WORKSHOP'S BOOKLET

European Research Council
Established by the European Commission

Table of Contents

01 ABOUT WEB DATA OPP

02 WORKSHOP'S PROGRAM

03 ABSTRACTS: DAY ONE

04 ABSTRACTS: DAY TWO

05 KEYNOTE

ABOUT

web data opp project

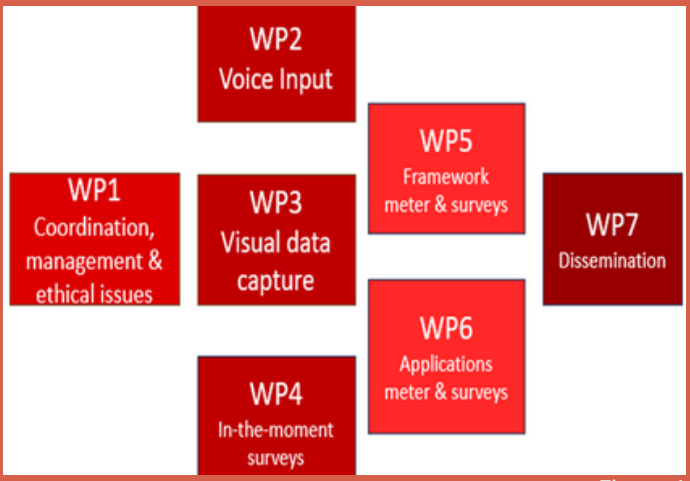

Figure 1

OUR METHODOLOGY

The project is structured in seven closely inter-related work packages (see Figure 1). Aside from the general management (WP1) and dissemination (WP7) packages, the project includes three work packages to test the potential data quality improvements obtained by using new measurement opportunities: voice input (WP2), visual data capture (WP3), and "in-the-moment" surveys (WP4).

In addition, it includes a theoretical work package that corresponds to the construction of a framework on both when/how to replace or combine metered and (mobile) web data (WP5) and, a more applied work package dedicated to illustrating how the framework can be used to answer key applied research questions (WP6).

WORKSHOP FOCUS

The expansion of the Internet and the development of a range of new active and passive measurement tools, particularly on mobile devices, present exciting opportunities for survey researchers, such as visual or geolocation data capture. Using these new measurement opportunities could reduce respondent burden, improve data quality, and extend measurement into new domains, allowing to answer questions that could not be answered so far and to improve the decisions of key actors.

However, while many people speak about these ideas, limited research has implemented such possibilities so far. Besides, researchers interested in using such data face a range of key issues, including technical and ethical ones, and very little is known about how to best deal with all these challenges.

Thus, the primary objective of this workshop is to engage in fruitful discussions surrounding these challenges presented by the integration of new data types within the context of web surveys.

ABOUT

web data opp project

Meet our team

Melanie Revilla
Principal investigator

Patricia Iglesias
Researcher

Carlos Ochoa
Researcher

Oriol Bosch
Researcher

Tracy Silva
Junior Researcher

María Acuña
Junior Researcher

Workshop's program

Monday March 18th				
Ciutadella Campus of Pompeu Fabra University (Ramon Trias 25-27, 08005 Barcelona). Mercè Rodoreda building, Room 24.009.				
Time	**Session**	**Chair**	**Presenter(s)**	**Presenter's affiliation**
10h30 - 11h00	Registration			
11h00 -12h00	Welcome and Opening		Melanie Revilla	IBEI
12h00 - 13h00	Session 1	Melanie Revilla	Jan Höhne	DZHW
			Joshua Claassen	DZHW
			Joris Mulder	Centerdata
13h00- 14h00	*Lunch*			
14h00-15h00	Session 2	Patricia Iglesias	Jessica Daikeler	GESIS
			Şükrü Atsızelti & Mehmet Fuat Kına	Koç University
			Simon Dickopf	Leibniz Institute for Educational Trajectories (LIfBi)
15h00- 16h00	Session 3	Patricia Iglesias	Christopher Antoun	University of Maryland
			Clemens Hetschko	University of Leeds
			Carlos Ochoa	RECSM-UPF
16h00- 16h30	*Break*			
16h30- 18h20	Session 4	Oriol Bosch	Patricia Iglesias	RECSM-UPF
			Ana Slavec	University of Primoska
			Jakob Ohme	Weizenbaum Institut
			Joan Garriga & Monika Falk	Centre d'Estudis Avançats de Blanes (CEAB-CSIC)
			Oriol Llaurado	GDPR consultant
18h20- 18h30	Closing remarks		Melanie Revilla	IBEI
20h	*Dinner*. Gastrobar Pipa - Carrer de Llull, 159			
Tuesday March 19th				
Ciutadella Campus of Pompeu Fabra University. Mercè Rodoreda building, Room 24.009.				
Time	**Session**	**Chair**	**Presenter**	**Presenter's affiliation**
9h00- 10h20	Session 5	Carlos Ochoa	Judith Gilsbach	GESIS
			Leah von der Heyde	Ludwig-Maximilians University (LMU) Munich
			Marc Asensio Manjon	University of Lausanne
			Thijs Carrière	Utrecht University
10h20-10h40	*Break*			
10h40-12h00	Session 6	Carlos Ochoa	Anna Bellido	GfK - NIQ
			Alexandru Cernat	University of Manchester
			Paulina Pankowska	Utrecht University
			Oriol Bosch	Oxford
12h00-12h10	Final words		Melanie Revilla	IBEI
12h10-12h30	*Change of room*			
Ciutadella Campus of Pompeu Fabra University. Mercè Rodoreda building, Auditori.				
12h30-13h30	Keynote. Concluding speech	Melanie Revilla	Mick Couper	University of Michigan
13h30-14h30	*Networking drinks & finger food*			

ABSTRACTS

DAY ONE

OPENING

- Melanie Revilla: “The WEB DATA OPP Project”

SESSION 1.

- Jan Höhne: “Respondent-Centered Incentives: Increasing Answer Provision and Data Quality when it Comes to Voice Answers to Open Questions”
- Joshua Claassen: “Inferring Respondents’ Emotional States From Text and Voice Snswers to Open Questions in a Smartphone Survey”
- Joris Mulder: “Can You Hear Me Loud and Clear? Advantages and Limitations of Voice Recorded Answers in an Online Survey Environment”

SESSION 2.

- Jessica Daikeler: “Assessing Data Quality in the Age of Digital Social Research: A Systematic Review“
- Şükrü Atsızelti & Mehmet Fuat Kına: “Linking Surveys and Social Media Data: Shaping the Future of Public Opinion Research with the Politus Project”
- Simon Dickopf: “Using Common Web Analytic Technologies for Observing Digital Behavior in Web Survey Approaches”

SESSION 3.

- Christopher Antoun: “Using an App for Real-Time Measurement of Transit Rider Satisfaction”
- Clemens Hetschko: “Collecting Panel Data on the Cognitive Wellbeing, Momentary Happiness and Physical State of German Job Seekers”
- Carlos Ochoa: “Navigating Challenges: Implementing In-the-Moment Surveys with Digital Trace and Geolocation Data”

SESSION 4.

- Ana Slavec: “Visual Data in Web Surveys: The Case of Bedroom Flooring”
- Jakob Ohme: “Augmenting Survey Data with Data Donations”
- Joan Garriga & Monika Falk: “Exploring Challenges in Citizen Image Classification for Mosquito Alert Project.”
- Patricia Iglesias: “Challenges when collecting photos of the books respondents have at home through a mobile survey”
- Oriol Llauradó: “Navigating Ethical Challenges in Data Collection”

OPENING

The WEB DATA OPP project

Melanie Revilla

Abstract:

This workshop has been organized in the frame of the WEB DATA OPP project, funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 849165). This project investigates how some of the new measurement opportunities that the growing presence of mobile devices made possible, namely visual data, voice data, metered data and in-the-moment surveys, can help scientists and practitioners to get more accurate and/or more complete insights. In the introductory talk of the workshop, I will present the main idea of this project and give an overview of the research that has been done so far in the frame of the project.

Presenter's bio:

Melanie Revilla is a survey methodologist researcher at IBEI. She is the PI of the WEB DATA OPP project, funded by an ERC starting grant. This project investigates new measurement opportunities linked mainly to the growing presence of smartphones, including the use of visual data or metered data, to complement or replace conventional survey data in order to get better or new insights. She is also Editor-in-chief of the journal methods, data, analyses (MDA). Before moving to IBEI, Melanie was Deputy Director of the Research and Expertise Centre for Survey Methodology (RECSM) and adjunct professor at Universitat Pompeu Fabra.

SESSION 1

Respondent-centered incentives: Increasing answer provision and data quality when it comes to voice answers to open questions

Jan Karem Höhne
Melanie Revilla
Mick Couper

Abstract:

During the last couple of years, survey researchers started to employ open questions with requests for voice instead of text answers. In doing so, they built on new advancements in communication technology to reduce respondent burden and increase data quality. Respondents can record their open answers through the microphone of their survey device by simply pushing a recording button facilitating narrations. Research indicates that voice answers, compared to text answers, are longer (more words and characters), consist of a higher number of topics, and result in higher criterion validity. Voice answers also produce shorter response times than their text counterparts suggesting that they may be less burdensome. However, a key challenge for practical use remains the high item-nonresponse rate associated with voice answers. The level of item-nonresponse varies between 20% and 60% indicating that a substantial minority of respondents may not be willing to provide voice answers. This high level of missing data can negatively affect survey outcomes and reduce the generalizability of results. In this study, we therefore investigate whether respondent-centered incentives help to increase voice answer provision, while still providing text answer options for reluctant respondents. For voice answers, we will provide additional incentives, depending on respondent characteristics (e.g., age and tech-savviness) and survey situation (e.g., survey device and third-party presence). We are planning to conduct a web survey (N ~ 1,000) in the Netquest online access panel (Spain) assigning respondents to different answer conditions. This study is work in progress and still needs methodological refinement since we struggle with appropriate assignment strategies (e.g., what respondents should receive additional incentives?) and group sizes (e.g., how to achieve equally sized groups including enough voice answers?). Despite its methodological challenges, this study sets the stage for tackling the threat of item-nonresponse when it comes to voice answers.

Presenter's bio:

Jan Höhne is a junior professor at the Leibniz University Hannover in association with the German Centre for Higher Education Research and Science Studies and head of the CS3 lab for Computational Survey and Social Science. His research combines survey research and computational social science. He employs Automatic Speech Recognition, Natural Language Processing, and generative AI applications for measuring political and social attitudes. This also includes exploring new, digital data sources and forms for empirical social research.

SESSION 1

Inferring respondents' emotional states from text and voice answers to open questions in a smartphone survey

Joshua Claassen
Jan Karem Höhne
Christoph Kern
Hayastan Avetisyan

Abstract:

A frequently neglected aspect in the process of answering survey questions are respondents' emotional states, such as joy and anger. Although emotions are key for the formation of attitudes and opinions (Kühne & Schemer, 2015), survey research faces difficulties in measuring emotions accurately without rigid scales. In this study, we build on advances in text-as-data methods addressing the following research question: How to infer respondents' emotional states based on the content of their answers to open questions? For this purpose, we ran a smartphone survey in Germany (N = 1,001) and randomly assigned respondents to text or voice answer conditions. We asked respondents one question about their in-situ feelings (baseline) and one question about their opinion of how the world would look like in ten years (target). Between these two questions, we randomly showed respondents a picture of a healthy or unhealthy environment, allowing us to infer emotional changes. Voice answers were transcribed and the text content of answers will be analyzed across conditions. In order to infer respondents' emotional states, we will employ sentiment analysis to evaluate general emotional tone (positive, neutral, or negative). Furthermore, we will utilize transformer models for emotion recognition to identify specific emotions (e.g., joy and anger) in respondents' answers. Additionally, we will look for predominant themes and associated emotional states in respondents' answers through topic modeling. As this study is still work in progress, we are currently facing the following challenges: 1) How to effectively combine sentiment analyses and emotion recognition to infer respondents' emotional states? 2) How to best infer emotion changes between the baseline and target questions? 3) Which model adjustments ensure an accurate and contextually sensitive emotion recognition system? Although still in progress, this research introduces a promising avenue for inferring respondents' emotional states from the content of their open answers.

Presenter's bio:

Joshua Claassen (claassen@dzhw.eu) is a research associate and PhD student at the German Centre for Higher Education Research and Science Studies (DZHW) and the Leibniz University Hannover, Germany. His research combines survey research and computational social science, placing a special emphasis on augmenting web surveys with digital data.

SESSION 1

Can You Hear Me Loud and Clear? Advantages and Limitations of Voice Recorded Answers in an Online Survey Environment.

Joris Mulder

Abstract:

Online survey research (CAWI) facilitates collecting research data quickly, efficiently, and at relatively low costs. This method of data collection has become indispensable in contemporary behavioural research: existing survey infrastructures systematically gather a variety of valuable data for researchers and policymakers.
One valuable data source commonly not collected in traditional online surveys is a respondent's voice. While often unnecessary for many studies, in research focusing on cognitive functioning, socioeconomic status, verbal reasoning abilities, or emotion analyses are important, a valuable source of information is lost. The respondent's tone, language proficiency, and vocabulary can provide valuable insights and can be used for NLP, topic modelling or sentiment analysis.
Two online studies were conducted within the Dutch LISS panel using speech-to-text software (CARI) to collect respondents' voices. In the first study (n=2,242 completed net response) an experimental design using three conditions was fielded: 20% of the sample received the text-only survey, 40% the audio response version, and 40% could choose between the two conditions. Approximately n=280 respondents successfully provided audio responses to multiple survey questions. The second study targeted respondents aged 20 to 49 (n=486 completed net response), of which n=100 respondents provided audio responses to open-ended questions. Those unable or unwilling to do so were asked to finish the text-only version. Audio responses of both studies were stored as audio files and text transcripts.
Despite the substantive potential for linguistic research, a variety of limitations remain when collecting voice responses in online survey research. Our focus is on discussing the (selective) non-response and drop-out rates, willingness to participate, practical and technical constraints, as well as privacy and security issues in these two studies. Our goal is to further enrich insights for the implementation of CARI in CAWI.

Presenter's bio:

Joris Mulder is a senior researcher at the non-profit research institute Centerdata (Tilburg University, the Netherlands). Joris has a background in computer science and social & economic psychology. He coordinates the LISS panel, the online research infrastructure in the Netherlands representative of the Dutch population and specifically setup to facilitate scientific researchers worldwide. Joris oversees the representativeness of the panel, the quality of the collected research data and promotes the awareness and use of the disseminated data, which is freely available through the LISS Data Archive.
quality.

SESSION 2

Assessing Data Quality in the Age of Digital Social Research: A Systematic Review

Jessica Daikeler

Abstract:

While survey data has long been the focus of quantitative social science analyses, observational and content data, although long-established, are gaining renewed attention; especially when this type of data is obtained by and for observing digital content and behavior. However, even the most innovative and extensive amounts of data are hollow if they are not of high quality. But what does data quality mean for modern social science data?

To investigate this rather abstract question the present study focuses on four objectives. First, we provide researchers with a decision tree to identify appropriate data quality frameworks for a given use case. Second, we determine which social science data types and quality dimensions are already addressed in the existing data quality frameworks. Third, we identify gaps with respect to different data types and data quality dimensions within the existing frameworks which need to be filled. And fourth, we provide a detailed literature overview for the intrinsic and extrinsic perspectives on data quality.

By conducting a systematic literature review based on text mining methods, we identified and reviewed 58 data quality frameworks. This contribution will especially emphasize on data quality of voice, visual and digital trace data and discuss pitfalls and possible data quality indicators for these data types.

Presenter's bio:

Dr. Jessica Daikeler is working as a post-doc at GESIS in the Survey Design and Methodology department, she was am part of GESIS' new strategic direction on digital behavioral data. She graduated and did her PhD at the University of Mannheim. Since November 2023 she is coordinating the KODAQS project at GESIS, the University of Mannheim and the LMU Munich to build a competence center for the data quality in the social sciences in Germany.

SESSION 2

Linking Surveys and Social Media Data: Shaping the Future of Public Opinion Research with the Politus Project

Şükrü Atsızelti
Mehmet Fuat Kına

Abstract:

ERC-funded Politus Project aims to revolutionize public opinion research by harnessing the power of micro-blogging data and AI, while adhering to GDPR standards. Traditional survey methods are increasingly challenged by declining response rates, diminished trust, and social desirability bias—where respondents may conceal or alter their views on certain topics. To address these issues, Politus proposes a blend of cutting-edge public opinion research method, including multimodal deep learning, network analysis, web surveys and privacy-preserving techniques. This approach seeks to gauge public opinion accurately and representatively without the biases inherent in traditional surveys or raw digital trace data.

By linking survey responses with user-generated content on Twitter, the project intends to develop models that can predict survey outcomes from social media activity. This innovative method involves collecting tweets from consenting survey participants, then analyzing these alongside survey data to build machine learning models capable of inferring user characteristics—such as education—from their online expressions. Regular online surveys will keep these models fresh, reflecting the dynamic nature of public sentiment.

Additionally, the project will employ this data in a tweet-to-user conversion process. We originally rely on analyzing individual tweets and then setting thresholds to determine the users' ideology, for example, labeling users as 'religious' if their tweets surpass a specific threshold. This conversion process, however, is filled with uncertainties, posing a challenge in accurately reflecting a user's true characteristics based on their tweets. The project will use the survey responses to evaluate the usefulness of thresholds.

Based on the knowledge gained during this project, this presentation aims to delve into the technical and ethical challenges encountered during the data gathering and linking processes and to discuss the potential solutions.

Presenters' bio:

Dr Atsızelti is Sociology PhD candidate at Koç University. Currently, he is engaged with the ERC-funded "Politus: AI-based Data Platform for Fair Social Polities" project. This initiative aspires to furnish real-time public sentiment data to pertinent social and political stakeholders. His primary role encompasses online survey design and gleaning demographic insights from social media datasets. Furthermore, he is delving into the utilization of large language models for tasks like stance and event extraction.

Dr Kına works as a postdoctoral researcher at the Center for Computational Social Sciences, and an instructor in the Computational Social Sciences MA program at Koç, conducting research and teaching in the fields of computational social sciences, social movements, and comparative analysis of welfare systems and social policy. Since beginning my doctoral studies (2018), He has been employed as a researcher in three European Commission-funded, large-scale, and innovative research projects: the ERC-funded Emerging Markets Welfare, the ERC-funded Politus, and the Horizon Europe-funded Social ComQuant projects.

SESSION 2

Using Common Web Analytic Technologies for Observing Digital Behavior in Web Survey Approaches

Dickopf, S.,
Bela, D.
Schneider, P.

Abstract:

Motivating respondents to participate in social science surveys is a growing challenge (Luiten et al., 2020). Considering the lengthy questionnaires of multi-purpose longitudinal surveys, such as the German National Educational Panel Study (NEPS), this issue gains even more importance. In such survey programs, non-response can easily lead to drop-outs of further panel waves. Motivational aspects are not only a crucial issue for recruitment and communication strategies. They also play a role while respondents process a survey. With regards to this, the situational specifics that may affect a respondent's burden while answering a questionnaire are key to successful participation. Among many other aspects, this burden is induced by the questionnaire itself, and especially its perception by the respondent (Fricker et al., 2014). Measuring the changes of an respondent's burden over the time of processing the
survey (Couper et al., 2013) is beneficial for adaptive designs that take the situational level of burden into account. In the context of self-administered questionnaires, that burden can, to some extent, be derived by observing digital behavior (Horwitz et al., 2017). A reliable measurement of digital behavior is therefore the crucial starting point to achieve this.
Using the NEPS web survey infrastructure, we can show how freely available web analytics technology has been implemented to observe digital behavior. We used this infrastructure in the first waves of a newly recruited panel of fifth-graders, as well as in context surveys for their parents and teachers, starting in 2022. The conducted measurement of digital behavioral data focuses on access-specific information (such as login timestamps, interruptions, window width and resolutions) as well as interaction-specific data (such as state and timestamps of any click, scrolling, keystroke event, changes of window size or focus).
With that measurement of observed digital behavioral data we aim on estimating the respondents burden component that is induced by the individual's perception of the questionnaire.

Presenter's bio:

Simon Dickopf is head of the working unit Survey Technology at the Center for Study Management at the Leibniz Institute for Educational Trajectories (LIfBi). which is responsible for setting up and continuously developing a largely automated and scalable infrastructure for computer-based surveys. His research interests include respondent burden, attitudes towards surveys, using metadata for optimizing surveys and using paradata for enriching respondents data.

SESSION 3

Using an App for Real-Time Measurement of Transit Rider Satisfaction

Christopher Antoun
Vanessa Frías-Martínez
Saad Mohammad Abrar
Naman Awasthi
Anthony Garove

Abstract:

Traditionally, online surveys of transit riders were administered with some delay after they made their trips. Today, by comparison, transit surveys are being integrated into smartphone apps, and riders can answer them on-board or immediately after their trips. In light of this change, it is critical to understand the measurement properties of these in situ app-based approaches and how they might be improved. In our study, Baltimore City residents living in lower-income neighborhoods were invited to download the BALTOApp (Abrar et al. 2023) and record their trips on public transit for two weeks. They were randomly assigned to complete satisfaction surveys either during or immediately after each trip, with an incentive provided either for each completed survey or for each day of participation. Data collection for this study is currently in progress. Analyses are planned to examine differences in participation (response rates) as well as several indicators of response quality (missing data, non-differentiation, and so on). Our presentation will highlight some practical challenges involving participant recruitment and app development that we encountered when implementing the study.

Presenter's bio:

Christopher Antoun is an Assistant Research Professor in the Joint Program in Survey Methodology and College of Information Studies at the University of Maryland. His research tries to understand the measurement properties of newer data collection methods that take advantage of recent technology and how those methods might be improved. Chris holds an M.S. and Ph.D. in Survey and Data Science (formerly Survey Methodology) from the University of Michigan and has completed a fellowship at the U.S. Census Bureau. His work has been funded by NSF, NCSES, and USDA. He served as associate editor for the *Journal of Survey Statistics and Methodology* between 2020–2023 and serves on the editorial board for *Public Opinion Quarterly*.

SESSION 3

Collecting Panel Data on the Cognitive Wellbeing, Momentary Happiness and Physical State of German Job Seekers

Clemens Hetschko

Abstract:

This presentation introduces the German Job Search Panel (GJSP), a monthly survey that follows people who registered as job seeking over the course of up to two years. The focus of the survey is on the wellbeing and health of jobseekers, with special emphasis on workers affected by mass layoffs. The use of an innovative survey app allows for frequent measurement every month and for conducting the experience sampling method (i.e., momentary assessment) to measure affective wellbeing. The collected data may be linked to administrative records of the Federal Employment Agency subject to participant consent. For a subsample of surveyed jobseekers hair cortisol levels are available as a biomarker for chronic stress.

The presentation elaborates on the opportunities and challenges of a data collection using smartphone apps to gather experience sampling data. It also presents the results of several survey methodological contributions based on GJSP paradata. This includes a study of the feasibility of hair cortisol sampling in a non-face-to-face setting as well as an experiment on the impact of high-frequency study participation on labour market behaviour.

Presenter's bio:

Clemens Hetschko is an Associate Professor in Economics at the University of Leeds. He obtained his PhD from Free University Berlin (2015), Germany, and continued to work there until 2020 as a post-doc as well as at the main German institute for labour market policy research (IAB Nuremberg). His research is at the intersection of economics and psychology and analyses worker wellbeing with a view to informing labour market policy. Another current focus of his work combines economic modelling with survey data to study the determinants and impacts of risk and time preferences. Clemens has been part of the team that collected the German Job Search Panel and contributed a number of survey-methodological analyses around participant recruitment strategies, biomarker collection and the Hawthorne effect. Clemens has published in major journals across fields, such as the Economic Journal, the British Journal of Industrial Relations, Journal of the Royal Statistical Society Series A and the Journal of Personality and Social Psychology.

SESSION 3

Navigating Challenges: Implementing In-the-Moment Surveys with Digital Trace and Geolocation Data

Carlos Ochoa

Abstract:

This presentation focuses on in-the-moment surveys, defined as web surveys sent to a subset of online panel members who already share some passive data whenever an event of interest is detected through such data. This method allows for the collection of information that cannot be gathered passively while simultaneously reducing memory errors that often impact conventional surveys.

Recent technological advancements have made the automated execution of in-the-moment surveys triggered by digital trace and geolocation data possible. This includes the detection of events of interest, such as website interactions or physical location visits, along with the timely dispatch of survey invitations to potential participants.

However, the implementation of in-the-moment surveys involves the execution of specific costly tasks compared to conventional surveys. For surveys triggered by metered data, compiling a comprehensive list of relevant websites and specific URLs is essential, requiring ongoing updates throughout the project. Similarly, surveys triggered by geolocation data demand the identification of GPS coordinates corresponding to specific locations of interest. Testing procedures for these surveys markedly differ from traditional approaches.

This presentation delves into the insights gained from a successful in-the-moment survey sent to members of an online job panel who applied for a job online. Additionally, it explores the intricacies of setting up a similar survey planned for visitors to bathing areas in Spain during the summer of 2024, shedding light on the preparation, execution, and lessons learned from these innovative survey methodologies.

The presentation elaborates on the opportunities and challenges of a data collection using smartphone apps to gather experience sampling data. It also presents the results of several survey methodological contributions based on GJSP paradata. This includes a study of the feasibility of hair cortisol sampling in a non-face-to-face setting as well as an experiment on the impact of high-frequency study participation on labour market behaviour.

Presenter's bio:

Carlos Ochoa is a researcher at the Research and Expertise Centre for Survey Methodology (RECSM). He holds a degree in Telecommunications Engineering from the Universitat Politècnica de Catalunya (UPC) and is currently pursuing his PhD under the guidance of Melanie Revilla at UPF, working in the Web Data Opp project.

Before joining RECSM, Carlos developed his professional career in the private sector, being part of the entrepreneurial team of Netquest for 16 years. Netquest is an online panel company that provides valuable data to researchers spanning various disciplines, including social and political science, as well as market research.

Carlos's recent research endeavors have been centered around the utilization of innovative data types for research purposes. Specifically, his doctoral work is dedicated to exploring the feasibility, advantages, and potential drawbacks of in-the-moment surveys triggered by passive data collection. Additionally, he maintains a keen interest in subjects such as survey methodology, online opt-in panels, willingness to participate, and conjoint analyses.

SESSION 4

Challenges when collecting photos of the books respondents have at home through a mobile survey

Patricia Iglesias

Abstract:

In social sciences-related surveys it is usual to inquire about the numbers of books at home, however most types of questions regarding this present several limitations (e.g., wide intervals and potential impact of social desirability bias). Those limitations might be overcome by asking respondents to send photos of the books in their home: photos might increase the accuracy of the responses, while also allowing to collect additional information, such as the location(s) where the books are stored, and the presence of books in different languages.

In this study, we asked respondents for photos of the books they have at home to know the number of books for children who do not know how to read by themselves, for literate children and teenagers, and those aimed to a general audience. Photos also allow knowing about the books' storage and languages. In addition to the photos, respondents were asked to type the number of books in each category, and answer questions regarding books' storage and languages.

Throughout the collection, classification, and analysis of the photos, various challenges emerged. These challenges made the overall process more intricate compared to conventional questions, and presented researchers with additional decisions in order to make the best of the photos and the information contained in them.

Overall, the collection of photos is complex and presents many challenges, which should be considered for researchers to assess whether is more convenient to use photo-based or conventional questions.

Presenter's bio:

Patricia A. Iglesias is a researcher at the Research and Expertise Centre for Survey Methodology (RECSM) and a pre-doctoral student at the Department of Political and Social Sciences in Universitat Pompeu Fabra (UPF) in Spain. She is developing her thesis within the project "WEB DATA OPP: New opportunities to enhance or extend (mobile) web survey data and get better insights." Patricia holds a Research Master in Sociology and Demography from UPF and a degree in Sociology from the Universidad de Chile. Before joining RECSM, she worked in the National Statistics Institute of Chile, coordinating the National Labor Survey transition from a pen-and-paper personal interview (PAPI) to a computer-assisted personal interviewing (CAPI) methodology, and participating in the update of the survey's questionnaire. Patricia is interested in the capture and sharing of images from web survey respondents to replace or complement conventional questions.

SESSION 4

Visual data in web surveys: the case of bedroom flooring

Ana Slavec

Abstract:

Flooring materials have distinct properties which affect not only the comfort of residents but also their comfort and their well-being. Notably, natural materials like wood can contribute to better indoor air quality. Moreover, their lower carbon footprint in production and disposal renders them environmentally friendlier compared to their synthetic counterparts. However, official data usually lacks specific information regarding the materials used for flooring and other household furnishing.

In August 2023 we conducted a residential renovation survey on an online market research panel in Slovenia. A total of 3765 panelists were invited to participate, and 1009 completed the questionnaire in full. The survey included a question regarding the type of flooring material used in bedrooms, along with a subquestion to measure how confident they were in their answer. This is important as some participants might not be able to identify the material, especially if they have not furnished the apartment themselves.

Furthermore, we prepared a brief follow up-survey that entailed respondents to upload a photo of their bedroom floor. They were instructed to respond while being at home and out of the 1009 invitees, 380 participated, with 314 of them uploading an image. In cases where the upload did not happen, we inquired why. In addition, all respondents were asked if they would be willing to provide other types of visual and other data.

The research team coded the images and removed those that did not reveal floors. Additionally, Google image search was used to verify their authenticity. The remaining photos were reviewed by a panel of experts who helped identify the type of material, which were then compared to survey responses. The study critically evaluates the potential of respondent photos to enhance data collection, address key methodological challenges and lay out recommendations for future research utilising visual data.

Presenter's bio:

Dr. Ana Slavec is an assistant professor at the Faculty of mathematics, natural sciences and information technology at the University of Primorska and a research associate at the InnoRenew CoE Renewable Materials and Healthy Environments Research and Innovation Centre of Excellence in Izola, Slovenia. She holds PhD in social science statistics from the Faculty of Social Sciences at the University of Ljubljana. Her dissertation was on the topic of improving survey question wording using language resources. In her current research work she focuses on the use of surveys in sustainable building research. She has been the principal investigator on a postdoctoral project on Using survey questionnaires to measure attitudes and behaviours of building users and a bilateral research project on Innovation activities of Austrian and Slovenian companies in the wood-value chain in collaboration with the University of Graz, both funded by the Slovenian research and innovation agency. In her career she has co-authored 8 original scientific articles, 6 chapters in scientific monographs issues by international scientific publishers, and various other publications, including several conference presentations at international events.

Augmenting Survey Data With Data Donations

Jakob Ohme

Abstract:

This talk addressed how data donations can augment information collected via surveys. It describes the process of the collection, augmentation, and analysis of the different data sources and highlights the value of combining observed with self-reported data. Based on experiences with visual data donations collected as screenshots and videos from mobile and digital platforms, the challenges and limitations of this approach are discussed, presenting an overview of different possibilities on how digital trace data can inform media research on an individual data level.

Presenter's bio:

Jakob Ohme is Head of Research Group at the Weizenbaum Institute for the Networked Society in Berlin, where he leads the 'Digital News Dynamics' research group.

His research interests center around the impact of digital and mobile communication processes on political behavior and news flows, as well as generational differences in media use and political socialization. His work focuses on developing and utilizing digital methods in political communication and journalism research.

Prior to his work in Berlin, he worked as a Postdoctoral Researcher at the Amsterdam School of Communication Research (ASCoR), University of Amsterdam, and as Assistant Professor at the Centre for Journalism, University of Southern Denmark, where he also earned his PhD degree. Jakob earned a Master's degree from the Department of Media and Communication at the Technische Universität Dresden in Germany. He is a Fellow at the Digital Communication Methods Lab at the University of Amsterdam.

SESSION 4

Exploring Challenges in Citizen Image Classification for Mosquito Alert Project.

Joan Garriga
Monika Falk

Abstract:

Mosquito Alert (MA) is a citizen science initiative focused on the surveillance and control of mosquito disease vectors (i.e. Aedes albopictus, Aedes aegypti, Aedes japonicus, Aedes koreicus and Culex pipiens), engaging citizens via a smartphone app to report the presence of these mosquitoes. Increased spatial resolution and real-time monitoring of the distribution of these species are crucial to trigger the necessary response from the public health authorities. With this aim, MA has shown itself as a low-cost and more effective tool than traditional surveillance methods. However, the quality standards required to this end are high. Therefore, MA depends on a solid validation protocol based on a group of entomology experts who visually inspect all the reports uploaded to the system. This validation protocol has a direct effect on the average response time of the system.

AIMA is the artificial intelligence system powering the Mosquito Alert project and plays a central role in automatically classifying the images uploaded with citizen reports. AIMA's automatic classification contributes some fundamental features:

- Early warning: the system raises an alert when it identifies a targeted species in a non-colonized area;
- Improved citizen engagement: citizens receive immediate feedback from the system;
- Improved system response: the system filters off-topic reports and dispatches the ones of interest to the attention of the experts.

However, citizen images are challenging due to their unpredictable heterogeneity in terms of content, position, focus, zoom, light and background. In addition, the content is highly unbalanced in relation to the species of interest and not perfectly matched to them. As a result, the raw image dataset does not provide the necessary suitability to train a robust model for species recognition. In our presentation, we will address the complexity arising from this unpredictable heterogeneity and discuss the solutions we have developed to overcome these challenges.

Presenters' bio:

Monika Falk is an IT professional with a strong background in web development and artificial intelligence. With dual master's degrees, one in Computer Science focusing on Multi-media Technologies and Artificial Intelligence Methods, and another in Environmental Engineering concentrating on Heating, Ventilation, and Atmosphere Protection, Monika boasts a diverse educational background. She transitioned her career focus from web development to an expertise in artificial intelligence several years ago. Monika is currently pursuing a doctoral degree, focusing on emotion recognition for social robots. Her aim: AI understanding of human emotions."

Joan Garriga is an electrical engineer (ETSEIT, 1992) with over 15 years of experience in freelance programming. He started a Ph.D. program in Artificial Intelligence (UPC, 2002) and joined the Theoretical and Computational Ecology group (CEAB-CSIC, 2011). As a data scientist and machine/deep learning expert at CEAB, Joan has contributed to several projects in the field of movement ecology aimed at understanding the structural complexity and generative mechanisms of foraging behavior and monitoring the migratory movements of small animals via satellite multi-sensor movement data. He also developed two innovative R packages, EMbC and bigMap, tailored for unsupervised analysis of large data sets. He currently leads AIMA, the artificial intelligence system powering the Mosquito Alert project.

SESSION 4

Navigating Ethical Challenges in Data Collection

Oriol Llauradó

Abstract:

In this presentation, we explore key ethical considerations in data collection, especially web surveys, from fundamental principles like informed consent to addressing specific challenges with the four new data types investigated in the WEB DATA OPP project. We will present real applications, covering issues such as shared devices, privacy in metered data, and handling sensitive images. You will gain practical insights for fostering responsible data practices and mitigating risks.

Presenter's bio:

Oriol Llauradó is a social scientist who has devoted his career to promote the ethical and legal use of new technologies. After spending +15 years in the data industry, in 2019 he has started his consultancy firm and acts as an independent privacy advisor for digital companies. Certified as Privacy Manager by the IAPP since 2015. Member of the Ethical Committee of the Spanish society for social and market research.

ABSTRACTS

DAY TWO

SESSION 5.

- Judith Gilsbach: “Quality Improvement or Disruption of Participant Behavior?”
- Leah von der Heyde: “Can large language model estimate how people vote? Evidences from Germany”
- Marc Asensio Manjon: “Measuring Smartphone Usage Through Data Donations: Challenges and Best Practices”
- Thijs Carrière: “Best Practices in Data Donation: A Workflow For Studies Using Digital Data Donation”

SESSION 6.

- Ana Bellido: “Acquisition and Methodology to Derive Insights from an Enriched Panel with Digital Passive Measurement”
- Alexandru Cernat: “Estimating Measurement Quality in Digital Trace Data and Surveys Using the MultiTrait MultiMethod Model”
- Paulina Pankowska: “Measuring Facebook Use: The Accuracy of Self-Reported Data vs. Digital Trace Data”
- Oriol Bosch: “The Uncertainties of Working with Web Tracking Data, and on How to (Maybe) Tackle Them”

KEYNOTE

- Mick Couper: “New Data Types and Surveys: Opportunities and Challenges”

SESSION 5

Quality improvement or disruption of participant behavior?

Judith Gilsbach
Joachim Piepenburg
Frank Mangold
Sebastian Stier
Bernd Weiss

Abstract:

Our study links web tracking data with surveys. We decided to run identical incentive experiments in two data collections. In each data collection half of the participants were assigned an unconditional incentive of 5€ with the invitation to the web tracking study. Additionally, each participant was assigned one out of four conditional incentives. The conditions were ranging from no incentive to 40€. Participants were told that they will receive the conditional incentive only if they show minimum 30 days of activity. Participants were recruited into a panel via Meta ads and via a German general population survey (ALLBUS).

Another decision we made is to send inactivity reminders. The tracking tool works via a browser plug-in for Chrome, Firefox and Edge and scrapes visited URLs and HTML snapshots passively once the visit of a website is terminated. Participants need to install the plug-in in the beginning of the study and log-in once with their credentials. After that no further interaction with the tool is needed. Every second week we established a "reminder day". Each participant who had not sent any data the week prior to the reminder day received an activity reminder.

Both measures, the minimum activity threshold to receive the postpaid incentive and the activity reminders, can potentially interfere with participants online behavior and encourage them to show behavior they would not show naturally. Additionally, the reminders raise the awareness of being tracked. But what is the larger concern for data quality: missing digital traces or recoding artificial behavior that is only imitated by participants to simulate activity? We can present some measures of activity after reminders and depending on the incentive conditions. We would like to discuss how other researchers handle this issue and whether conditional incentives and reminders are rather disruptive or a valid measure to reduce missing data.

Presenter's bio:

Judith Gilsbach is a 2nd year PhD student at GESIS Leibniz Institute for the Social Sciences in the Computational Social Science department. She holds a master's degree in sociology. Her lab builds a web tracking infrastructure as a service to other social scientists and conducts research on recruitment, data quality and other methodological topics. For her dissertation she analyzes the web tracking data as bipartite social network data. The presented work is joint work with the above-mentioned authors.

Can large language model estimate how people vote? Evidences from Germany

Leah von der Heydel
Anna-Carolina Haensch
Alexander Wenz

Abstract:

The recent development of large language models (LLMs), such as OpenAI's GPT, have spurred discussions about whether these models might serve as a novel, efficient method of collecting public opinion data. As LLMs are trained on large amounts of internet data, potentially reflecting attitudes and behaviors prevalent in the population, LLM-generated "synthetic samples" could complement or replace traditional surveys. A number of studies has prompted LLMs to mimic survey respondents, finding that the responses closely match the survey data. However, most of these studies focus on the US population. Several factors might affect the generalizability of such findings, including the prevalence of native-language training data, structural differences between the target population and the population reflected in the training data, and a country's political and social structure. In this study, we investigate whether LLMs can be used for estimating public opinion outside of the United States.

To generate a "synthetic sample" of the voting-eligible population in Germany, we use survey data from the 2017 German Longitudinal Election Study (GLES) to create personas matching the respondents' individual characteristics. Prompting GPT-3.5 with each persona, we ask the LLM to predict each respondents' vote choice in the 2017 German federal elections. We compare the LLM's predictions to the survey data on the aggregate and subgroup level. We find that GPT-3.5 does not correctly predict citizens' vote choice, exhibiting a bias towards progressive parties, and making better predictions for more typical voter subgroups, such as political partisans.

By investigating the contextual conditions under which LLMs can be leveraged for studying public opinion, our study contributes to the growing body of research about how LLMs can be used in the social sciences. The findings underscore the limitation of applying LLMs for election prediction without accounting for the biases and potential limitations in their training data.

Presenter's bio:

Leah von der Heyde is a PhD candidate and research associate at the Social Data Science and AI Lab (SODA) at LMU Munich. She holds a bachelor's and master's degree in Political Science from the Universities of Munich and Mannheim with specializations in quantitative methods, survey research, and political sociology. Leah was also a visiting student at Georgetown University, where she focused on polling, public opinion, and electoral campaigns. During her studies, she worked on projects such as analyzing the potential of Google Trends for predicting European Parliamentary elections.

Back at her alma mater since 2021, Leah is investigating the potential and pitfalls of new data sources, such as large language models and social media data, for learning about public opinion. She is especially interested in data justice, giving a voice to groups that are typically overlooked or being biased against in data collection and analysis, the prediction of political behavior, as well as data storytelling.

Measuring smartphone usage through data donations: Challenges and best practices.

Marc Asensio Manjon
Oriol J. Bosch
Caroline Roberts

Abstract:

Given the widespread adoption of smartphones, it is vital to understand smartphone-usage patterns, and their effect on online and offline phenomena such as mental wellbeing. Although survey self-reports are used as the main instrument to obtain smartphone usage estimates, evidence has led researchers to doubt about their quality. Hence, there is an increasing interest in directly observing individual's behaviors using digital trace data. A relatively unexplored alternative to collect individuals' digital traces, known as data donations, is to ask participants to share data that their devices and services already collect from them, during the course of the survey. This should stablish a transparent and self-administrable dynamic that tracking software procedures do not offer.

Data donations can take different forms, all with their own challenges, the most prominent being the expected low willingness to engage in a task that might be perceived as burdensome and sensitive. In this study, we specifically focus on the challenges and best practices when collecting already saved information about participants' daily screentime, number of pickups and specific app usage, as reported in 'Digital Wellbeing (Android) / Screentime (IOS)' tools of their smartphones.

To study this, we conducted a survey experiment in Switzerland using an online pane. Specifically, respondents were randomly asked to share this information in three separate ways: (1) taking and uploading several screenshots of the tools; (2) taking and uploading several video recordings; and (3) manually checking and reporting the information.

We discuss the challenges associated with implementing such tasks in an online panel and provide insights on how this can be done with (almost moderate) success. We address the problematic on how to integrate different devices (and different interfaces / tools) and how to instruct respondents when no knowledge is assumed. Additionally, we examine response rates to draw conclusions about the best approach.

Presenter's bio:

Marc Asensio is a PhD candidate at the University of Lausanne and a RECSM external collaborator. The main focus of his research is the integration of smartphones on the web survey data collection process. Particularly, how to facilitate and normalize smartphone participation on web surveys and how to extend data collection by using them. Marc has many more qualities as human being, but these are unrelated with his research career.

SESSION 5

Best practices in data donation: A workflow for studies using digital data donation

T.C. Carrière
L. Boeschoten
B. Struminskaya
H.L. Janssen
N.C. de Schipper
T.B. Araujo

Abstract:

Digital trace data are gaining popularity in social scientific research and form a good complement to other data types, such as survey data. Digital trace data can be collected through several methods. Current methods exhibit disadvantages, such as high costs and lack of consent of people using the data. Data donation forms an alternative collection method for digital trace data, accounting for these disadvantages. In data donation, research participants obtain their own digital trace data and actively consent to donating (part of) these data to research.

The General Data Protection Law (GDPR), legislation in Europe on personal data, which digital trace data are classified as, enables data donation. The GDPR mandates any entity that collects personal data to provide a copy of this data to the people concerning the data upon request. Larger data controllers often comply to such request by means of Data Download Packages (DDPs), which are folders containing files with all the requested data. Boeschoten et al. (2022) introduced an approach to data donation relying on these DDPs. In the approach, privacy of participants is preserved by extracting and processing data from DDPs locally before participants decide on donating the data.

Setting up a data donation study presents researchers with numerous challenges and considerations. In our paper, we introduce a workflow that guides researchers in conducting a data donation study. This workflow is based on the approach to data donation by Boeschoten and colleagues, data donation studies conducted earlier, and the total error framework for data donation. The workflow consists of six steps that are encountered in setting up data donation studies, and we distinguish between five areas of expertise needed for the workflow. We illustrate that the workflow takes into account potential sources of error and therefore helps in preserving quality of future data donation studies.

Presenter's bio:

Thijs Carrière graduated in 2022 from the research master in Methodology and Statistics at Utrecht University. He wrote his master thesis on usage of process data in explaining differential item functioning, at the Center for Educational Testing in the Netherlands. After graduating, Thijs worked for 2 years as a junior researcher on the VIDI grant awarded to Daniël Oberski, mainly being involved in the D3I project. This Digital Data Donation Infrastructure project focusses on setting up an infrastructure for data donation studies. Thijs' research focused on best practices in data donation studies, and on guiding researchers in setting up a data donation study while retaining study quality. Furthermore, he was involved with setting up two pilot data donation studies and leading the data download packages volatility project, which studied the stability of data download packages. From January 2024 onwards, Thijs started as a PhD student on the VIDI grand awarded to Bella Struminskaya. This project aims at building software that improves digital behavioral data collection for social sciences.

SESSION 6

Acquisition and methodology to derive insights from an enriched panel with digital passive measurement.

Ana Bellido

Abstract:

Passive measurement over devices with access to Internet can be a source of very valuable insights on the population digital consumption patterns. This speech will cover an overview of an entire path to those insights applied in the industry. We'll start with the acquisition of data, which involves different data sources like a panel, census measurement and population benchmarks. We'll go through some data science models being used to be able to deliver a representative view on our target. And ending up in the reporting tool from which our clients can consume the derived KPIs.

Presenter's bio:

Anna Bellido is currently working as a Senior Data Scientist and a Product Owner for GfK-NIQ. There she has been working for the past 4 years in developing pipelines to deliver a view on digital media consumption for specific markets. She previously worked in a research lab conducting psychological studies using virtual reality (Event Lab, Universitat de Psicologia, UB). Her background is on Mathematics and Computer Science, with a master's degree in computing

SESSION 6

Estimating measurement quality in digital trace data and surveys using the MultiTrait MultiMethod model

Alexandru Cernat

Abstract:

Digital trace data are receiving increased attention as a potential way to capture human behavior. Nevertheless, this type of data is far from perfect and may not always provide better data compared to traditional social surveys. In this study we estimate measurement quality of survey and digital trace data on smartphone usage with a MultiTrait MultiMethod (MTMM) model. The experimental design included five topics relating to the use of smartphones (traits) measured with five methods: three different survey scales (a 5- and a 7-point frequency scale and an openended question on duration) and two measures from digital trace data (frequency and duration). We show that surveys and digital trace data measures have very low correlation with each other. We also show that all measures are far from perfect and, while digital trace data appears to have often better quality compared to surveys, that is not always the case. Finally, we find that the measures of duration of smartphone usage both in surveys and digital trace data have the best quality out of the methods we compared.

Presenter's bio:

Alexandru Cernat is an associate professor in the social statistics department at the University of Manchester. He has a PhD in survey methodology from the University of Essex and was a post-doc at the National Centre for Research Methods and the Cathie Marsh Institute. His research and teaching focus on: survey methodology, longitudinal data, measurement error, latent variable modelling, new forms of data and missing data. You can find out more about him and his research at: www.alexcernat.com

SESSION 6

Measuring Facebook Use: The Accuracy of Self-Reported Data vs. Digital Trace Data

Paulina Pankowska
Ruben Bach
Florian Keusch
Alexandru Cernat

Abstract:

The recent growing availability of digital trace data has prompted social scientists to rely more on these data sources in their research. Information obtained from these data is increasingly used in current studies to replace surveys, or as a benchmark to validate and assess the quality of survey- based measures. These studies often rely on the unrealistic assumption that log-data are error-free. However, like any other data source, digital traces are likely to be subject to non-negligible error. Research related to social media and its impacts is a domain in which these concerns are particularly prominent. Therefore, in this paper, we examine the quality of self-reported and digital-trace measures of Facebook use simultaneously, while allowing each of the sources to contain measurement error. To assess and compare the error in these two sources, we apply hidden Markov models to a sample from a non-probability online panel in Germany. The self-reported measures are based on a longitudinal survey, and the digital-trace measures are based on information from tracking apps that were installed on the respondents' smartphones and/or computers.

Our results suggest that both sources measure Facebook use rather consistently for about two thirds of the sample. For the remaining one third of the respondents, we observe large inconsistencies between the survey and the digital-trace measures. Namely, in this group, respondents are highly likely to report using Facebook daily, while the tracking data shows no, or close to no, Facebook use. Our results suggest that the inconsistencies can be (partially) due to the tracking data being systematically incomplete for some individuals, as potentially the device or devices on which Facebook is used most often is not tracked. A further confirmation for our findings comes from estimates obtained from Facebook donated data, which is available for a sub-sample of our respondents.

Presenter's bio:

Paulina Pankowska is an Assistant Professor at the Sociology Department of Utrecht University. Previously she was a postdoctoral researcher at the Department of Communication Science and the Department of Sociology, Vrije Universiteit Amsterdam. She is currently leading a project that aims to design and set-up a social science benchmark. Her research relates primarily to data and methods quality in the social sciences. She has also published in the field of official statistics, survey research methods, and machine learning.

In November 2020 Pankowska defended her PhD dissertation titled: 'Measurement error: estimation, correction, and analysis of implications', which investigated the feasibility of using hidden Markov models (a latent variable modelling technique) to account and correct for measurement error in survey and administrative data. The project was conducted in collaboration with Statistics Netherlands.

SESSION 6

The Uncertainties of Working with Web Tracking Data, and on How to (Maybe) Tackle Them

Oriol J. Bosch

Abstract:

When working with web tracking data, researchers face many challenges that are hardly acknowledged and discussed. Many of these challenges add a level of uncertainty to web tracking research that is hardly found when collecting and using survey data. For example, when collecting data for a specific behaviour, if a researcher realises that the values observed are completely different than expected based on the literature and common sense, a big question arises: were my priors completely wrong, or is this supposed gold standard surprisingly off? The complexity of the process of collecting and processing web tracking data makes it hard to answer this question, many times leading researchers to simply put their face in the big black box that web tracking data can be.

In this presentation, I will discuss some of the challenges that I have identified over the years which introduce uncertainty to the results obtained with web tracking data. These challenges can be both on the data collection side (tracking undercoverage or shared devices), on the extracting and processing side (errors from the panel provider when creating the datasets), or on the conceptualisation stage (do I even understand what I want to measure?). Additionally, I will show some of the strategies that I have been using to deal with this uncertainty, either to reduce it or to embrace it, while showing the limitations that I have found when applying them in real life. Finally, I will present some of the results that I have found over the years that have perplexed me, making me doubt about the value of web tracking research, in the hopes that transparency might lead to understanding.

Presenter's bio:

Oriol is a Postdoctoral Researcher at the Leverhulme Centre for Demographic Science, University of Oxford. He is also a fellow at Nuffield College and at RECSM, Universitat Pompeu Fabra. Before joining Oxford, Oriol completed a PhD in Social Research Methods from the London School of Economics. As a methodologist, Oriol focuses on understanding how to better collect and analyse attitudinal and behavioural data for the social sciences. He specializes in topics related to web and mobile surveys and the use of digital trace data and sensors to enhance or substitute surveys. His work, 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 donations 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 new data donation infrastructure.

KEYNOTE: New Data Types and Surveys: Opportunities and Challenges

Mick Couper

Abstract

The rapid development of new technologies and data types brings exciting opportunities for expanding and enhancing survey data collection beyond traditional modes and questions, as evidenced by this workshop. But the introduction of these new technologies into survey data collection also raises questions about the fundamental goals of survey research. Survey design often involves trade-offs, e.g., between measurement and representation, or between standardization and customization. In this talk, I plan to discuss some of these trade-offs, especially with regard to new technologies and data types, using both an historical and a forward-looking perspective. It is more of a philosophical presentation than an empirical one. My goal is to challenge those of us working in this area to think beyond the current new technology or tool.

Bio

Mick P. Couper is a Research Professor in the Survey Research Center, Institute for Social Research, at the University of Michigan. He has a Ph.D. in Sociology from Rhodes University, an M.A. in Applied Social Research from the University of Michigan and an M.Soc.Sc. from the University of Cape Town. He has been doing surveys and research on surveys for over 35 years. He is author of Designing Effective Web Surveys (2008), co-author of The Science of Web Surveys (2013), co-author of Nonresponse in Household Interview Surveys (1998), co-author of Survey Methodology (2 nd ed., 2009), and chief editor of Computer Assisted Survey Information Collection (1998) and has published extensively in a variety of journals. His research has focused on the application of technology to the survey process, the design of computer-assisted surveys, and the data collection process, including issues of coverage, nonresponse, and measurement. His current research focuses on the design and implementation of Internet surveys (including smartphones and wearables); alternative modes of data collection; informed consent; and recruitment and retention to panel and cohort studies.

Reach out!

https://www.upf.edu/web/webdataopp
webdataopp@upf.edu