

Protection against re-identification under linkage of data

Lukas Plätz
Chair of Theoretical Computer Science
Supervisor: Maïke Buchin

Jan Schmutzler
Chair of Cultural Psychology and Anthropology of Knowledge
Supervisor: Estrid Sørensen

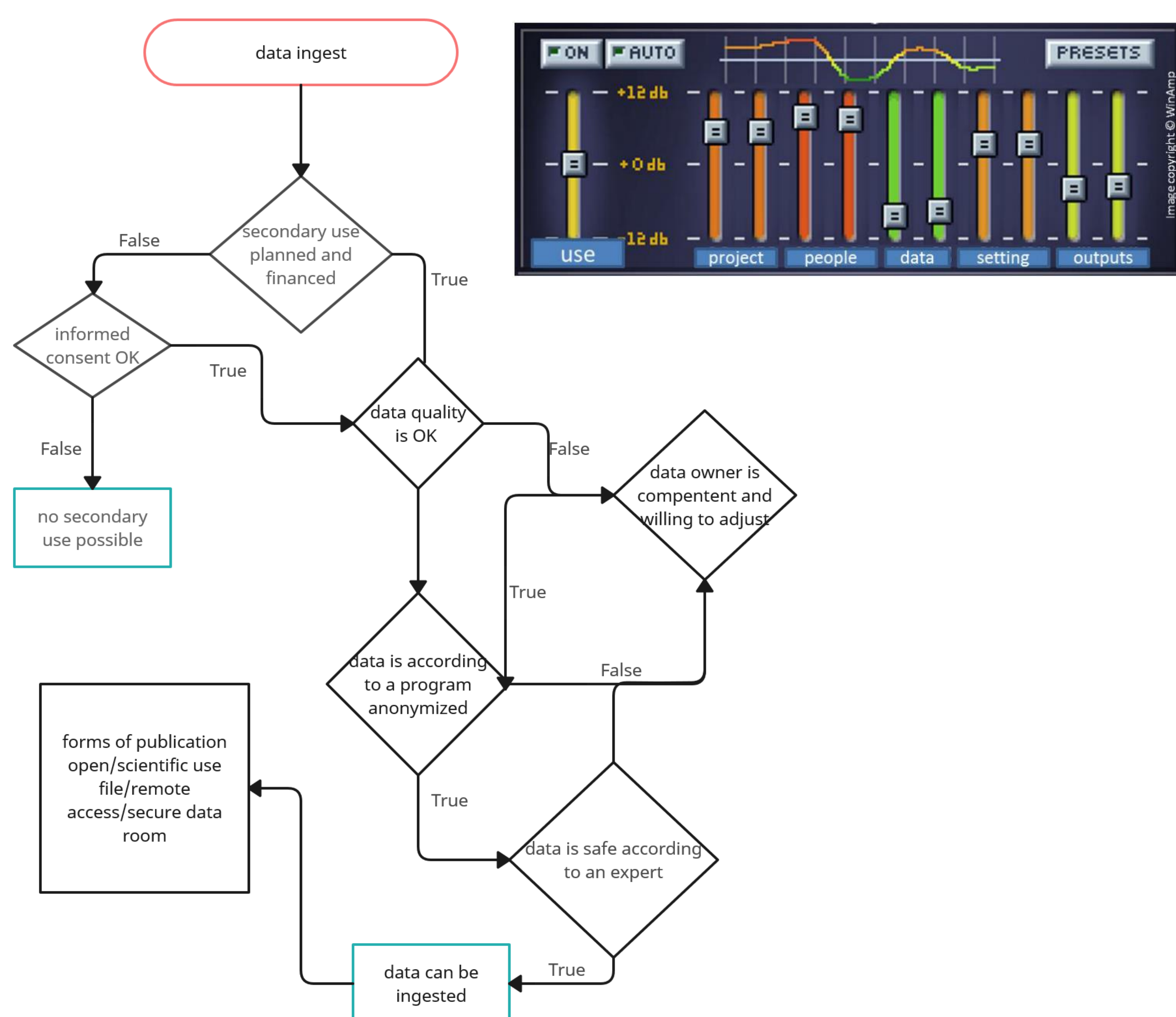
Privacy in research

Privacy is a central topic in research. It allows studies to be robust and reproducible. It also allows participants to build trust in the research team. With the General Data Protection Regulation and its implementation, the DSGVO in Germany, it is now necessary to enlighten participants about the data flow. In the publication, each participant must be guaranteed anonymity. Often this is done by pseudonymity. Here personal data which is identifying is deleted and replaced by a single not related but unique attribute. The pseudonyms allow the researcher to remove a single participant if they request it. The research in k-anonymity showed the existence of quasi-identifying attributes. Quasi-identifying attributes can lead to a re-identification when linked to a new data set.

To prevent these kinds of attacks, a series of safeguards is implemented by the data institute that forms a idiosyncratic and contingent testing cascade which built the base for our sociological analysis.

In the mathematical part of the project we tested the identification risk of quasi-identifying attributes of a large German survey in cooperation with the publishing data institute. After we identified geo-data as especially sensible we started research into trajectory data. Here we establish a new test to determine where two trajectories can be matched with some outliers. Another project is to use the r-gathering clustering algorithm to establish a strategy to k-anonymize trajectory data with a minimal loss of information.

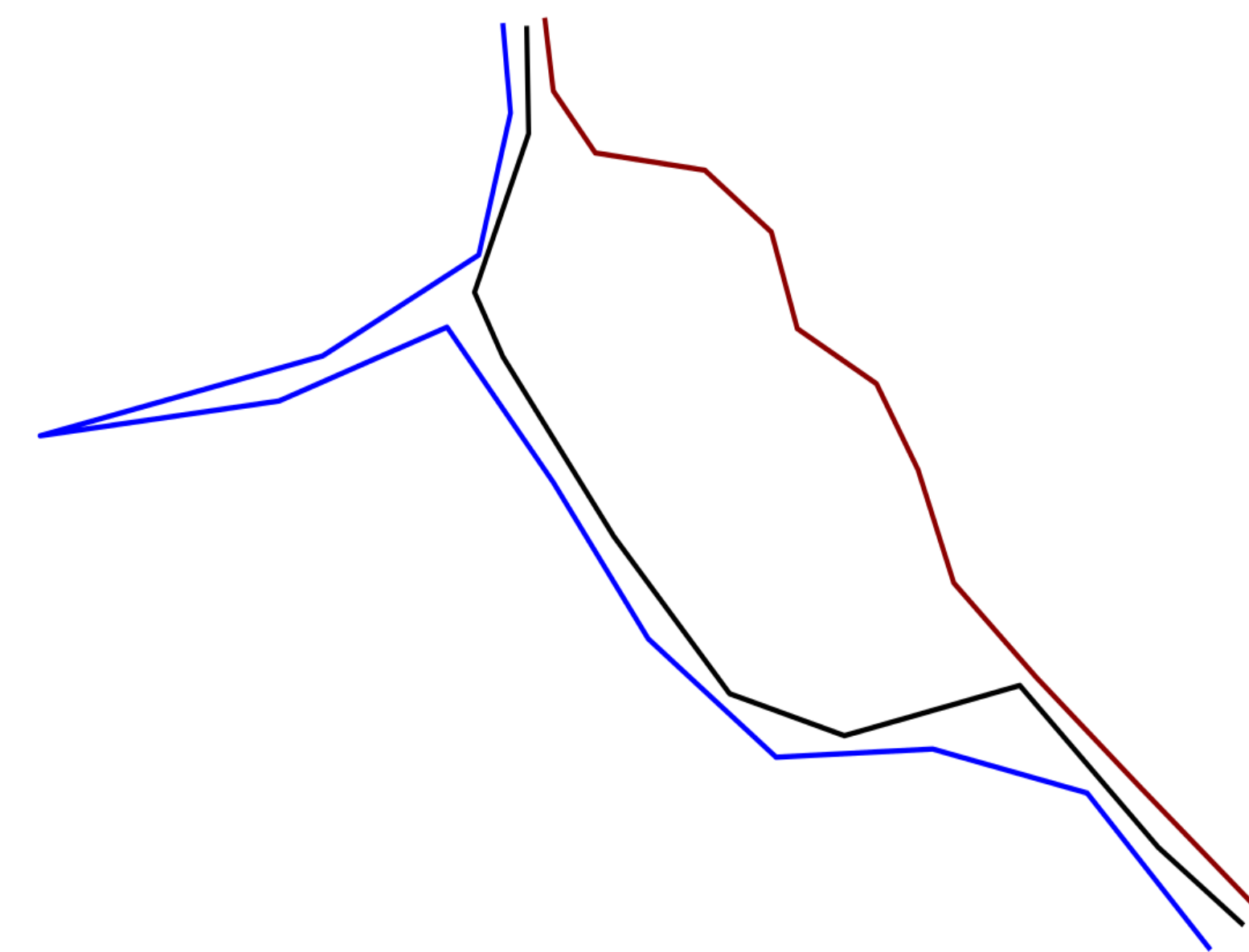
The 5 Safes as testing cascade



The 5 Safes framework is often imagined as 5 interdependent controls like on an equalizer but this picture fails to illustrate the complex testing cascade that leads to certain configurations of safety. Also a flowchart can only capture a little more complexity since the social and technical dimensions of each test can barely be simplified. Our research focusses on in-depth studies of privacy protecting practices in research in order to get a better understanding how they settle and how they evolve. Being observer and inventing own tests at the same time is the very challenge and specialty of this tandem.

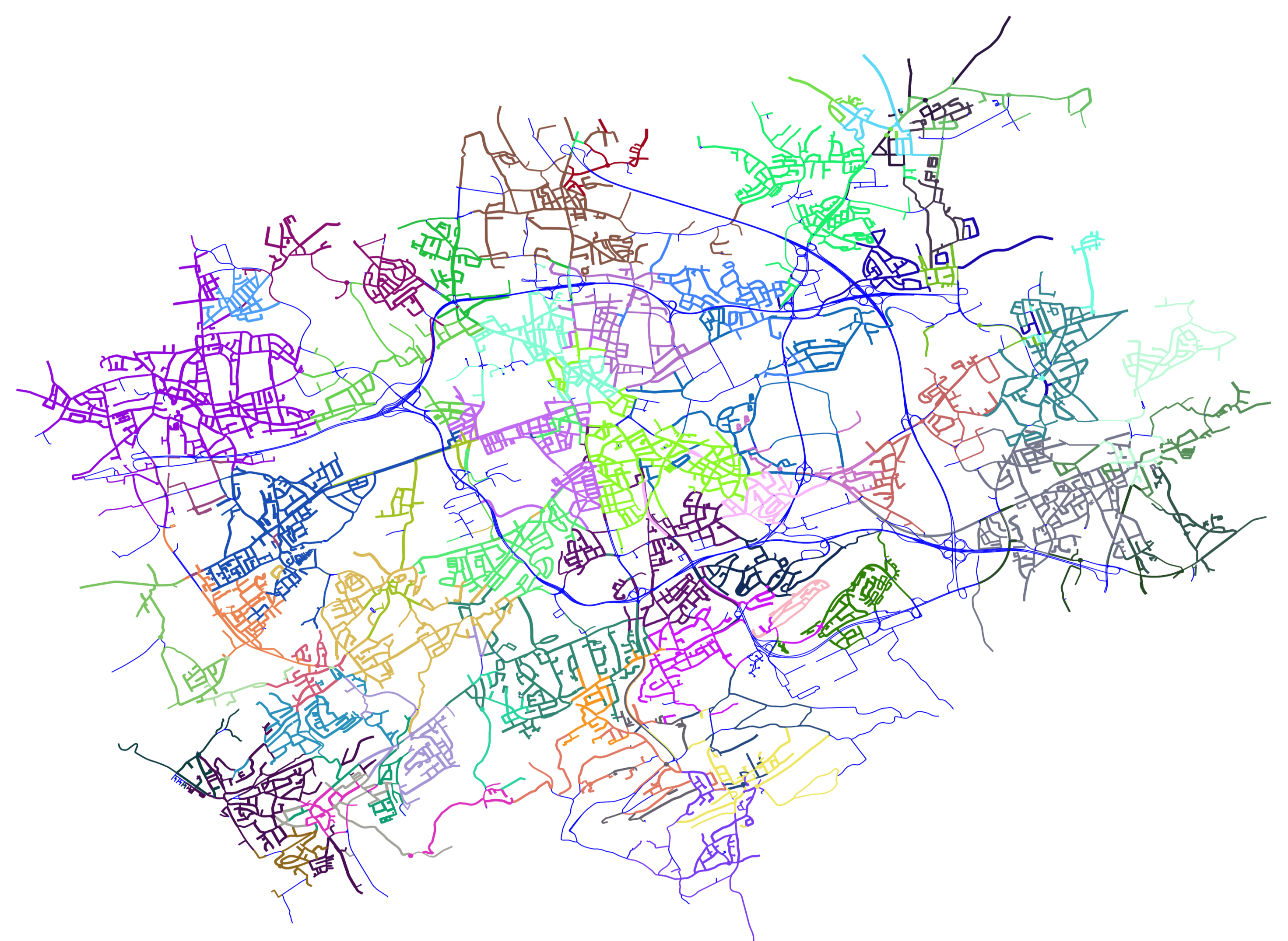
Comparing Trajectories

We produce trajectories in our daily life. Each one could be analyzed to reveal some personal information. We started to work on a distance measure for trajectories which could distinguish between individual outliers and common trajectory without the outliers.



Applying k-anonymity in navigation

Starting and Endpoints of trajectories are the most sensitive part of each journey. It is therefore essential to trim down that information before publication. To save as much as possible we tried to do that minimal via clustering points of interest with a minimum size.



Funded by:

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen



Project partners:

Fachhochschule Dortmund
University of Applied Sciences and Arts

tu technische universität dortmund



Further partners:

gesis
Leibniz-Institut für Sozialwissenschaften