), we must be aware that this practice violates the GDPR principles. An example that was cited are instances of (mis?)use

of a centralized Corona app in Germany², an app that was quasi-obligatory for the use of several services (such as going to the restaurant), so this implied a huge coverage of the system. Therefore, special caution must be taken and ensure that the collected data will not be used for other purposes during or after the emergency in the context of which they were originally collected.

Unlike the data-protection principle of purpose limitation, principles of open science allow and encourage secondary use; the drawback is that stored datasets can often be de-anonymized quite easily. This is one reason why access to open datasets is conditioned on some constraints, such as the definition of a specific research project. As an example, in the Netherlands, it is possible, for researchers working with a university or other institution associated with the national statistics office, to access the comprehensive and highly detailed Dutch Microdata. To get access, researchers have to submit their project idea. Once approved, they can access the data in a secure environment provided by the statistics office. Payment is per month plus per export (because each export is checked for meeting common privacy standards).³

A solution to this problem is to build and use synthetic data.

During the working group discussion, we focused on concrete use cases, such as local buses management (in terms of both real-time position and capacity) and journey planners. Such planners strongly depend on the starting point and destination of the user, on the means of transport (which can also be suggested by the service), and on the time of the day. You can also consider “mobility as a service application”, where you can also have other utility integration (e.g., buying tickets) that involve other stakeholders (e.g., bus companies), and possible second uses of data. This helped us highlight different contrasting dimensions:

Tracing people vs. not tracing people. Both tracing and not tracing can be approaches to solving a problem, but these problems are different in that they have different needs and challenges. Some services can be obtained even without the tracking of people in a privacy-preserving way, so not tracing people could still be a good option. In the case of buses, for example, sensors can be installed on buses that can give the real-time aggregated occupancy, a value that can be exploited to avoid further data gathering. Clearly, when we do not track the location of buses or people inside it, we do not know the actual route. For this kind of service, avoiding the tracking we are actually applying data minimization: we do not need to track people and we will not do it. The difference in terms of privacy in the different scenarios is due to the need to know an immediate location characteristic (e.g., a bus position or whether a person is on a bus) versus the need to know the actual movements and, thus, to actually trace people (of course, only in this case we have to deal with personal data).

Historical data vs. real-time data. Historical data are data collected in the past and generally used to build some models; these models can be used at a later time for several uses. Here, already existing methodologies can be applied, such as randomization or k-anonymity. Once the model has been computed on historical data, the model will likely be used. Indeed, one must share his/her position to take advantage of it, resulting in a real-time scenario. While hiding persons in the crowd during the building of the model is relatively simple, when

² This was the app “Luca”, not to be confused with the “main” German tracing app CWA, which is decentralised. Regarding the adoption rate of such measures in general, and the importance to provide incentives, see [5].

³ <https://www.cbs.nl/en-gb/our-services/customised-services-microdata/microdata-conducting-your-own-research>

the system is in use you have more constraints, because the environment is both dynamic and distributed. There is probably the need to define a new concept of privacy in this case. Indeed, in the real-time context, it is more difficult to compute the privacy risk since, in general, we do not have the general overview of the situation. The most promising approaches in this context seem to be *randomization* or *federated learning in location data*. The limitation of the first solution is that when you are the only person in an area, even when noise is consistently being added, you will always be re-identifiable; thus, you could have no privacy and no utility at all. Federated learning approaches are similarly limited: when the peers have very different location data, there are no guarantees of protection.

Service quality vs. data quality. To assess the utility, we usually rely on service quality, measuring whether we are able to perform similar analyses with or without privacy. This represents what the end user loses with a privacy-preserving service. In the bus scenario, an example could be to suggest a longer route to reach a destination in order to preserve privacy: the privacy-preserving route could be more expensive in terms of more time but also more money for gas or the bus ticket. Therefore, we must be able to measure the difference in performance (e.g., accuracy

or cost-sensitive performance metrics) between the model with and without privacy (i.e., with private data or raw data).

The “kind of risk”. This dimension is related to the risk you are taking into consideration. Examples are the re-identification risk or the disclosure of sensitive attributes (e.g., profiling). We can also consider if the risk is data-centric or real-life-centric. Regarding the risk of re-identification, there are some practical tools⁴, based on the work presented in [6, 7].

Multi-criteria utility function. Utility is often equated with accuracy of the model, but accuracy is not the only objective. For example, there are the *fairness* trade-off or the *transparency* trade-off. We call these options *simple model vs. complex model*: one can prefer an explainable model over one with perfect accuracy. While on the one hand, we need to protect privacy, on the other hand, transparency is a valuable aspect that could be in conflict with privacy. How to achieve both *and* maintain utility? In this sense, we could start from the work done in [8], where some contrasting dimensions are considered (even if here the utility is missing), and [9], where it is highlighted that there are values that help each other and there are values that harm each other; it also depends on the choice you made.

Participation vs. non-participation. This is another trade-off that needs to be considered. We should be able to measure the trade-off between benefits and opportunity costs, consider disincentives, and be sure that subjects really understand the risk of participation. Moreover, it happens very often that a certain service cannot be pursued if the adoption rate does not reach a certain threshold. Thus, we are led to thinking that there is a strong need to collect large quantities of data. However, this thinking may be the result of assuming *false trade-offs*: usually, a lot of data means a high utility, and this leads to a high risk; on the contrary, no data means no utility but also no risk. However, this need not be the case, and with some transformed (e.g., partially or badly anonymized) data one may obtain something in the middle: no utility and full risk [10].

Stakeholder diversity. During the discussion, we agreed on the existence of three different stakeholders for both privacy and utility, aiming at different kinds of utility:

⁴ <https://github.com/scikit-mobility/scikit-mobility>

- *subject*, who provides the data or uses the service. Regarding the privacy aspects, subjects who provide data have the specific need that their data cannot be re-identified. At the same time we want to be able to maintain some data utility. In general, we can refer to *individual utility* that is how good is the service or *collective utility* meaning that a user can share his/her data if there are benefits for the society, such as reducing traffic in a city. Regarding the final user of the service, we can use other metrics. For example, as said before, a user can take more time by using a path different from the shortest with the purpose of preserving privacy.
- *data owner*, who is usually the provider owning the data, and who possibly permits to third parties to access its data. This subject has interest in keeping the ownership of the data and prevent uses by third parties without their consent; their utility is usually the profit. We can consider the utility from the perspective of data collector: you should check the difference of performances (e.g., accuracy) of the model with or without privacy (i.e., with private preserved data or raw data).
- *data analyst*, who is in charge of analyzing data and providing the service. This subject wants the confidentiality of results. Its utility can be recognized in profits if it is related to a private company or in benefits for the society if it corresponds to a public entity. More information about these three kinds of privacy can be found in [11]. Moreover, these different kinds of perspectives can be easily related: for example, when data is privacy protected, this is good not only for the subjects but also for the owner of the service. It is important to consider in this analysis also the economic concept of externalities [12].

To conclude, the last part of the discussion focused on the accountability and the responsibility, especially at the societal level, of the actual implementation and on the possible implications of the utility computation. In particular, we discussed the use of checklists, where the outcome is not a specific value but rather the starting point for the evaluation. These checklists come in the form of a list of questions to the user, such as “Did you consider this aspect?”.

Finally, about the application of this data utility mechanism: even supposing that we have the perfect formula for utility computation, who is going to implement it?

References

- 1 George Danezis, Josep Domingo-Ferrer, Marit Hansen, Jaap-Henk Hoepman, Daniel Le Metayer, Rodica Tirta, and Stefan Schiffner. Privacy and data protection by design-from policy to engineering, 2014. ENISA Report. <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design>.
- 2 Anna Monreale, Gennady Andrienko, Natalia Andrienko, Fosca Giannotti, Dino Pedreschi, Salvatore Rinzivillo, and Stefan Wrobel. Movement data anonymity through generalization. *Transactions on Data Privacy*, 3, 2010.
- 3 Anna Monreale, Salvatore Rinzivillo, Francesca Pratesi, Fosca Giannotti, and Dino Pedreschi. Privacy-by-design in big data analytics and social mining. *EPJ Data Science*, 3(10), 2014. <https://doi.org/10.1140/epjds/s13688-014-0010-4>.
- 4 Michael Beauvais. GA4GH GDPR Brief: The public interest and the GDPR, 2021. <https://www.ga4gh.org/news/ga4gh-gdpr-brief-the-public-interest-and-the-gdpr-february-2021/>.
- 5 Mirco Nanni et al. Give more data, awareness and control to individual citizens, and they will help COVID-19 containment. *Trans. Data Priv.*, 13(1):61–66, 2020.
- 6 Luca Pappalardo, Filippo Simini, Gianni Barlacchi, and Roberto Pellungrini. Scikit-mobility: a python library for the analysis, generation and risk assessment of mobility data, 2019. <https://arxiv.org/abs/1907.07062>.

- 7 Francesca Pratesi, Anna Monreale, Roberto Trasarti, Fosca Giannotti, Dino Pedreschi, and Tadashi Yanagihara. Prudence: a system for assessing privacy risk vs utility in data sharing ecosystems. *Trans. Data Priv.*, 11(2):139–167, 2018.
- 8 Marit Hansen, Meiko Jensen, and Martin Rost. Protection goals for privacy engineering. *2015 IEEE Security and Privacy Workshops*, pages 159–166, 2015.
- 9 Josep Domingo-Ferrer and Alberto Blanco-Justicia. Ethical value-centric cybersecurity: A methodology based on a value graph. *Sci. Eng. Ethics*, 26(3):1267–1285, 2020.
- 10 Bettina Berendt. Better data protection by design through multicriteria decision making: On false tradeoffs between privacy and utility. In Erich Schweighofer, Herbert Leitold, Andreas Mittrakas, and Kai Rannenberg, editors, *Privacy Technologies and Policy – 5th Annual Privacy Forum, APF 2017, Vienna, Austria, June 7-8, 2017, Revised Selected Papers*, volume 10518 of *Lecture Notes in Computer Science*, pages 210–230. Springer, 2017.
- 11 Josep Domingo-Ferrer. A three-dimensional conceptual framework for database privacy. In Willem Jonker and Milan Petkovic, editors, *Secure Data Management, 4th VLDB Workshop, SDM 2007, Vienna, Austria, September 23-24, 2007, Proceedings*, volume 4721 of *Lecture Notes in Computer Science*, pages 193–202. Springer, 2007. <https://crises-deim.urv.cat/web/docs/publications/lncs/457.pdf>.
- 12 Bogdan Kulynych, Rebekah Overdorf, Carmela Troncoso, and Seda F. Gürses. Pots: protective optimization technologies. In Mireille Hildebrandt, Carlos Castillo, L. Elisa Celis, Salvatore Ruggieri, Linnet Taylor, and Gabriela Zanfir-Fortuna, editors, *FAT* '20: Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, January 27-30, 2020*, pages 177–188. ACM, 2020.

4.2 Mobility data anonymity

WG scribe and other members: Francesca Pratesi, Jeanna Matthews, Anna Monreale, Florence Chee, Ioannis Kontopoulos, Karine Zeitouni

License © Creative Commons BY 4.0 International license
 © WG scribe and other members: Francesca Pratesi, Jeanna Matthews, Anna Monreale, Florence Chee, Ioannis Kontopoulos, Karine Zeitouni

In this working group, the discussion focused on multiple aspects of data anonymity, which is extremely hard to reach [3] for mobility data [1], especially considering the fact that mobility patterns are usually highly predictable. As experts, can we recommend (and develop) ways to use technology still but with some evasions of location tracking (e.g., download offline maps and navigate with GPS), and do our best to increase people’s awareness, especially considering how we are exposed and easily manipulable [2].

Moreover, there is a need to consider the danger of predatory data collection. Sometimes, the choices are false choices (you “must” download an app due to peer pressure or to participate in ordinary social life). Therefore, we should encourage the definition of different levels of consent. Another possibility is to implement tools to help people to analyze the permissions required from mobile phone apps⁵ or from web pages⁶.

However, as said in Section 4.1, it is worth noting that the choice to participate in a project or to share something affects the final goal: the goodness of the model depends on the number of individuals that participate. And this has an impact on the individuals

⁵ <https://privacyflag.eu>

⁶ <https://github.com/chatziko/location-guard>

themselves. In the literature a scenario that evaluates this trade-off is missing: what is the impact (in terms of privacy loss) if you participate? And what is the impact (in terms of service quality) if you do not participate? Another challenge is that this must be computed and evaluated in real-time. Moreover, in the mobility context, it is not enough to have a certain percentage of the population, but we need to evaluate also the geographical area; it seems that the only possibility is to simulate this computation.

The discussion ended up with a possible trend of personal data management, providing decentralized solutions. There are solutions actually existing, such as personal data cloud, where you locally store data, but to be usable, they need suitable infrastructures and ad hoc protocols. This is not implementable today because there is no pressure neither from governments and surely not from companies. We need to ask ourselves if we have the power to influence choices, both at individual and at companies levels.

A follow-up article on the concerns of surveillance and the best practice we can adopt to contrast them has already been published [4].

References

- 1 Hui Zang and Jean Bolot. Anonymization of location data does not work: a large-scale measurement study. In Parmesh Ramanathan, Thyaga Nandagopal, and Brian Neil Levine, editors, *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking, MOBICOM 2011, Las Vegas, Nevada, USA, September 19-23, 2011*, pages 145–156. ACM, 2011.
- 2 Jeanna Matthews. How fake accounts constantly manipulate what you see on social media – and what you can do about it. *The Conversation*, 2020, Jun 24. <https://theconversation.com/how-fake-accounts-constantly-manipulate-what-you-see-on-social-media-and-what-you-can-do-about-it-139610>.
- 3 Alex Hern. “anonymised” data can never be totally anonymous, says study. *The Guardian*, 2019, Jul 23. <https://www.theguardian.com/technology/2019/jul/23/anonymised-data-never-be-anonymous-enough-study-finds>.
- 4 Geoffrey Rockwell, Bettina Berendt, Florence M. Chee, Jeanna Matthews, Sébastien Gambs, and Chiara Renso. Ottawa’s use of our location data raises big surveillance and privacy concerns. *The Conversation*, 2022, Jan 27. <https://theconversation.com/ottawas-use-of-our-location-data-raises-big-surveillance-and-privacy-concerns-175316>.

4.3 Ethics on mobility data: what is unique? Which guidelines?

WG scribe and other members: Geoffrey Rockwell, Christine Ahrend, Florence Chee, Thierry Chevallier, Maria Luisa Damiani, Peter Kraus, Fen Lin, Fran Meissner, Alessandra Raffaetà, Chiara Renso, Paula Reyero Lobo, Yannis Theodoridis, Karine Zeitouni

License © Creative Commons BY 4.0 International license

© WG scribe and other members: Geoffrey Rockwell, Christine Ahrend, Florence Chee, Thierry Chevallier, Maria Luisa Damiani, Peter Kraus, Fen Lin, Fran Meissner, Alessandra Raffaetà, Chiara Renso, Paula Reyero Lobo, Yannis Theodoridis, Karine Zeitouni

The question of data ethics and technology ethics has been a hot topic for some time. Over the past decade following an explosion of the availability of various and voluminous data sets often referred to as big data, a common laminate has been that technology is often developing faster than the regulations that might keep the implications of those technologies and data uses at bay in step with the innovations in the field. The mobility data analysis field is not excluded and efforts are underway to engage with and “update” ethics thinking to

new data and technology realities. Those efforts raise a number of questions. In particular, we discussed the questions described in the following Sections 4.3.1–4.3.3. As a result of the discussion, we discussed existing guidelines and drew up a tentative set of guidelines tailored to the needs of a graduate student doing a mobility data project, provided here in Section 4.3.4.

4.3.1 What is unique about Mobility Data from an ethics perspective?

In many ways the field of mobility data aligns with many other fields that have moved from relative data scarcity to a relative data abundance. So is there anything unique about the type of research? We here argue that while there are many parallels and there is a lot to learn from the ongoing wider debates, our argument here is that there is indeed something, if not unique, at least specific about the mobility data if viewed from an ethics perspective. Much of that uniqueness derives from the type of data needed for the mobility analysis.

One of the peculiarities of mobility data is that we can deduce/infer knowledge. Mobility data is first of all characterized by spatial-temporal information and this allows us to relate this data to the context, by considering the environment where the moving object is traveling. In this way, we can detect the points of interest that the moving object is crossing, the time spent in each place, their semantic categories. Based on the time and frequency of the places, we can infer the workplace and/or home. Moreover, we have the possibility to understand the interaction among the moving objects, if they are moving in a group, who they are in a meeting and to have an idea of the social network.

Some features of mobility data merit highlights.

- **Scale:** the amount of information from smartphones/sensors is massive. Such data expand the notion of temporal and spatial scales that used to be developed from the “small-and-static” data. Moreover, the granularity of the spatial and temporal information can vary significantly (e.g. days vs. seconds, km vs. cm).
- **Associated Metadata:** depending on the provider, mobility data can often come with associated metadata about the user or about the application.
- **Auto-correlation:** the data are not independent and they have incredible details correlated in time and space. The temporal component, especially, seems to add a new dimension to the ethics.
- **Relational:** the mobility data is beyond categorical information, it also reveals the relational behaviour – who were close to, who were you likely talking to, what might you have been doing.
- **Behavioural prediction:** due to the temporal and spatial detail, Mobility Data can be used to infer everyday behaviour in ways that many other sources of data don’t. If ethics is about guidelines for behaviour, then having such detail about day-to-day behaviour raises questions about how behaviour can be surveilled, managed, and manipulated in new ways.

As a result, mobility data is much more than “simple” geographic data, and these data are incredibly valuable both for fine-grained analysis but also in economic terms. This value of mobility data brings with it an array of different interests in how and for what purposes mobility data ought to be used. At the same time it is also hard to anonymize full mobile datasets, and this creates challenges when it comes to sharing data for open science purposes.

Such features of mobility data imply that there is an array of normative questions that mobility researchers are faced with today. These questions go beyond, but also accompany, more traditional concerns such as how to preserve privacy of individuals in the data, as

elaborated on in Section 4.1. This also means that due to the nature of mobility data is engaging with, articulating and addressing those questions is exceptionally important. The myriad of ethical considerations that need to enter the mobility data field have gained attention that is too fragmented to create clarity, for researchers, about applicable ethics guidelines and how to use ethical considerations as a means to engage in responsible research. The following takes the uniqueness of mobility data we here postulate as a starting point to engage with some important questions that are often not very clearly spelled out.

4.3.2 What is the role of context and culture in Mobility Data Ethics?

We recognize that ethical practices vary across contexts and cultures. For example, the MIT moral machine experiments that examined moral decisions of self-driving cars across 42 countries revealed a cultural divide in ethics [1]. The use and regulation of mobility data might therefore need to integrate contextual situations and normative ethical principles. At the same time it may be necessary to recognize that some technical solutions that work in one context might have adverse effects if applied in another context, making awareness for the context of analysis and application crucial in assessing ethics concerns.

Even though big data technology in general, and technology using mobility data specifically, enables researchers to efficiently understand the people's behaviour, which might offer insightful suggestions for policy-makers to develop better governance, it inevitably invites the conundrum of balancing utility and personal privacy. How to make such a balance might also vary by contextual situation (thus adding to the many other factors to take into account that are discussed in Section 4.1) and across cultures. The contextualized ethical practices might also be manifest in various stages of operationalization of mobility data: data collection, data analysis, data publication, data sharing and data storage. Such contexts might also bring issues that are related to ethics but go beyond ethics. For example, when a private citizen attends a public event, to what extent do they own their behaviour and trajectory data in the public event?

We recognize that there are communities that do not want to be surveilled and, in fact, want to remain hidden from the research gaze. Likewise, we recognize that researchers are not necessarily neutral or objective observers who have only the good of the community in mind. It is the responsibility of the researcher to engage communities that they are going to surveil in a dialogue and be willing to *not* gather data in some cases or to redesign their research in others. Further, there are now contexts in which it would be an "appropriation of voice" to do research which is not co-designed with the community. Indigenous communities in particular have made it clear that no research should be conducted without their involvement and approval.

On these debates much can be learned from ongoing debates about data justice. For example, Taylor [2] argues that one useful approach might be to take a capabilities approach – meaning to think through how individual and group capability is hindered or facilitated – to think through data justice. This requires thinking through at least three conundrums. (1) what sort of visibility is being made possible with the data and can access to representation as well as access to information privacy be maintained. (2) how is an engagement with the technology made possible that allows both to share the benefits of data and allow for an autonomy in technology choices; and (3) how can a principle of non-discrimination be adhered to in a way that allow for challenging bias and preventing discrimination. Facilitating capabilities is also an relevant aspect of debates that link data technologies to the questions of ethics.

4.3.3 What are the benefits and risks of ethics (and too much ethics)?

There is a balance between research utility and addressing various ethical concerns. In this sense, ethics should be part of the research process as a dialogue to ensure there is a minimum risk of harming people and this goes beyond complying with legal regulations.

Several factors involved with the use of mobility data advocate for the need for an ethics assessment to be in place before and during the whole lifetime of the project. Privacy preservation is a fundamental aspect, but not the only one. While there are questions about data manipulation and control which relate to both ethics and privacy, we discuss other important aspects of data ethics that differentiate these two concepts. The impact of the project needs to be understood to be able to make decisions such as which data to use or how to use it, in order for it to have the best possible outcomes. This is an intricate task, as the impact may not always be straightforward, for example, if the data being used may not be directly linked to an individual. However, is it important that researchers have a legitimate interest and data subjects are aware that their activities are being recorded and the purpose for which they are used?

Such question invokes conversation that were circled around the tensions between a) the value that we, as researchers, put on openness and the sharing of data and b) the privacy rights of individuals or groups whose data are collected. This is particularly a problem in mobility data research due to the difficulty involved in anonymizing datasets. In fact, sharing the data and the methods used in a research project is essential in open science. It allows the community to check and reproduce the results that support the findings. It is the basis of evaluating and benchmarking different approaches.

Some of the approaches discussed that could mitigate this tension include:

- Minimize your data to that which is needed for replication.
- Researching and applying new practices in anonymization. We recognize that this is itself a research field where approaches are changing.
- Embargo your data for a number of years until the privacy challenge is lessened.
- Ask people to apply for access to your data and provide it with clear terms of use.
- Provide synthetic data instead of the real data.

4.3.4 A Tentative Guideline

On the second day we discussed guidelines. We were inspired by the Locus Charter [3]. We agreed that the guidelines could be organized by audience and different sets of guidelines might be proposed at different stages of processing data, such as data collection, data analysis, data storage and data sharing. We proposed a tentative guideline for a graduate student doing a mobility data project.

Guidelines for a Graduate Student doing a Mobility Data project

Policy. Inform yourself about the research ethics and data privacy policies of your university and research lab. If there is a research ethics board to which you need to apply to proceed with your project then you should engage that board. Don't think of the process as an obstacle, but as a chance for dialogue.

Data. Identify where you are getting your data from. Are you gathering the data yourself, or are you getting data from a mobility provider, or have you inherited a data set? Make sure you fully document the provenance and structure of the data even if you have inherited it.

Transparency and Consent. If you are gathering data you should consider how you could get consent or at least be transparent about what you are gathering and why. You might, for example, have a notice at the location where you gather data or you might publish a web site that provides information on the project. As part of your transparency you should describe what you are gathering, what you will do with it, and how someone might hold you accountable.

Security. Develop a data security plan. Check it with your supervisor and data security colleagues. This is especially important if you are working with data that can identify individuals or is sensitive in some fashion.

Check for Bias. Ask yourself if your dataset is biased. Who is represented? Who is not? Does this make a difference to the research? Familiarise yourself with the literature on bias so you understand what might be reasons for bias. Test your dataset for bias if you can.

Research Goal. Ask yourself what you are trying to achieve with this research and whether the research project itself is ethical. Are you trying to do good, or could there be unintended harms from misuse of your data or research? Is the new knowledge generated helpful, useful, or welcome?

- (a) A good starting point is to identify your values, including the often unexpressed values embedded in Western modes of research. What do you think is good such that your research will make a difference.
- (b) It is a good idea to identify the stakeholders in your research, both those whose data is gathered and others (including the research team). Engage the stakeholders in a dialogue about the design of the research.
- (c) Consider how they might have different values. They may not value the research you are doing or value it differently.
- (d) Make sure you are transparent about the goals of your research so people understand the purpose for the data gathering, the analysis, and publication.
- (e) Consider the ethics of the methods and analytics used in the project. Could your methods be unethically applied in a different context, or with different data?
- (f) Audit the research. Share your research goals and plans with other people, starting with your supervisor and colleagues in order to get input on the ethics of the means and the goals. If there is an ethics board, or people with a designated ethics role, you should get their feedback early and often.

Data Management Plan. Develop a long-term DMP that describes how you will process your data to minimize it, what you will deposit in the university research data repository for future use, whether there should be an expiry date to the data, and whether it should be embargoed for a while. Consult with research data librarians about the plan. Ask yourself and others how your data could be used for good or misused. Document your wishes for the data (and include this with the data deposited) so you are clear how you hope it will or will not be used.

Ask yourself about possible second use cases. Even if your own research may adhere to best research standards you should audit your data and project. Could your data, the methods you develop or the findings you plan to publish lead to adverse effects? If so, what are those effects and can you think of and actively implement mitigation strategies that will prevent adverse effects and promote socially responsible reuse of your work?

Care and Repair. Maintain and repair your data, methods and research publications.

References

- 1 Edmond Awad, Sohan Dsouza, Azim Shariff, Iyad Rahwan, and Jean-François Bonnefon. Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proceedings of the National Academy of Sciences*, 117(5):2332–2337, 2020.
- 2 Linnet Taylor. What is data justice? the case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2):2053951717736335, 2017. <https://doi.org/10.1177/2053951717736335>.
- 3 EthicalGEO. Lotus charter, 2021. <https://ethicalgeo.org/locus-charter/>.

4.4 Mobility Data Analysis Ethics beyond the data

WG scribe and other members: Fran Meissner, Fen Lin, Florence Chee, Chiara Renso, Paula Reyero Lobo, Yannis Theodoridis

License © Creative Commons BY 4.0 International license

© WG scribe and other members: Fran Meissner, Fen Lin, Florence Chee, Peter Kraus, Chiara Renso, Paula Reyero Lobo, Yannis Theodoridis

Debates about the ethics of mobility analysis tend to focus on the data and what it reveals. They are often framed in light of privacy and the problems with guaranteeing privacy given the specificity of spatio-temporal data compared to other categorical data. To take a step back from this perspective, we explore what kinds of ethics questions surface if we think about the ethics of mobility data analysis beyond the data. Engaging in such an exercise is helpful to consider mobility data and its analysis as part of mobility information infrastructures. These infrastructures include broader institutional infrastructures that make mobility research possible (think servers, data specialists, multi-origin data sources) and political and economic infrastructures that influence research priorities.

As with many other fields, mobility data analysis used to grapple with relative data scarcity. New data collection and processing capabilities have transformed how mobility research is done. More and more actors interested in technological innovation and monetising data are entering the field. Such changes bring opportunities and concerns. These require a broader engagement with ethics – with what it means to do good. The lure of innovation itself can be a concern as it often trumps a real need for innovation. Use cases are frequently found after the fact. Given economic interests that undergird the field, it is not immune to a mentality of “move fast and break things” that leaves little room for considering ethics and long-term implications of new mobility data ecosystems. The concerns raised in our discussion centred around questions of consent, equitable access, reproduction of power asymmetries, and fundamental questions about who should benefit from mobility research how.

Our discussions involved thinking about issues linked to a digital divide, both in terms of data produced and the mobility tools that data is used to create. New ideas about mobility as a service, for example, are frequently thought about in terms of creating sustainability and efficiency. Still, social aspects require considering a diversity of stakeholders who might be disadvantaged by hyper-efficient and environmentally sustainable systems. City building always tends to happen both as a byproduct of natural growth and the engineering of cities – cities are arenas where there is an interplay between policy and new knowledge. At the level of ethics, this raises the question of what kind of city we want to live in. It seems complicated to imagine a truly ethically ‘optimised’ mobility data application without concerted efforts engaging with that question.

The question of third-use scenarios and the pivoting of services that might have been conceptualised with good intentions is another concern that looms large in the field. Beyond this, there is an increasing growth of hybrid applications that take advantage of mobility data information infrastructures. One prominent example is how developers might draw on mobility choice data for identifying investment areas – accelerating problems with already tight housing markets in many urban areas across the globe.

Questions of access and equity also matter for how mobility data becomes bound up in platforms and projections of how mobility should work in the future. For example, while visions for developing so-called EU “mobility data spaces” are framed as openly accessible platforms, de facto who registers to those spaces is limited, as technology knowledge and resources required to access such platforms are pretty high. This might perpetuate inequities as it consolidates the prominent role of large corporations and creates dependencies that may not necessarily be just or desirable.

There are different technological solutions envisaged to bridge these different access inequalities. For example digital identity management may offer inclusion opportunities also to people without legally recognised identification [1]. At the same time, a drive towards digital identity verification is often under scrutiny for their surveillance potential. In terms of facilitating access to data, debates hone in on interoperability standards or so-called mobility data marketplaces.

Privacy remains a significant value. Privacy and how its protection is interpreted requires critical evaluation. Issues that arise are those linked to reproducing prevailing power inequalities. For example, in debates on who ought to access and share what kinds of mobility data, customer data is often excluded as too sensitive. In effect, this means that this data stays with those who can harvest it. It also means that those with access to more detailed customer data have a competitive advantage over others as “innovative” mobility platforms grow. While it remains unclear how to counter this type of development, it is important to re-emphasise that debates about a multiplicity of principles – with privacy being one – have to continue to ensure that mobility research does not consolidate undesirable social patterns.

At the same time, innovations in the field are also shaking up who the relevant players are. Some startups – like, for example, Citymapper – might perform better at facilitating mobility for their subscribers than the transport operators themselves. This eventually also gave them a stake in the mobility debates, even though their practices might not fall under the same regulatory frameworks as public service providers. This is not to say that such changes are necessarily ethically suspect – various business models are emerging, some of which sincerely have social and ecological sustainability at their heart. Examples here might be the establishment of data trusts in some Canadian cities that conceptualise mobility data as a public good. Similarly, the ideals of Gaya X at the EU level are pointing us to a rethinking of priorities. Still, a prevailing pattern is strong incentives to cut corners to stay competitive.

References

- 1 McKinsey Global Institute. Digital identification: A key to inclusive growth, 2019. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/digital-identification-a-key-to-inclusive-growth>.

4.5 Mobility Data Analysis Ethics beyond humans only: Tracking animals and moral agency

WG scribe and other members: Alessandra Raffaetà, Bettina Berendt, Maria Luisa Damiani, Stan Matwin, Chiara Renso, Geoffrey Rockwell

License  Creative Commons BY 4.0 International license
© Chiara Renso, Bettina Berendt, Stan Matwin

In this group, we considered issues around the dignity and rights of non-humans when it comes to location data. We track, for example, land animals (individuals and herds) and fish, for all sorts of reasons. The location data can be used to study them or to hunt/fish them. This raises a number of ethical questions.

Rights from dignity. Animals with dignity implies the existence of rights for such animals. For example, pets are considered by many to have dignity, so there are laws that punish their mistreatment or abandonment. We can say that rights are proportional to dignity: more dignity, more rights. It is also a matter of relationships. You are tied to your cat or dog so you recognize that they have some rights.

Rights from suffering. The philosopher Peter Singer [1, 2] states that all beings capable of suffering have to be worthy of equal consideration and that giving lesser consideration to beings based on their species is no more justified than discrimination based on skin colour. He argues that animals' rights should be based on their capacity to feel pain more than on their intelligence. Animals can suffer so they should have rights, hence they should be protected.

What about fish? We do not consider fish as individuals but as species. Our attitude is to take care of the species' survival in order to maintain the ecosystem. We have an utilitarian approach with respect to fish. We want to avoid over-exploitation because this can provoke loss of work (for the fishers) or loss of food for some populations. In this context, domain experts like ecologists are useful to establish policies for sustainable ecosystems; such policies are utilitarian rather than tied to the deontological notion of dignity.

What about insects? We usually kill mosquitoes, and in agricultural ethics it is argued that individual insects do not have a "right to life." ⁷ Animal ethics depend on culture and religion. Let us think about Buddhists or Jains – they try to do no harm to *any* animal.

Issues concerning research and animals

We also discussed issues around the ethics of research and animals.

Tracking animals. We track animals to understand their behaviour; is that ethical or not? We use datasets reporting the movements of animals assuming that they do not have any privacy/ethical concerns. Is it really true? Lennox et al. [3] highlight the risks associated with animal tracking. In fact, bad actors, such as poachers or also photographers, could intercept animals directly using tracking hardware or indirectly from research results, databases and maps that provide the positions of vulnerable animals, and they could chase or simply disturb them. Thus, uncontrolled access to such location data may ultimately compromise the welfare of wild animals and the recovery of species. In [3] some techniques are discussed to protect animal data from misuse, such as data blurring, data aggregation, and data hiding.

⁷ https://en.wikipedia.org/wiki/Insects_in_ethics

However, it is worth noticing that anonymization approaches used with human data are less useful in this context because the identity of an individual animal is rarely important. A good trade-off to delivering research data could be providing time-varying density location data extracted from trajectories. In this way, the whole trajectories are not shared and they are protected.

Bridging movement ecology and human mobility. Movement ecology is a relatively new discipline in the field of ecology that studies the spatio-temporal patterns and processes at the basis of animal movement [4]. While data and analytical methods are similar between movement ecology and human mobility, there is surprisingly little interdisciplinary awareness of these similarities. Developing new methods to integrate human mobility data (e.g. road traffic or human recreational activities) in the study of movement ecology would crucially improve the ecologists’ understanding of the tight relationship between animal movement and human activities, as for example to unveil the effect of COVID-19 human lockdowns on animal movement and behavior. The vision of an integrated science of movement bridging the two research fields is however hampered by the limited availability of open data on human mobility.

Sustainability of research. As a more general issue, we need to reflect on what research looks like, how it is conducted, and also what its impacts on the environment are. To what extent should we consider not only humans and animals, but also ecosystems and the planet as a whole as moral agents (or patients)? It is important to consider alternative or hybrid formats of conferences that are more inclusive and ultimately more sustainable. In fact, having virtual conferences can allow researchers who are excluded for geographical and financial reasons from participating and at the same time the carbon footprint is reduced [5].

References

- 1 Peter Singer. *Animal Liberation*. Harper Collins, 1975.
- 2 Peter Singer. *In Defence of Animals: The Second Wave*. Blackwell, 1985.
- 3 Robert J Lennox et al. A Novel Framework to Protect Animal Data in a World of Ecosurveillance. *BioScience*, 70(6):468–476, 2020.
- 4 Federico Ossi, Fatima Hachem, Francesca Cagnacci, Urska Demsar, and Maria Luisa Damiani, editors. *HANIMOB@SIGSPATIAL 2021: Proceedings of the 1st ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility, Beijing, China, 2 November 2021*. ACM, 2021.
- 5 Chelsea Miya, Oliver Rossier, and Geoffrey Rockwell. *Right Research: Modelling Sustainable Research Practices in the Anthropocene*. OpenBook, 2021.

5 Online Interactions in a COVID-19-era Dagstuhl Seminar: Design, Experiences, and Reflections

Bettina Berendt (TU Berlin, DE)

License  Creative Commons BY 4.0 International license
© Chiara Renso, Bettina Berendt, Stan Matwin

This three-day Dagstuhl Seminar took place in January 2022, which for many participants was month 23 of a string of more or less harsh COVID-19 lockdowns and home-office conferences. The seminar had originally been planned as an on-site and then as a hybrid workshop. Given the skyrocketing COVID-19 infection counts throughout the world and the suddenly quickly increasing number of participants who were forced to or chose to participate online, we decided, in the week before the seminar, to hold it fully online. We were quite concerned

about this decision, given the stark contrast between many participants' fond memories of the interactivity of pre-COVID Dagstuhl Seminars and the growing Zoom fatigue that we had been witnessing in ourselves and in our colleagues. We therefore contemplated what we could do to make the most of our three days together.

We (the organisers) were happy about the results of our strategy, and we received very positive feedback from several participants. Of course, our choices could not fully reproduce the experience of an on-site Dagstuhl Seminar. However, we believe that online and hybrid meetings have many advantages and will be an important part of how we all will be working in the future. Therefore, in this section, we want to share our experiences and observations. We hope that others can profit from them.

5.1 What do we consider “success”?

There are many possible measures of a Dagstuhl Seminar's success. Based on our personal experience, we especially hoped for (a) academic output (especially of the type(s) aspired to by the participants of the seminar), (b) a pleasant working atmosphere: interactive, inclusive and with the right balance of focussed academic exchange, loose brainstorming, and everything in between, (c) meeting new colleagues or interacting with known colleagues in new ways, so as to enter into meaningful conversations (which often begin as loose brainstormings, often about personal matters), and (d) retention of participants over the course of the seminar.

We believe that we have met goal (a) (with, for example, an op-ed in *The Conversation* [1] about seminar topics published two weeks after the seminar's end; other activities are ongoing). This is related to the self-reported goal of a majority of participants at the beginning of the seminar (“an extended paper abstract”). All participants who gave us feedback expressed (b), and this is in line with our (the organisers') own impressions. We did not ask participants about (c) but can confirm this to be true for ourselves.

With regard to (d), most participants attended the seminar all day, which we found remarkable given our observations that project meetings and workshops tended to become shorter over the course of the pandemic. Some participants (especially senior ones) had to leave the seminar repeatedly for other duties, and we observed how hard it was for participants in substantially different time zones to participate in all sessions. However, due to the informal organisation this did not present serious problems, and in fact some participants who were not able to participate throughout still managed to be central for their working group's progress and outputs, even if in a partly asynchronous way. We are also especially impressed by and grateful to those participants from +7 hours to -8 hours time zones who participated in all sessions. So in sum, we believe we were reasonably successful along this criterion. Since we do not have statistics about participant retention in other online formats, it is even possible that comparatively speaking, we were very successful.

5.2 What did we do, and what did we learn from it?

We believe that four elements of interaction design and participation choices were key.

5.2.1 Zoom + Gathertown + jointly edited documents

A specific combination of two different types of videoconferencing software, in our case Zoom and Gathertown, proved very helpful for engagement. Our hypothesis had been that Zoom fatigue is not only a physiological effect and the lack of the full breadth of “real” interaction, but also very much a consequence of the types of interaction favoured by Zoom (or similar

VC systems that are essentially generalisations of bilateral video telephony): a meeting at a specific time in which there is one speaker at a time. Zoom breakout rooms allow for a restructuring of a meeting into smaller meetings, but again they create public fora for well-defined groups with set start and end times. Parallel conversation threads are possible (in private chats), but usually only with one other participant at a time, and they always carry a certain feeling of clandestine conversation. So in sum, Zoom appeared to us to be an excellent tool for presentations, for goal-oriented discussions in “work mode” (i.e. under strict time settings), and suitable for the informal interactions needed to get to know people (or get to know them better) at most in small groups in which trust is established through a common goal or between people who already know each other well. Importantly, group membership and partaking in a conversation are binary choices, marked by entries and exits that are rendered very visible by the software.

Some of us had attended work meetings in Gathertown before. Gathertown allows for a categorically different type of being in a group and partaking in a conversation. First, Gathertown employs a spatial metaphor: groups congregate in locations, and one can *see* a group (as a set of avatars, rather than a list of names as in Zoom) and *move* there (and be watched moving by those in the group and others) rather than be “teleported” as in Zoom. Membership is *gradual*: the closer one gets to another person, the better one sees them (and is seen by them), and the better one hears them (and is heard). By positioning one’s avatar at the “outskirts” of a group, one can signal a certain looseness of belonging, and if one does this, does not take part actively in the conversation, and then moves on, it feels a lot less abrupt and impolite than dropping out of a Zoom breakout room. In addition, Gathertown’s Minecraft-like environment may particularly appeal to people who enjoy video games, but also for others it offers gratuitous (but still potentially effective) signals of shared environment luxury: the cute watercooler, the plants, the sofa, rather than only the often bleak and distorted camera view of somebody else’s home, office, or virtual background. We considered these elements key enablers of the pleasant and informal and fluid conversations that are so generative of creative exchanges with (hitherto) strangers in physical Dagstuhl meetings. We also considered the *spatial* and *motion*-based interaction of Gathertown to be particularly suited to our topic of *mobility* data mining.

We were privileged to have a colleague who is an expert of Gathertown designs and was enthusiastic to help us: we are incredibly grateful to Adem Kikaj of KU Leuven – his work has been a major factor in making this Dagstuhl Seminar a success! We are also grateful to people from whom we learned interaction techniques (in former Dagstuhl Seminars as well as in the planning of the current one). We designed an action model for the seminar and a spatial manifestation of this action model (with combinations of frontal presentations, fishbowl interaction, posters that people would be able to come back to to remember other participants’ introductions of themselves and have bilateral or small-group discussions, etc.). In line with the real Dagstuhl experience, we also designed a lobby (with small seating groups, implemented as Gathertown private spaces, to mimic the availability of various meeting rooms in Dagstuhl castle) and a dining hall. We also planned a sports room for joint gymnastics exercises, a feature one of us had seen and appreciated as an activity during breaks in a different conference, and a nod to Dagstuhl’s sports equipments/rooms. Adem built this world for us and enabled us to apply the finishing touches (which would only be possible during the seminar and/or live) ourselves.

Another core element in the media mix were jointly editable documents and folders. We used Google docs/folders for the landing page, the participant-to-working-group assignment, shared resources such as participant self-descriptions and tutorial slides, and working group results (each working group had one folder that they were free to fill).

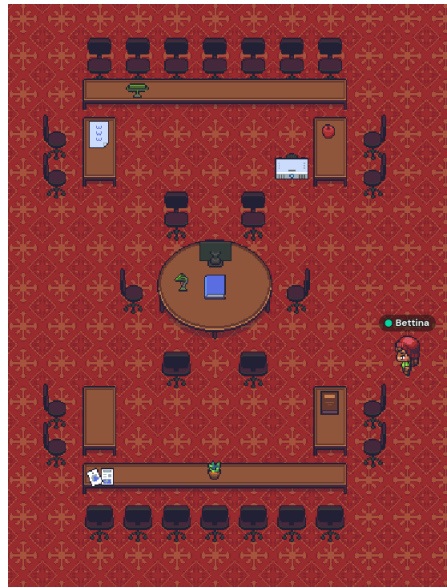


■ **Figure 1** Gathertown: entrance area (map view).

The services themselves of course come with their advantages and disadvantages. We chose two services that are widely known and easy to use (Zoom and Googledocs) and one that, as far as we know, is the only one to provide a certain type of interaction (Gathertown). Familiarity is, in our experience, important, but it does raise questions with regard to providers and their (e.g. data protection) policies. Future designs could prioritise other features more (such as open source or provider characteristics).

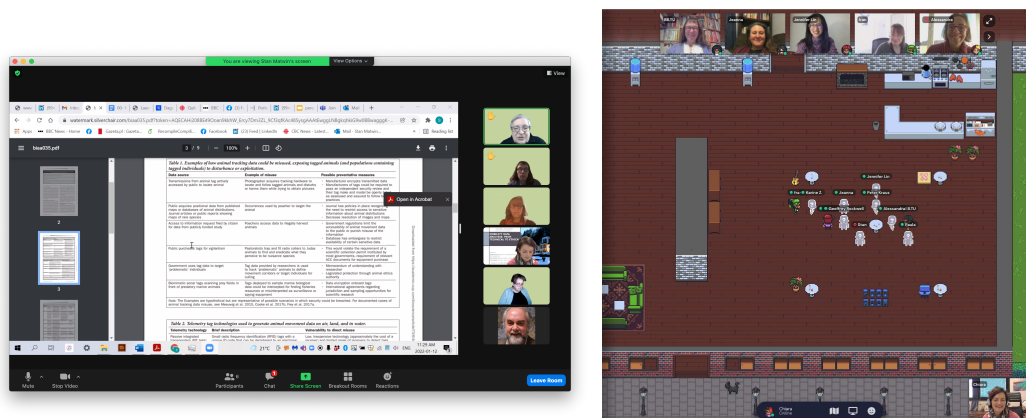
5.2.2 Less is more

As explained above, the Gathertown world was quite elaborate and beautifully supportive of different types of planned interactions. In order not to overwhelm participants, we however decided to do the first day fully in Zoom and with Googledocs (for familiarity) and gave participants a first tour of the Gathertown world only during the “reception”, an extra one-hour slot after the close of the official “work” programme. Many (but not all) participants accepted this “invitation”, and we did our best to make this a tour, picking people up at the entrance (and returning to meet individual latecomers), showing them around, explaining the functionalities, etc. Observing their interactions, we (the organisers) quickly decided to use Gathertown for the breaks only and keep the “work” programme in Zoom, and we enforced neither the use of all rooms in the Gathertown world (for example, we supplanted the originally planned poster-room by a simple folder on Googledocs) nor of all functionalities (such as shared whiteboards, private spaces, and the fishbowl). We also dropped the idea of our gymnastics offer, instead only making it known that we’d be in the Gathertown for the “morning coffee” (half an hour before the official start every day), the two coffee breaks and the lunch break, and the “reception”. The “morning coffee” break was taken over by one of



■ **Figure 2** Gathertown: room designed for supporting fishbowl discussions (partial view).

us who likes getting up early, and the “reception” by one who tends to work later during the day. In sum, an even simpler Gathertown world would have sufficed (and may have required less bandwidth).⁸



■ **Figure 3** “Work” and “break”.

We believe that this adaptivity to the actual interactions was important and prevented a sense of over-structuring that we had observed in earlier online events we had taken part in, a structuring which may contribute to work productivity but can feel stifling at times.

There was one element that we *added* (while all other adaptations were removals): one participant reminded us of the Dagstuhl tradition of randomising (but then enforcing) seating at meals. We implemented this with a simple hack that illustrates that complex Gathertown

⁸ We learned after the seminar that a previous Dagstuhl collaborator created a virtual replica of Dagstuhl itself. While we were intrigued by this and probably would have used it had we known about it, based on our experiences, we believe it is not functionally necessary and may even prove somewhat overwhelming. This question should be explored in future work.

spaces can also be pragmatically adapted under time pressure: we put different ready-made objects (a plant, a guitar, ...) next to the different tables, created a random assignment of participants to tables, wrote this on the landing page, and showed it just before the lunch break, asking people to go to their assigned table. These arrangements proved to not work as deterministically as in the physical Dagstuhl, but they did create the same effect: unexpected gatherings and chitchat over lunch with participants one had not known before.

These little tricks encouraged also timid people to interact more with other participants they did not know before. It was not as effective as in physical Dagstuhl, since some people preferred to take a real break from the online event and also had to prepare their food, but a start.

To future organisers who consider using meeting software that is not widely known, we recommend trying out an interaction space such as Gathertown intensively. We did this in a “dress rehearsal” with two organisers and two participants in the week before the seminar, but at the time, we did not anticipate the dining seating arrangements. It was only during the breaks that we realised we had positioned the “tables” too close to each other relative to the visibility/audibility radius in Gathertown, such that conversations at different tables mixed in a confusing manner, and we had to artificially “move into the corner behind the table” to avoid disturbing the neighbouring table. This was amusing rather than really annoying, but we would certainly improve this design element next time.

As a consequence, there was one great simplification: Conceptually, there was one space for “work” (Zoom and shared documents) and one space for “breaks” (Gathertown). Within the “breaks” space, only the two rooms near the entrance (dining hall and lobby) were actually used, but this may have been a consequence of the number of users and the space being sufficient for them. The two organisers who were responsible for “hosting the breaks” complemented this conceptual separation by a technical one, running the “work” space on one computer and the “breaks” space on a different one. This improved audio and video quality considerably and made it easier to mentally control and survey the spaces. One participant remarked on this clean separation by relating their experiences from another online event, which employed a very elaborate Gathertown world with dedicated spaces for every activity. They had found that too complicated and overwhelming and welcomed our simplicity.

5.2.3 Structure + flexibility

We combined a simple outer structure with guidance on the one hand with freedom in the details of how working groups were to function. The outer structure consisted of (i) a consistent timetable for all three days, (ii) a rough specification of the functions and expected outcomes of each half-day slot, (iii) a self-presentation round in which each participant was asked to fill in the same Powerpoint slide consisting of three questions to present themselves, with slides sent to the organisers ahead of the seminar,⁹ (iv) a tutorial round in which three relevant angles on the seminar topic were introduced by experts in the respective fields (see Section 3), (v) working-group formation and reporting rounds in the plenum, and (vi) working-group sessions (see Section 4).

⁹ The slide asked for: name, affiliation, “my question” (that I would like to see answered at the seminar), “expected seminar output” (4 choices + free-form), and “an image of an object that symbolizes your interest, motivation, curiosity, doubts, ... regarding this seminar”. All templates that we believe could be helpful for future seminars can be downloaded at <http://www.master-project-h2020.eu/dagstuhl-materials/>.

Media use was kept simple: one landing page (a Googledoc that was adapted during the course of the seminar) which pointed to all other resources, one Zoom main room which was always open and either in use by at least one person or equipped with a slide that indicated the current status (such as “lunch break”), and the Gathertown world which was also always open. A first version of working-group topics was created by the organisers clustering participants’ answers to a question about the personal goal for the seminar (included in the self-presentation slide), but then modified in an open group discussion. This as well as the assignment of people to working groups used a jointly edited spreadsheet.¹⁰ Sessions for five working groups were planned adaptively to also take into account when participants in American time zones would be able to attend (better). The working groups were numbered and met in Zoom breakout rooms of the same number, which were only opened during the times previewed in the time plan. Participants were encouraged to also meet with their working groups or with others outside these slots, and the use of Gathertown for such meetings was explained. This last option was however, as far as we know, not used much, since the “break” times were needed for relaxation, and remaining online for socialising after a full day of online work was not really an option for most participants.

One organiser monitored the Zoom main room throughout and another one did so in the Gathertown world. During working-group sessions, this presence was a secondary activity from a different computer. This ensured that participants who arrived at an unusual time (e.g. due to other commitments or time zones) were never “alone” and could always be briefly informed of current activities and locations.

Working groups were free to choose any type of media they wanted and any type of reporting, and other structures (such as the dining seating arrangements) were not enforced. This flexibility was used and appreciated by participants.

Visual consistency was ensured by a graphics expert’s providing slide templates for self-presentation and organiser-provided slides: we are very grateful to Beatrice Rapisarda of ISTI/CNR, who also created the “group photograph” included at the beginning of this report from a collection of screenshots we had taken throughout the seminar.

5.2.4 “I am not here”

Many of us had experienced, over the course of the pandemic, a “densification” of work: an increasing number of work meetings from the home office, with meetings often taking longer than previously and positioned back-to-back. Attendance at conferences, which now involved no physical travel, was regularly disturbed by the assumption of colleagues that all the usual work could be done in parallel. Many senior colleagues we spoke to concluded that at some point, they just stopped going to conferences because it was “too much” or “not worth it”. We believe that this overload contributed at least as much to Zoom fatigue as the physiological factors and the restricted interaction mentioned above.

Without having coordinated this in advance, we found that (at least) two of the organisers and one other participant had independently decided to break this pattern for this seminar. They had told their colleagues and superiors that they would be “at” Dagstuhl and not

¹⁰ Interestingly, the spreadsheet was chosen by participants over an initially planned process involving spatial movement that one of us had observed as an efficient group formation process in earlier Dagstuhl Seminars. We had envisaged a replication of this technique in the Gathertown spaces, but after one participant proposed the spreadsheet, everybody immediately agreed that this would be the best method. It supported the same functionalities: simultaneous and mutually influencing self-assignment and observation of other participants’ choices.

available during the three days of the seminar, and severely restricted their email monitoring and processing. One participant from the Americas had even put themselves into the Dagstuhl time zone, ensuring support from their family for this “absence”. While the three days were still physiologically very taxing, all three of us agreed that this had been very sensible choices, needed for sustainable participation in future seminars, conferences, etc.

5.3 A sense of place in a hybrid world – and other parallels between the medium and the message

One element stands out from these observations, and it echoes the seminar’s themes. Media use and interactions provided a “sense of place” that helped structure both (a) a *prima facie* non-spatial online world and (b) the hybrid nature of contemporary work. The unique and permanent “entrance points” (the Web-based landing page, the main Zoom room, the Gathertown world) provided stability and simplicity. The clear association of Zoom with “work” and Gathertown with “breaks” (or, in some cases, also corridor talk in parallel to the official working group meetings) helped participants to focus respectively relax. Gathertown’s fluid interaction supported informal chats. Once inside a space (in particular in the breakout rooms), individuals and groups had all the freedom of acting there. The additional choice made by some participants to effectively “remove” themselves from their main place of work, in order to be fully at the seminar place, proved beneficial.

The sense of place echoes the *mobility* part of the seminar’s themes, *Mobility data analysis: from technical to ethical*. We believe that the other key elements were also reflected: the *technical* obviously in the explorations of online-media design and uses, and the *ethical* through our general approach to research design, which we base on an ethics of care. This involves the recognition of relationships between people as at least as fundamental as autonomy and a desire to create, sustain and honour such relationships in research design. This meta-notion of care and ethics, first explored by the author of this section – together with two participants of this Dagstuhl Seminar – in [2], informed the choices described in the present section. In this sense, the present section is not only an exploration of different videoconferencing systems, but also a complement to the section author’s work as the MASTER project’s Independent Ethics Advisor.

References

- 1 Geoffrey Rockwell, Bettina Berendt, Florence M. Chee, Jeanna Matthews, Sébastien Gambs, and Chiara Renso. Ottawa’s use of our location data raises big surveillance and privacy concerns. *The Conversation*, 2022, Jan 27. <https://theconversation.com/ottawas-use-of-our-location-data-raises-big-surveillance-and-privacy-concerns-175316>.
- 2 Todd Suomela, Florence Chee, Bettina Berendt, and Geoffrey Rockwell. Applying an ethics of care to internet research: Gamergate and digital humanities. *Digital Studies/le Champ Numérique*, 9(1), 2019. <http://doi.org/10.16995/dscn.302>.

6 Conclusions

Chiara Renso, Bettina Berendt and Stan Matwin

License  Creative Commons BY 4.0 International license

© Chiara Renso, Bettina Berendt, Stan Matwin

We believe that this seminar represents a successful first step in building a community of scientists around the mobility data ethics theme. The five topics that we identified and discussed in the Working groups are stepping stones from which the community can extrapolate new research topics. Indeed, at many points of our three-day journey, we realized that more research and reflections are needed to properly address these issues. For example, many discussions in the working groups revolved around drilling down into and challenging points that had been topics of the tutorials.

We also saw, in the tutorials as well as the discussions and through the interest in our work that greeted us at The Conversation Canada and in reaction to the article we published there directly after the seminar [1], the extent to which the general public perceives the urgency and importance of the issues we discussed, and the extent to which they expect explanations, answers, and better technology from the research community. It is our responsibility and aim to continue to provide these.

We would like to close with conclusions for future work on interaction design. After the second COVID-19 winter and with political choices worldwide geared towards making us “live with COVID”, all of us find ourselves in a transition to working routines that will likely involve more online interactions than in the past. Some of us believe that online meetings are the future, some of us believe that we should also invest more time and energy into creating successful hybrid meetings, which present additional challenges. Others are more cautious and long for a return to “normality”. Most believe in the need to find a new balance, such as obviating extensive travel for one-day meetings but travelling for longer meetings. (These discussions were also had at our Dagstuhl Seminar.)

We believe in the virtues of online meetings, at the minimum because they promise more inclusiveness (e.g. of mobility-restricted or otherwise vulnerable people, individuals without access to travel funding, persons with care duties, etc.) and are environmentally more sustainable. We also agree that hybrid meetings are particularly challenging, in particular when it comes to forming working groups that involve both on-site and online participants. But through the experience of this seminar, we also understood better that even so-called “fully online” meetings are in fact hybrid – in the sense that every participant remains engulfed in their home-office and/or regular-place-of-work spatial spheres. Creating a “sense of place” that is psychologically, physiologically and socially sustainable, will be a necessary element of future meetings, whether they are “fully online”, “hybrid”, or “on-site” in standard terminology. The sometimes-used term “offline” may actually be a key element for disconnecting from an overload of simultaneous duties. The simple and often ad-hoc uses of existing technology towards these ends that we have described in this text may have worked well in this particular setting, and others will be needed for different settings. But we hope that our lessons learned can serve as an inspiration to others on this trajectory.

Acknowledgements

This work is supported by the MASTER project that has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie-Słodowska Curie grant agreement N. 777695. Bettina Berendt thanks the German Federal Ministry of Education and Research (BMBF) – Nr. 16DII113f for funding for this seminar.

References

- 1 Geoffrey Rockwell, Bettina Berendt, Florence M. Chee, Jeanna Matthews, Sébastien Gambs, and Chiara Renso. Ottawa's use of our location data raises big surveillance and privacy concerns. *The Conversation*, 2022, Jan 27. <https://theconversation.com/ottawas-use-of-our-location-data-raises-big-surveillance-and-privacy-concerns-175316>.

Participants

- Darren Abramson
Dalhousie University, CA
- Christine Ahrend
TU Berlin, DE
- Bettina Berendt
TU Berlin, DE
- Florence Chee
Loyola University Chicago, US
- Thierry Chevallier
Akka Technologies, FR
- Maria Luisa Damiani
University of Milan, IT
- Josep Domingo-Ferrer
Universitat Rovira i Virgili –
Tarragona, ES
- José Antônio Fernandes de
Macedo
Universidade Federal do Ceara –
Brazil, BR
- Sébastien Gambs
University of Montreal, CA
- Ioannis Kontopoulos
Harokopio University –
Athens, GR
- Peter Kraus
European Data Protection Board
– Brussels, BE
- Fen Lin
City University –
Hong Kong, HK
- Jeanna Matthews
Clarkson University – Potsdam,
US
- Stan Matwin
Dalhousie University –
Halifax, CA
- Fran Meissner
University of Twente, NL
- Anna Monreale
University of Pisa, IT
- Francesca Pratesi
ISTI-CNR – Pisa, IT
- Alessandra Raffaetà
University of Venice, IT
- Chiara Renso
ISTI-CNR – Pisa, IT
- Paula Reyero-Lobo
The Open University –
Milton Keynes, GB
- Geoffrey Rockwell
University of Alberta –
Edmonton, CA
- Yannis Theodoridis
University of Piraeus, GR
- Konstantinos Tserpes
Harokopio University –
Athens, GR
- Karine Zeitouni
University of Versailles, FR

