

Workshop "Digital Trace Data in Social Science Research"

Essen, 06/26 - 06/28/2024

Lecturers: Prof. Dr. Frauke Kreuter, Prof. Dr. Christoph Kern,
Dr. Anna-Carolina Haensch, Clara Strasser Ceballos

Organizers: Dr. Birgit Zeyer-Gliozzo, Dr. Carolin Kunz, Prof. Dr. Martina Brandt

The workshop is structured into two main sections:

1. Overview of Digital Trace Data: This section introduces digital trace data, exploring various sources like e-learning systems, websites, smartphone apps, and sensors in wearables. Key aspects covered include the typical characteristics of these data, data quality, and their potential for social and cultural science research, along with the prerequisites for leveraging these potentials. Special attention is given to social media data from platforms like YouTube, Reddit, and TikTok. The session includes both theoretical discussion and practical data collection exercises using the statistical programming language R.

2. Analysis of Digital Trace Data: The second section delves into the analysis of the data discussed earlier. It begins with an introduction to supervised and unsupervised machine learning, covering use cases and methods. It then focuses on specific applications: text classification models (an example of supervised learning) and topic modeling (an example of unsupervised learning). Participants will engage in practical R exercises to consolidate their learning. The workshop concludes with a forward-looking segment that explores the application of these methods to other data formats, such as analyzing open-ended responses in traditional survey data.

Register here until May 31, 2024: workshop.fair@tu-dortmund.de Please provide your name, institution (e.g. TU Dortmund), faculty (e.g. Faculty of Social Science), status (e.g., doctoral researcher, Postdoc) and a short overview of your experience and your goals regarding digital trace data in social science research (max. 100 words).

Funded by:

Ministry of Culture and Science
of the State of
North Rhine-Westphalia



Block 1: Overview of Digital Trace Data		
06/26/2024	Social media data	
10-11:30 am	Overview of data sources and data collection	Caro Haensch
	<i>Lunch break</i>	
	General insight into digital trace data	
1-2:30 pm	Data quality	Frauke Kreuter (online via Zoom)
	<i>Coffee break</i>	
3:30-5 pm	Overview of data sources and data collection II	Caro Haensch

06/27/2024	Data collection	
10 am-5 pm	Hands-on data collection through Scraping and APIs	Caro Haensch & Clara Strasser Ceballos

Block 2: Analysis of Digital Trace Data		
06/28/2024	Analysis	
10-11:30 am	Overview: supervised and unsupervised machine learning	Christoph Kern & Clara Strasser Ceballos
	<i>Lunch break</i>	
1-2:30 pm	Classification models for text (Supervised ML)	Christoph Kern & Clara Strasser Ceballos
	<i>Coffee break</i>	
3-5 pm	Topic modeling (unsupervised ML)	Christoph Kern & Clara Strasser Ceballos