「**2021** 調查研究方法與應用」學術研討會：
調查方法的挑戰與新興議題

專題演講

時間：**9** 月 **2** 日（四） **15:00-16:00**

講者：**Dr. Frauke Kreuter**

講題：**Combining Data from Multiple Sources: Examples from Economic and Public Health Research Studies**

摘要：

Combining data from different sources will be key for social scientists to take full advantage of the data deluge resulting from the increasing digitalization of society. Currently we see many attempts at using single (big data sources) with mixed results, the most exciting projects rely on a combination of different data, some still collected with traditional modes. This talk will highlight a few approaches and provide a framework with which researchers can think about creating new data products. An important element in this endeavor is however the respect of people's privacy. While different cultures have different norms about the collection on specific types of data for specific purposes, the notion of contextual integrity still holds. Learning how to design data collections for new insights in a more holistic way, will be the overarching theme of this talk.

In the talk I will be using several Economic and Public Health Research examples, in particular the IAB-SMART research project to discuss privacy issues and the approaches to create high quality combined data sources. See the attached paper for details on the privacy part. In brief: The IAB-SMART study combines data from administrative records, surveys, and digital traces from smart phones. The digital trace data are collected via an app. The purpose of the IAB- SMART study is to measure the effects of long-term unemployment on social integration and social activity, as well as the inhibiting effects of reduced social networks and activities in finding reentry into the labor market. To create measures of social integration access to the phone's address book and usage is required, as well as sensory data from accelerometer and geoposition. For valid population estimates statisticians need to account for potential coverage bias and bias due to nonresponse and measurement error. Using the case study, I will demonstrate how we approached these problems.

The 2nd example from the Global CTIS survey, a partnership between Facebook and academic institutions to create a global COVID-19 symptom survey. The survey is available in 56 languages. A representative sample of Facebook users is invited on a daily basis to report on symptoms, social distancing behavior, mental health issues, and financial constraints. Facebook provides weights to reduce nonresponse and coverage bias. Privacy protection and disclosure avoidance mechanisms are implemented by both partners to meet global policy and industry requirements. Country and region-level statistics are published daily via dashboards, and microdata are available for researchers via data use agreements. Over 1 million responses are collected weekly. We will discuss problems such partnerships face, skills needed for such large survey data collections, as well as early results from the new vaccine module.