The challenging implications of inconsistent data

aiming at higher efficiency in the digital search of historical persons and places, the Swiss project histHub is working on the linking and standardising of data to complement existing infrastructures and services in the area of digital humanities.

In the 14th and 15th centuries, the dynasty of the counts of Werdenberg-Sargans was very influential in today's Canton of Graubünden. In 1388, for example, Hartmann IV was elected Bishop of Chur. While the date of his election is clearly documented, the conditions surrounding it are not. Hartmann IV's year of birth is also ambiguous, his day of death, however, is considered verified to be September 6, 1416. It is quite typical for historical persons from the late Middle Ages or the early Modern Period to show a mix of clearly documented and unclear or even contradictory dates side by side. A variety of different spellings, bynames, surnames and names of places or functions is also characteristically coexisting. Hartmann IV von Werdenberg-Sargans is for example also known as Hartmann de Sanegans or, making the danger of confusion significantly higher, as Hartmann II von Chur.

Unique and linked

This example highlights a problem that typically arises within the context of historical data on persons: It is frequently very difficult to make a unique identification. In addition, many persons have been entered into various different databases. Multiple existences of database entries equally concern data on places, organisations or keywords. Linking exis-ting information to the maximum extent would therefore make it possible to compare data across suppliers as well as increase the visibility of the databases.

histHub is currently working on a solution which allows to uniquely assign and link data. The platform thereby pursues two strategies complementing each other. On the one hand, the portal supplies standard data on historical persons, places, organisations and concepts. This is achieved on the basis of typologies and controlled vocabularies. The entities are modelled in a way as to make it possible to enter even the slightest historical differences with their respective sources. Permanent IDs ensure that the data is unique for IT processing.

From Switzerland into the world

The second strategy pursued by histHub is aimed at linking existing databases in an automatic or semi-automatic way. Concordances are stored, making it possible to link "like with like". All the while, data sovereignty remains entirely with the supplier. By using innovative technologies, histHub can search and link a large number of data from third parties. The histHub linking tool for place indications for example currently generates links from more than ten international suppliers for "Sargans". Even though the project currently focuses on data pertaining to the history of Switzerland, the tools, data models and ontologies developed by histHub are certainly of interest to the international community.

Alban Frei is a historian and research manager. He acquired a doctorate from ETH Zurich and has been Managing Director of histHub since 2017.

© private