

---

# Large Card Indexes in Historical Research – Contemplating Historical Big Data

Sebastian Bondzio\*<sup>1</sup>

<sup>1</sup>Osnabrück University – Germany

## Abstract

The administrative reforms at the beginning of the 20th century implemented new cultural techniques that were conceived as rationalization, in the everyday work of German administrations and authorities. Card indexes replaced lists and books as leading storage media, were intended to provide access to information without the need for a registry and allowed a more efficient and flexible use of the knowledge stored on the card indexes. Whenever the modern state considered it necessary to collect personal information, card indexes were set up and their diligent work became more and more important.

Recently, two card indexes have been the object of two historical research projects at the University of Osnabrück:

Since spring 2018, in the research project *Überwachung. Macht. Ordnung. Personen- und Vorgangskarteien als Herrschaftsinstrument der Gestapo* the Osnabrück Gestapo card index is researched with regard to the surveillance, repression and other practices of the Nazis secret state police as well as the significance of knowledge and knowledge production in this context. The card index is one of only six Gestapo card indexes that have been preserved in whole or in part and the first to be thoroughly examined. It comprises about 49,000 index cards on about 48,000 persons, which were created between 1928 and 1945, first by the political police of Prussia and then by the Nazi secret police. In addition to biographical information and details on places of residence, it contains the 'facts' recorded by the Gestapo, in which Nazi State Police briefly recorded what made the registered person relevant to the Nazi-ideologically motivated work of the Gestapo. With the so called Osnabrücker Ausländermeldekartei<sup>1</sup>, since January 2020, the research project *Massendatenbasierte Langzeitmodelle migrationsinduziert wachsender Diversität im urbanen Kontext. Ausländerkarteien als Kulturgut und Grundlage reflexiver Migrationsforschung* for the first time ever researches a card index of a foreigners' registration office to explore the potential of this source for reflexively sensitized historical migration research and historical social geography. Between 1930 and the 1980s, in the card-index around 60,000 persons read as non-German, who had stayed in the city of Osnabrück for more than two weeks and who had become subject to registration, were registered. In addition to biographical information, the addresses were recorded, which now allow for modelling the persons itineraries through the city and the analyses of the urban space. Due to the long duration of the foreigners' registration card file of about 50 years, the changing categorisations of the municipal administration can also be traced and, by analyzing the production of knowledge, statements can be derived about its constructed reality.

In the Department of Modern History and Historical Migration Research, we understand these card indexes as functional media that modern states have used in order to cope with phenomena perceived as a social challenge and to establish social conditions that were historically considered desirable. Card indexes became part of the repertoire of Western cultural

---

\*Speaker

techniques in the 20th century, were repeatedly used to register certain individuals or groups of individuals considered relevant and to provide efficient access to them. Therefore, where they have been handed down, card indexes are unique sources for researching into the historical processes they monitored.

In addition to the space- and time-critical modelling of the processes depicted in the card indexes (e.g. with the help of GIS and other forms of KDD), we are also interested in taking a reflective position and to work out the working practices<sup>[1]</sup> of the institutions in charge of the card indexes as 'agents' (Gestapo/ a Federal Republican city administration) as well as their constructions of reality and world experiences materialized in the card index.

These work steps require a comprehensive, exploratory examination of all information stored in the respective file, which, given the amount of information, can only be done digitally. For this reason, we not only digitized the two card indexes, but had all the information recorded on the index cards extracted by an OCR- and HTR service provider, who is using recurrent neural networks and other AI. The historical mass data obtained in this way are now available to the research projects in machine-readable form and will be modelled in corresponding databases in such a way that they allow for a digital simulation of historical operation of the card index.

In addition to the guarantee of producing digital and machine-readable replicas of the sources, which ideally differ from the original sources only in their materiality, it was particularly important to us to make any further processing, abstraction and categorization of the information transparent and to be able to return to the original card file at any time (whether in the database or to the physical source).

This data modelling, which happens in extremely close proximity to the source, on the one hand open up our classifications during data preparation for scrutiny and enables the reflexive work steps of the research projects, on the other hand guarantees the reusability of the data sets. Once the project periods have expired, subsequent historians can use the digital replicas of the card indexes in state archives as a substitute for the actual source work, integrate the data sets in whole or in part into their research, carry out their own data classifications and thus, almost without restriction, pose their own, as yet unthinkable questions to the data.

In my paper I would like to take the opportunity to discuss the data aspect of the two research projects of the Department of Modern History and Historical Migration Research. This essential part of digital history research activity has so far been little discussed and often accepted uncritically as a prerequisite for the findings produced. Therefore, I would like to present our workflow from source to database as a practical case study of the generation and modelling of Historical Big Data, explain the historiographical ideas and needs underlying our data modelling, and finally, by presenting some empirical results, show the importance and potential of an explorative Data Driven History for the future of historiography.

'Practices' and 'Agents' are understood as Elements of Giddens 'Structuration Theory'. See: Anthony Giddens, *The Constitution of Society. Outlines of the Theory of Structuration*, Berkeley 1986.

**Keywords:** Data Driven History, Historical Big Data, Card Files, Serial Sources, Digital History, Modelling, Space, Time, Agents