Course syllabus: Measuring Citizen's Digital Behaviours Using Web Trackers and Data Donations

Date and time: 13 - 15th March, 2024, 10:00 - 13:00.

Room: 24.009 of the Mercé Rodoreda Building, UPF Ciutadella Campus.

Format: In person and online

Instructor: Oriol Bosch

Course description:

The rapid expansion of the Internet, coupled with the capabilities of modern connected devices, has unleashed a wealth of data that holds tremendous potential for understanding individuals' digital behaviours. As individuals access the internet and interact with digital devices, they leave behind digital traces. Collecting these traces can offer profound insights into human behaviours and thoughts, from understanding the impact of the internet on mental well-being to evaluating how smartphone usage affects sleep patterns, and from studying commute patterns to exploring the relationship between pollution, health, and mood.

Simultaneously, recent technological and legal developments have opened up new possibilities for collecting these digital traces. Advanced tracking technologies can now record what people do on their devices and browsers, and regulations like GDPR grant individuals greater control over their data, enabling researchers to access a wealth of digital traces. This course is dedicated to two promising data collection methods:

- **Web Trackers:** These technologies capture data about visited URLs and app usage when individuals are online. By having participants willingly install these technologies on their devices, researchers can collect information about the websites and apps used, and sometimes even accessing the content individuals interact with on the Internet.

- **Data Donation:** Users voluntarily provide researchers with data that has already been collected by their devices or is accessible through the platforms they use. This encompasses data stored by platforms like Facebook, Instagram, or Twitter, data from wearable devices like Fitbit, device and search behaviours from Google, and even genetic data from companies like 23andMe.

However, like any data source, collecting digital traces is complex and susceptible to errors. Therefore, it is crucial for researchers to not only understand best practices when using these approaches but also be aware of the challenges and errors that may arise. If not collected properly, conclusions drawn from web tracking data and data donations may be significantly biased.

Consequently, this course empowers students with essential skills to effectively measure digital behaviours using these innovative approaches, seamlessly integrating them with online surveys. It provides students with practical knowledge necessary to adeptly design, manage, and analyze projects focused on data collection through web trackers and data donation. Furthermore, the course imparts expertise in recognizing and effectively addressing errors that may surface at each stage of the data collection and analysis process. Specifically, the course has the following learning objectives:

1) Develop an understanding of what web tracking data and data donations are.

2) Learn how web tracking data and data donations can be collected and analysed, and how they can be combined with surveys.

3) Recognise the challenges and errors that might arise in every step of the process of collecting and analysing both data sources.

4) Develop best practices when using this type of data, specifically, strategies to quantify, minimise, and report potential errors.

5) Evaluate the limits of their own and others' web tracking and data donation collection strategies

To support these educational goals, students will have access to the innovative TRI-POL web tracking dataset. Armed with this resource, they will get hands-on experience with computational techniques such as Monte Carlo simulations and Random Forests, enabling

them to assess the quality of digital trace data comprehensively, in a state-of-the-art fashion. They will also get hands-on experience on how to transform the unstructured data that is normally collected with data donations (e.g., images, json documentation), into useful datasets and insights.

Learning schedule:

Day 1	10h00-11h30	Introduction to web tracking data: how to collect, process and analyse it.
	11h30-11h45	Break
	11h45-13h00	The challenges and errors of web tracking data: from identification to correction
Day 2	10h00-11h30	Practice with real web tracking data: learn how to simulate biases and conduct multiverse analysis.
	11h30-11h45	Break
	11h45-13h00	Introduction to data donation: how to collect, process, and analyse them.
Day 3	10h00-11h30	Design principles for a successful data donation project
	11h30-11h45	Break
	11h45-13h00	Practices with real data donations: going from data, to insights

Prerequisites: Only basic knowledge of survey methodology is required. The course is based on an adaptation of the Total Survey Error framework, familiarity with this framework will be helpful.

Software: Some knowledge of R is desirable, in order to fully understand the practical activities of the course. But it is not mandatory: there is no need to create new code, just to

apply already existing code. Hence, it is possible to follow the logic of the methods applied without knowledge in R.

Readings:

Bosch, O. J., & Revilla, M. (2022). When survey science met web tracking: presenting an error framework for metered data. Journal of the Royal Statistical Society Series A: Statistics in Society, 185(Supplement_2), S408-S436.

Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. Computational Communication Research, 4(2), 388-423.

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2023). Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. Communication Methods and Measures, 1-18.

Bosch, O. J., Kuha, J., & Revilla, M. (2023). Uncovering digital trace data biases: tracking undercoverage in web tracking data.

Instructor bio:

Oriol is a Postdoctoral Researcher at the Leverhulme Centre for Demographic Science, University of Oxford. He is also a Research Fellow at the Research and Expertise Centre for Survey Methodology (Universitat Pompeu Fabra). As a computational methodologist, Oriol specialises in the use of survey and computational methods to understand how scientists can best collect and analyse new sources of data, such as digital trace data. He specialises in topics related to web and mobile surveys and the use of digital trace data and sensors to enhance or substitute surveys.

His work has been published in journals such as Social Science Computer Review or the Journal of the Royal Statistical Society.

Oriol currently focuses on how social scientists can best collect information about citizens' online behaviours using data donation and web trackers. Through a combination of survey and computational methods, his research explores how to quantify and minimize digital trace data errors, while comparing them with the ones of surveys. In his current role in Oxford, he is working on the development of a *state-of-the-art* data donation infrastructure.

Before joining Oxford, Oriol completed his PhD in Social Research Methods at The London School of Economics, an MSc in Survey Methods for Social Research from the University of Essex, and a BSc in Political Science from Pompeu Fabra University. Additionally, Oriol has extensive research experience, having worked as a researcher or consultant for The Alan Turing Institute, Wellcome Trust, University of Southampton, University of Mannheim, and the Institute for Social and Economic Research.