Deterministic data linkage between organizations: a verified procedure applied in Austria

500 words; Deadline: 13. March 2016
Florian Endel, Synthesis: Anna Dremsek, Jürgen Holl, Michael Wagner-Pinter HVB: Gottfried Endel

Background

DEXHELPP ("Decison Support for Health Policy and Planning"), a COMET K-project funded by FFG, is a collaboration between small and medium-sized businesses, the University of Technology in Vienna, Gesundheit Österreich GmbH a subsidiary of the federal ministry of health, and the main association of Austrian health insurance institutions. Analysing, sharing and linking sensitive, indirectly identifiable personalized data is a required but highly delicate process in this project.

In the federal state of Austria, responsibilities concerning the health and social security system are spread among various public institutions. As a result, data collections are detached and although a unique person identifier ("Sozialversicherungsnummer" in German) exists, legal and organisational obstacles have to be tackled to utilize or even link data.

To prepare secure and reliable data linkage of several sources at a central facility, we defined and implemented a thorough process.

Methods

The designed procedure is divided into two distinct steps. First, project dependent identifiers (called 'common ID') holding no usable information are exchanged. Second, involved data sources prepare a defined extract including the common ID and independently transfer it to the project's secure computing facility where linkage is performed.

Exchanging the common ID is the most sophisticated part of this process. Several participants are involved: 2 project partners managing research data, 2 trust centers liable for pseudonymisation and data custodians managing the collaboration.

At the beginning, one data center creates a unique random 'common ID', and transmits it, including the associated pseudonyms to the center's own data trustee. The trustee replaces the pseudonym with the identifying unique person identifier and securely transmits both variables to the second data trustee. After pseudonymisation, the second source receives the data center, consisting of the common ID associated with its own pseudonym.

Finally, a dataset including the common ID and the local pseudonym is available at both data sources. As a result, they are potentially linkable without revealing any information during this process.

Results

The whole procedure has already been performed successfully in a pilot study restricted to a certain part of Austria. Two large data sources from different collaborating organizations, potentially covering the entire Austrian population (~8.5 million inhabitants) for a period of several years took part in this first proof of concept. Tough questions about security, permission and liability have been discussed during the process and have been sufficiently answered.

It is now possible to securely transfer data to the project's independent research facility where deterministic linkage on an individual basis can be performed directly.

Conclusion

An important milestone of the DEXHELPP project has been reached by demonstrating data linkage in a complex environment. The described process not only allows to share and link data of different sources in a reliable manner, but also achieved trust of all participating organizations while ensuring privacy beyond legal requirements. Confidence and understanding of the presented procedure turned out to be a key point of this project.