# Harnessing event and longitudinal data in industry and health sector through privacy preserving technologies

Matthias Templ , Murat Sariyar

01.11.2023 – 31.10.2026

<https://data.snf.ch/grants/grant/211751>

# Overview

**Grant number** 211751

**Funding scheme** Bridge - Discovery

**Call** Bridge - Discovery Full proposal 2022

**Approved amount** 710,208 CHF

**Status** Ongoing

**Research institution** Zurich University of Applied Sciences - ZHAW

**Institute** Institut für Datenanalyse und Prozessdesign ZHAW Zürcher Hochschule für Angewandte Wissenschaften

An essential component of modern applications in the age of digital transformation is the analysis of data and the application of modern machine and deep learning methods. Many of the most interesting applications that affect us as individuals, either indirectly or directly, also involve the collection and processing of personal data which covers almost all aspects of our lives. Such data holds enormous potential for developing new technologies and creating added value in research and industry. Besides the use of cross-sectional data, analysis of longitudinal information - repeated observations of the same individuals over a period of time - is crucial for generating added value through the creation of profiles of individuals. However, laws on data protection fundamentally restrict the use of personal data, see e.g., the Federal Act on Data Protection of Switzerland. From the Fichen scandal in Switzerland and especially since the first data protection scandals in the U.S. in the 1990s, legislators have also been aware that the removal or pseudoanonymization of directly identifying attributes such as names, exact addresses, and AHV numbers is generally not sufficient to prevent data protection violations. The core concept of data anonymization is to transform data in such a way that the risks of re-identifying individuals are (heavily) reduced, and the reduction of the risks is weighted against the benefit of re-identifying persons (de-facto anonymity). Statistical Disclosure Control (SDC) and synthetic data generation are complex tasks for anonymization that require the application and development of adequate methods to provide data with low disclosure risk. Diverse research in SDC and implementation of methods in various software implementations have been produced in the past for cross-sectional data, but there is a large research gap in the literature and application for data containing longitudinal information. For a wide variety of longitudinal personal data, for example, related to COVID19 tracking, health and surveillance of individuals over time, or mobile traffic, there is a great deal of uncertainty about how to make such data available for research. This becomes even more virulent when data must be shared in non-aggregated form to allow for detailed analysis at the individual level, such as interpreting the results of a regression analysis applied to person-related information. The development and advancement of anonymization methods that account for complex data with longitudinal information is the central aim of this proposal. In particular, this project generates a computational and methodological framework for the anonymization of longitudinal information and event history data and to evaluate their risk through complex simulation studies and their benefits.