Executive summary

The workshop produced a number of key outputs. A series of talks were delivered during the first day of the workshop, showcasing ongoing and state-of-the-art work on human mobility and behaviour, and the use of mobility data in urban science. Talks were delivered by researchers from six different institutions, including ISI Foundation, MIT, ETH, University of Leeds, University of Liverpool, UCL from four different countries. Presentations were recorded and are available on the GDSL YouTube channel.

Key reflections were obtained from discussions with 13 scholars with expertise in the use of mobility data during the second day of the workshop. First, participants identified key tools, methods and common practices in the use of new forms of data to capture human movement. Key issues around data access, validation, privacy and practises were identified, and potential solutions to address these issues were discussed. Second, a broad range of applications for mobility data were identified in the context of public health, urban transport and infrastructure planning, climate change and urban density patterns. Finally participants discussed key areas of future endeavours and potential collaboration, thinking of ways to exploit the spatial and temporal granularity of new forms of data, enhance and improve data access, privacy and practises, and the development of software and tools to handle and analyse the complexity of the data sources.

Participants also expressed a strong desire for closer integration and collaboration through formal networking activities. The Turing's urban analytics programme could potentially become a catalyst for providing a productive environment to foster these activities.

The remainder of the report is structured as follows. First, the background and objectives of the workshop are presented. Second, abstracts of the 6 talks held during day 1 of the workshop are shown. Third, the three discussion sessions that took place on day 2 are summarised. For each discussion session we report on the aims of the session, the questions that guided the discussion and the main points that emerged discussing each question. Finally, the list of participants and the full workshop agenda can be found in the appendix.

Edited by

Francisco Rowe
Senior Lecturer in Quantitative Human Geography, Geographic Data Science Lab, University of Liverpool

Alessia Calafiore
Postdoctoral Research Associate, Geographic Data Science Lab, University of Liverpool

Krasen Samardzhiev
PhD student, Geographic Data Science Lab, University of Liverpool

Martin Fleischmann
Postdoctoral Research Associate, Geographic Data Science Lab, University of Liverpool

Daniel Arribas-Bel
Deputy Director of Urban Analytics Programme, The Alan Turing Institute and Senior Lecturer in Geographic Data Science, Geographic Data Science Lab, University of Liverpool

Jacob MacDonald
Lecturer in GIS and Spatial Analysis, University of Sheffield
Background

Data have become a central pillar of society. Technological advances in computational power, storage and network platforms have enabled the production, processing, analysis and storage of large volumes of digital data. Information that previously could not be stored, or captured can now be digitally recorded. Digital data have become a key asset for government, businesses and individuals supporting their decision making processes and shaping human behaviour. A notable example has been the use of digital data to monitor the COVID-19 pandemic and inform the development of appropriate interventions.

A key data stream has been location data from mobile phones. These data have enabled monitoring the geographic spread of COVID-19 in near-real time with technological companies, such as Apple and Google releasing regular mobility reports. More generally, mobile phone data are a rich source of information offering a unique opportunity to capture human behaviour at an unprecedented geographic and temporal resolution. Yet, key challenges remain, such as issues about privacy, representativeness, biases and the use of large, noisy and complex data sets.

The report seeks to summarise the views of the participants of the “Using mobility data in urban science” online workshop on 27-28 October 2021, organised by the ITINERANT project. ITINERANT (InequaliTies IN Experiencing uRbAn fuNCTion) is a project funded by The Alan Turing Institute within the urban analytics programme and is led by a team of researchers at the Geographic Data Science Lab (GDSL) at the University of Liverpool. ITINERANT investigates how different populations experience and benefit from urban functions by leveraging mobility data at high spatiotemporal resolution.

The report has been edited by the ITINERANT team, a group of researchers with background expanding Artificial intelligence (AI), geographic data science, urban analytics and human mobility. The workshop aimed to bring together expert scholars in the use of new forms of data; provide a snapshot of their use to understand human mobility; and, identify common practises, methods, applications, tools and challenges.

The ITINERANT team felt there is a growing interest in the use of this data and the need to bring together experts to start developing best practises, established approaches, and sharing of know-how is becoming apparent. The workshop was organised as a part of The Alan Turing Institute’s urban analytics programme seeking to enhance their strategic position as a hub for expertise on data science and AI.

Objectives

We approached the workshop with the overarching goal of creating synergies and contributing to the growing community of researchers and practitioners in the area of new forms of mobility data. With this in mind, our key goals for the workshop were:

- To bring together leading scholars working with new mobility data in urban science to discuss ongoing, state-of-the-art research in this area;
- To contribute to building a community of practice around new mobility data, particularly strengthening links between ongoing projects at the Turing, and forging closer collaborative links;
- To pool different views, experiences and ongoing work in an effort to avoid duplication of work and unrealised synergies;
- To increase awareness of ongoing work exploiting the use of mobility data at the Turing and more widely across the academic community;
- To start a conversation around common knowledge, established best practice standards and agreed techniques to work with new forms of mobility data;

The workshop has evidenced the interest and scope of research on the use of new forms of data to understand human behaviour.
Workshop day 1 – presentation talks

The first day of the workshop consisted of a set of talks that are made available on the GDSL YouTube Channel. Six speakers were invited and what follows are the abstracts of each talk:

Charisma Choudhury (Institute for Transport Studies, University of Leeds)

Utilising passive data sources for predicting demographics
A major limitation of the passive data sources like smart card, GPS and mobile phone data is they are anonymous. But in mathematical models of travel behaviour, it is important to identify users who have similar sensitivity to level-of-service attributes like travel time and cost - so that we can use them to predict the demand levels in future scenarios with new modes and services. The presentation will feature two case studies where we harness passively generated mobility data (smart card and mobile phone call detail records) to cluster similar people using their historic travel behaviours and phone call patterns.

Laetitia Gauvin (ISI Foundation)

Inequalities in human mobility
Human mobility patterns in daily life and during crises reflect diverse socio-demographic factors. Mobile phone data analysis turns out to be a precious way to explore the multitude of such mobility patterns. We present two studies looking at inequalities in human mobility using mobile phone data: 1) how mobility is gendered in everyday lives 2) what have been the socioeconomic drivers of mobility during the first wave of the COVID-19 pandemic. Our first work uncovers a complex interplay between gendered mobility patterns and socio-economic factors. In the second work we explored the relationship between the behavioural responses to mobility restrictions and economic factors such as local structure of the labour force, or employment rate.

Esteban Moro (MIT-MediaLab/Universidad Carlos III de Madrid)

Understanding urban resilience through behavioural mobility data
Two thirds of the world's population will live in cities by 2050. The economic and social progress of our urban areas, our institutions and our work depend on the diversity and resilience of the social fabric in cities. Despite their importance, several major forces erode the diversity and strength of those social connections: from income or racial segregation to differences in education and work access. In this talk I will present our recent work to understand the fragility of the network of social connections in cities through the analysis of behavioural mobility data and its impact in experienced segregation.

Alessia Calafiore and Francisco Rowe (Geographic Data Science Lab, University of Liverpool)

ITINERANT – InequaliTies IN Experiencing uRbAn fuNcTion
ITINERANT is an ongoing project funded by The Alan Turing Institute that will develop methods, open data products, and policy-relevant insights on how differently population groups benefit from urban functions. Such understanding is key to develop policies that ensure these benefits are fairly shared across society. We will give an overview of the project aims and the methods we are planning to implement as well as some preliminary results.
Yanan Xin (Mobility Information Engineering lab, ETH Zurich)

Computational Movement Analysis for Sustainable Mobility

The increasing CO2 emissions in the transport sector and rapid expansion of road networks pose pressing issues to the sustainability of our society. A key to transit into sustainable mobility involves a deep understanding of human mobility patterns and the development of novel methods to analyse, predict, and visualise human mobility. This presentation will demonstrate the use of computational movement analysis to promote sustainable mobility through our studies conducted at the Mobility Information Engineering Lab at ETH Zurich. The primary focus is to share insights on the impact of new mobility concepts (e.g., Mobility as a Service), user mobility behaviour changes towards sustainable travel, and the potential benefits of applying mobility-aware smart charging for electric vehicles. The talk will also highlight the computational methods and the open-source package developed for conducting the spatiotemporal analysis of mobility data.

Maurizio Gibin (CDRC/UCL) and Jacob Macdonald (CDRC/GDSL-University of Liverpool)

Footfall and in-app mobile phone data research at CDRC

This talk will present new sources of human mobility data available for research purposes through the Consumer Data Research Centre (CDRC). Two data products related to human movement include measures of footfall and mobile phone mobility data. The presentation will present the datasets available for request through data.cdrac.ac.uk, along with the development of new research and validation ongoing related to the use of reliable, granular, human mobility measures.
Workshop day 2 – discussion sessions

Session 1: Methods and tools

Aims and session framework

This session focused on two key aspects of working with new forms of mobility data. First, we wanted to leverage on the collective knowledge and experience of the participants to identify the key challenges of working with new forms of mobility data, and to collect ideas about the state of the art in terms of methods to deal with such challenges. Second, once challenges and methods were identified and discussed, we examined what are the available tools that implement methods to deal with mobility data.

Guiding questions

The attendants were asked to respond to the following two questions:

Q1) What would you say are the main challenges of working with new forms of mobility data (i.e. data bias/representativeness, privacy concerns, mode of travel inference, individual trajectory reconstruction) and what state of the art methods would you consider more promising to address them?

Q2) What tools (platforms, software, libraries, etc.) would you consider essential to work with mobility data? Please identify the tasks you would use these tools for.

Outcomes

Question 1

Challenges in mobility data span different and intertwined aspects: from data quality to methodological, technological and policy issues.

One of the first issues encountered working with new forms of mobility data is related to access. Accessing new forms of data in general is non trivial both from an administrative and technical point of view. Specific agreements with the data providers are required and not so easily obtained; even when access is given, documentation is not always completed and clear, i.e. information on what the sensors are capturing and what pre-processing steps have been already covered is often not detailed.

Preserving privacy is a widely recognised challenge when working with mobility data. From the discussion one avenue which was considered to be promising in addressing this problem was the use of synthetic data. Synthetic data can be obtained through simulations or Generative Adversarial Networks; however, more research is needed to assess the extent to which synthetic data are effective in preserving privacy and how to properly integrate data and simulations. Relatedly, the importance of the Trusted Research Environment service to get secured access to data and the accredited researcher training has been highlighted.

The challenge of finding a ground truth to validate mobility data was also discussed in almost all groups. Surveys are often used as ground truth, however, unless accurately designed or adjusted and with a high number of respondents, they could also be biassed and therefore not providing a valuable source to validate mobility data. Calibrating mobile data indicators from multiple Mobile Network Operators could increase the reliability of these data, as well as combining mobility data with administrative data. A further complexity to find ground truth data could also emerge when data are labelled based on activity types/transportation mode. Overall, it is important to be aware of and explicitly state the data limitations. The relevance of addressing the ground truthing problem is due to the common presence of biases in mobility data. Data coverage is often not complete and there can be spatial and temporal errors / inaccuracies as well as GPS collection errors; therefore, analysing
missingness can help detect how the dataset is biased. Different datasets hold diverse types of bias and it would be important to develop frameworks that provide different adjustments depending on the specific bias characterising the dataset. Small area estimation methods anchoring mobility data to official statistics could help address the problem of bias. Also, some groups highlighted the importance of interaction across different research communities, i.e. computer science, transportation research, social science etc. as a path for innovation and disruptive solutions.

A topic that emerged across all discussion groups is the general lack of standard practices to analyse human mobility data, including the use of a common terminology. This creates several problems: from the reproducibility and comparability of the results, to limiting a common assessment of the methods employed. The participants also acknowledged that standardising practices when working with data provided by private companies and having remarkably diverse formats is more challenging. However, trying to converge towards a more uniform framework to pre-process mobility data is unanimously recognised as favouring the research community. It was also widely acknowledged by the participants that open science, reproducibility and open source software can significantly help moving towards more integrated practices.

**Question 2**

A variety of tools was discussed to deliver tasks related to:

- **Simulation**
  - ABM: Java (Repast, MASON), Python (Mesa), Netlogo, MATSIM (transport focus)
  - Traffic Flow/Transport simulation software: SUMO (Simulation of Urban MObility), PTV Vissim, Aimsun, MatSIM

- **Analysing Trajectories**
  - Python: scikit-mobility, MovingPandas(GeoPandas), Trackintel
  - Storing: MobilityDB (PostGIS)
  - Network analysis
  - Modelling: elastic net/xgboost, PuLP
  - Bayesian: Latent Dirichlet Allocation Topic Models

Overall, it was noted that more integration among different tools is needed, starting from a common terminology. Likewise, bridging tools/approaches used across different research communities, i.e. simulations/trajectories analysis can also be beneficial. Finally, the groups noted that existing tools still do not address all challenges that mobility data present, i.e. visualisation, data fusion.
Session 2: Application and policies

Aims and session framework

For this session, attendants were asked to focus on the broader role that mobility data can have on public policy. Specifically along the lines of what domains are most implicated, where these data could have the most impact going forward (especially considering the change in landscape since the onset of the pandemic), and how can research using these data be impactful and responsive to addressing social or public policy challenges.

Guiding questions

Participants were asked to discuss and comment on the following two questions:

- **Q1)** What policy areas do you think can be more significantly impacted by research leveraging human mobility data?
- **Q2)** What do you think are the resources required to achieve research impacting the policy areas you identified above?

Outcomes

**Question 1**

A complex picture of policy domains and industry sectors were discussed, with participants noting that new forms of data have a wide reaching impact, direct and indirect, across a number of areas. This range of applications across a variety of disciplines speak to the value of these sorts of data.

One of the most frequently discussed topics was the use of mobility data for emergency response, preparedness, and its overall usefulness over the course of the COVID-19 pandemic. First and foremost, from a public health point of view, being able to explore where and when individuals may have been in contact, or identify large congregations of individuals, can help in monitoring transmission, outbreaks and target local interventions. Mobility data have been commonly used to explore how individuals have responded to local lockdowns and openings in the context of public health orders. Along with information concerning public health changes in working from home and new commuting patterns, can be captured by granular human movement data. At the micro scale - individual level mobility and activity data can also be used to inform on personal health outcomes such as obesity or chronic illness.

Another domain where new forms of data have shown their potential in informing policy-making is that of urban and transport planning. These data sources can enable a better understanding of human mobility patterns influencing (for e.g.) the planning of residential or commercial development in areas where there is accessibility to local amenities or, by better understanding city centre or high street use. In urban areas particularly, mobility data are able to provide some indication about the visits and use of specific areas or points of interest. This could include, visitor or tourist patterns to monuments and landmarks, retail centres, the use of transportation hubs, and all the dynamic links in between these areas.

Mobility data can help inform on a number of specific transportation issues including commuting choice alternatives between private cars, public transport or walking and cycling. This includes further the application of this data towards the related infrastructure (e.g. lights, roads, bike baths) and respective route planning.

It is also noted that energy efficiency and a sustainable use of resources can be made more of a priority with better data on who uses what, where, and when. Planners across a range of sectors can use mobility data to target their environmental and efficiency gains. This could help (for e.g.) in building more sustainable, walkable, neighbourhoods, or realising efficiencies in traffic planning for emissions reductions.

No matter which general policy area is being analysed, it is important that when using mobility data the limits and details of the data itself are discussed.
**Question 2**

The priorities discussed here revolved around staying connected, in communication, and in collaboration - and importantly, not only amongst researchers but with all local stakeholders, practitioners and policy makers in the area.

More **resources and support to access and use** these data would be key in generating more impactful research. With respect to data access researchers would benefit from plain language guides on setting up legal arrangements or using secured research environments for analysis. Trusted Research Environments (Data Trusts) can play a role in making granular (sensitive) data available for research use. These secured infrastructures can provide the overarching data security governance and centralise the data agreement and sharing process - significantly facilitating administrative and logistical burden on researchers. Furthermore, it is noted that the access and availability of these data, research and resources should be democratised. It should not simply be the largest or most privileged institutions who are gaining access to these types of data (e.g. having data available for not just developed country capital cities and the largest urban areas).

As it also emerged in the previous session it is challenging to work with new forms of data because they often do not have a comprehensive documentation or standardised data format and terminology. Well documented open and research ready data products or indicators and reports would facilitate research by removing several preprocessing steps. At the same time, the development of open data products as well as open source tools and software needs to be valued and supported. Additional avenues for training or resources dedicated towards skills improvement for working with such potentially complex spatio-temporal data would also be beneficial.

**External Communication and Stakeholder Engagement** is crucial for making sure that local domain expertise is helping to guide any work being developed, and that the insights and measures being taken from mobility data are sensible and beneficial. Including decision makers throughout the research process ensures that priorities and outcomes are aligned. It is also needed to make sure that there is no duplication of efforts and to build on the existing methods, terminologies or frameworks in other domains (e.g. healthcare, physics).

There needs to be an emphasis on **good external communication**. This includes making sure that accurate, robust and unbiased results are being presented through academic research, media channels and the general public. There should be a particular emphasis on producing socially beneficial research and public good. Involving the community and individuals may also be useful towards building these conventions of trust surrounding the use of mobility data. Citizen science and overall science communication can be used to engage local members of the community, and further there is potential to use this avenue for data crowdsourcing or validation. This should be done keeping in mind the hesitation that many individuals in the public have with mobility type data being tracked and stored. Increased communication among stakeholders can further help with understanding what is needed between all parties involved. Oftentimes, the data insights and analysis required by policy makers differs from the analysis undertaken by academic researchers.

**Internal Networking and Collaboration** was recognised during the discussion for the particular value gained by participants in bringing together different voices of researchers working on similar projects and with similar challenges as they pertain to mobility data.
Session 3: Final thoughts

Aims and session framework

For this final session, participants were asked to discuss the next steps and moving forward in this line of research.

Guiding questions

Participants were asked to discuss and comment on the following closing remarks:

In which aspects of mobility research, do you see collaboration/networking among scholars as a driver to advance research on mobility data, and how do you think the Turing can play a role/has capacity to push things forward?

What areas of mobility research (methods, tools, applications) do you think need to be advanced more than others, if any? And why?

Outcomes

Question 1

There was a particular appetite for continued and ongoing collaborative events and arrangements between researchers in this field. Annual or semi-regular workshops or technical meeting groups are a powerful way to encourage the continued collaboration and on-board new researchers to use mobility data and frontier methods in the field. A particular set of academic conferences were mentioned as being useful for networking and presenting ongoing work related to mobility data science (e.g. GIScience Advancing Movement Data Science (AMD’21)\(^1\) or Spatial Data Science Conference\(^2\)). Also, as we gain more data on mobility, a more complex and multifaceted representation of reality emerges from research. This requires knowledge that cuts across disciplines, involving geographers, urban planners, neuroscientists, computer scientists and beyond. Along with conferences and workshops, asynchronous communication i.e. mailing lists, chat groups, should also be encouraged.

The Alan Turing Institute could play a central role in facilitating these collaborations and to bridge gaps across different stakeholders - academic, industry, policy markers, local organisations, developers, etc. The institute is well placed to foster and develop the communication between all interested parties.

Question 2

There are a number of research areas related to the use of mobility data that should be prioritised in the near future which the participants highlighted during the final discussion.

Near real time mobility data are now more and more available and can be applied to different contexts; novel research is needed to develop methods able to deal with the scale and velocity of near real time data. Also, it is recognised that as a society we need to transition towards a net zero emissions future to tackle climate change. Research using mobility data can help inform these transitions by; supporting modal change towards active travel and favouring more energy efficient solutions. These avenues for future research should be incentivised.

---

\(^1\) [https://somayehdodge.info/amd-giscience/amd2021/](https://somayehdodge.info/amd-giscience/amd2021/)

\(^2\) [https://spatial-data-science-conference.com](https://spatial-data-science-conference.com)
In line with the challenges that emerged in the first session, participants underlined that among the areas of mobility research that need to be more advanced in the near future we can list: bias, fairness and privacy. They are the key to creating sustainable and robust collaborations between companies, researchers, and policy-makers. It was noted that during the pandemic the volume of mobility data research increased in large part due to its purported benefits outweighing privacy concerns. In order to keep doing mobility research at a high pace and scale outside of such emergencies privacy concerns have to be better addressed. Similarly, the research community should make a collective effort to try and find disruptive solutions to address issues of fairness and bias, such as unequal access to technology.

Explainability of model results and their generalisability across different geographic contexts should also be prioritised in future research. As models increase in complexity, the explainability of the constructed predictive or generative results becomes a problem in itself. In addition to validating the model results, explainability is important in detecting potential biases. Similarly, knowledge about the generalisability of models across different geographic contexts, ensures that the results better reflect human behaviour.

Finally, another aspect the research community working with mobility data should try and develop further is that of converging towards common standards and definitions. More effective collaboration in the development of open source research tools was identified as one way to achieve this.
Appendix

A1. List of participants

Day 1 – List of speakers

1) Charisma Choudhury (Associate Professor, Institute for Transport Studies, University of Leeds)
2) Laetitia Gauvin (Senior Research Scientist, ISI Foundation)
3) Esteban Moro (Professor, MIT-MediaLab/Universidad Carlos III de Madrid)
4) Alessia Calafiore, PDRA & Francisco Rowe, Senior Lecturer (Geographic Data Science Lab, University of Liverpool)
5) Yanan Xin (Postdoctoral researcher, Mobility Information Engineering lab, ETH Zurich)
6) Maurizio Gibin (Senior Research Fellow, CDRC/UCL) & Jacob Macdonald (PDRA, CDRC/GDSL-University of Liverpool)

Day 2 – List of participants

1) Charisma Choudhury (Associate Professor, Institute for Transport Studies, University of Leeds)
2) Ed Manley (Professor of Urban Analytics, School of Geography, University of Leeds)
3) Neil Walton (Reader, School of Mathematics, The University of Manchester)
4) Luca Pappalardo (Researcher, Institute of Information Science and Technologies (ISTI), Italian National Research Council (CNR))
5) Stefano Iacus (European Commission, Joint research Centre)
6) Anita Graser (Scientist, AIT Austrian Institute of Technology GmbH)
7) Vanessa Frias (Associate Professor, Urban Computing Lab, University of Maryland)
8) Henry Martin (Doctoral student in Geoinformation Engineering, ETH Zurich)
9) Yanyan Xin (Postdoctoral researcher, ETH Zurich)
10) Maurizio Gibin (Senior Research Fellow, CDRC/UCL)
11) Jacob Macdonald (Postdoctoral Research Associate, CDRC/GDSL-University of Liverpool) - Workshop Facilitator and Report Co-author
12) Laetitia Gauvin (Senior Research Scientist, ISI Foundation)
13) Esteban Moro (Senior Research Scientist, ISI Foundation)

ITINERANT Team

1) Daniel Arribas-Bel, Senior Lecturer, Geographic Data Science Lab, University of Liverpool and Deputy Director, Urban Analytics Programme, The Alan Turing Institute; Workshop Coordinator & Facilitator
2) Francisco Rowe, Senior Lecturer, Geographic Data Science Lab, University of Liverpool; Workshop Coordinator & Facilitator
3) Alessia Calafiore, Post-Doctoral Research Associate, Geographic Data Science Lab, University of Liverpool; Workshop Coordinator & Facilitator
4) Krasen Samardzhiev, PhD student, Geographic Data Science Lab, University of Liverpool; Workshop Facilitator
5) Martin Fleischmann, Post-Doctoral Research Associate, Geographic Data Science Lab, University of Liverpool;
A2. Workshop agenda

Day 1 – October 27
14:00 – 14:10 Introduction

14:10 – 15:40
Charisma Choudhury (Institute for Transport Studies, University of Leeds)
Laetitia Gauvin (ISI Foundation)
Esteban Moro (MIT-MediaLab/Universidad Carlos III de Madrid)

15:40 – 15:50 Break

15:50 – 17:20
Alessia Calafiore & Francisco Rowe (Geographic Data Science Lab, University of Liverpool)
Yanan Xin (Mobility Information Engineering lab, ETH Zurich)
Maurizio Gibin (CDRC/UCL) & Jacob MacDonald (CDRC/GDSL-University of Liverpool)

17:20 – 17:30 Closing Remarks

Day 2 – October 28
14:00 – 14:15 Introduction & Presentations

14:15 – 14:50 Session 1: Methods and tools
    14:15 – 14:30 Breakout Rooms (with the support of Mural)
    14:30 – 14:50 Collective Discussion

14:50 – 15:00 Break

15:00 – 15:35 Session 2: Applications and impact
    15:00 – 15:15 Breakout room (with the support of Mural)
    15:15 – 15:35 Collective Discussion

15:35 – 15:55 Session 3: Final thoughts
15:55 – 16:00 Conclusions – Next Steps